



**HAL**  
open science

# Error Bounds for a Kernel-Based Constrained Optimal Smoothing Approximation

Laurence Grammont, François Bachoc, Andrés F. López-Lopera

► **To cite this version:**

Laurence Grammont, François Bachoc, Andrés F. López-Lopera. Error Bounds for a Kernel-Based Constrained Optimal Smoothing Approximation. 2024. hal-04645086

**HAL Id: hal-04645086**

**<https://hal.science/hal-04645086v1>**

Preprint submitted on 11 Jul 2024

**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.

# Error Bounds for a Kernel-Based Constrained Optimal Smoothing Approximation

Laurence Grammont<sup>1,\*</sup>, François Bachoc<sup>2</sup> and Andrés F. López-Lopera<sup>3</sup>

<sup>1</sup>Univ. Jean Monnet, CNRS, ECL, INSA Lyon, UCB Lyon 1, ICJ UMR5208, F-42023 St-Etienne, France.

<sup>2</sup>Institut de Mathématiques de Toulouse, Univ. Paul Sabatier, F-31062 Toulouse, France.

<sup>3</sup>Univ. Polytechnique Hauts-de-France, CERAMATHS, F-59313 Valenciennes, France.

\*Corresponding author

## Abstract

This paper establishes error bounds for the convergence of a piecewise linear approximation of the constrained optimal smoothing problem posed in a reproducing kernel Hilbert space (RKHS). This problem can be reformulated as a Bayesian estimation problem involving a Gaussian process related to the kernel of the RKHS. Consequently, error bounds can be interpreted as a quantification of the maximum a posteriori (MAP) accuracy. To our knowledge, no error bounds have been proposed for this type of problem so far. The convergence results are provided as a function of the grid size, the regularity of the kernel, and the distance from the kernel interpolant of the approximation to the set of constraints. Inspired by the MaxMod algorithm from recent literature, which sequentially allocates knots for the piecewise linear approximation, we conduct our analysis for non-equispaced knots. These knots are even allowed to be non-dense, which impacts the definition of the optimal smoothing solution and our error bound quantifiers. Finally, we illustrate our theorems through several numerical experiments involving constraints such as boundedness and monotonicity.

**Keywords:** Gaussian processes, inequality constraints, maximum a posteriori, reproducing kernel Hilbert space, basis function approximation, approximation error, asymptotic convergence.

## 1 Introduction

Consider a nonempty set  $\Omega$  of  $\mathbb{R}^d$  and a set  $E$  of functions from  $\Omega$  to  $\mathbb{R}$ . Given data  $(x_i, y_i)_{i=1}^n \in \Omega \times \mathbb{R}$ , the **smoothing problem** is to find a function  $\hat{u} \in \mathcal{H}$  solution of

$$\min_{u \in \mathcal{H}} \|u\|_{\mathcal{H}}^2 + \frac{1}{\tau} \sum_{i=1}^n (u(x_i) - y_i)^2, \quad (1)$$

where  $\mathcal{H}$  is the reproducing kernel Hilbert space (RKHS) defined by a kernel  $K$  on  $\Omega \times \Omega$ , with  $\tau > 0$ . Hence,  $\mathcal{H}$  is a Hilbert space included in  $E$ , and we let  $\|\cdot\|_{\mathcal{H}}$  be its Hilbert norm. Considering an RKHS allows the solution of (1) to be interpreted as a Bayesian estimator, involving a Gaussian process (GP)  $Y$  with covariance function  $K$ :  $K(x, x') = \text{cov}(Y(x), Y(x'))$ . We refer for instance to [Stein, 1999, Rasmussen and Williams, 2005, Karvonen, 2023] for references on GPs. In the Bayesian framework,  $\tau > 0$  is the noise variance term. Kimeldorf and Wahba [1970] prove that  $\hat{u}$  is the mean of the GP  $Y$  conditionally to noisy observations:

$$\hat{u}(t) = \mathbb{E}[Y(t) | Y_1 = y_1, \dots, Y_n = y_n],$$

where  $\mathbb{E}$  denotes the expectation of random variables. For the noisy case, the Bayesian model is  $Y_i = Y(x_i) + \mathcal{E}_i$  for all  $i = 1, \dots, n$ , where  $\mathcal{E} = (\mathcal{E}_i)_i \sim \mathcal{N}(0, \tau I)$  is an independent centered Gaussian vector. Here,  $I$  is the  $n \times n$  identity matrix. For simplicity, we omit the subscripts to the identity matrices as their sizes will be clear from the context. The solution  $\hat{u}$  is then given by

$$\hat{u}(t) = k_n^\top(t) (K_n + \tau I)^{-1} y,$$

where  $k_n(t) = [K(t, x_1), \dots, K(t, x_n)]^\top$ ,  $K_n = (K(x_i, x_j))_{1 \leq i, j \leq n}$  and  $y = [y_1, \dots, y_n]^\top$ .

If an additional constraint is added to (1), given by a closed convex set  $C$  of functions, we obtain the **constrained smoothing problem** of finding a function  $\hat{u}$ , in  $\mathcal{H} \cap C$ , solution of

$$\min_{u \in \mathcal{H} \cap C} \|u\|_{\mathcal{H}}^2 + \frac{1}{\tau} \sum_{i=1}^n (u(x_i) - y_i)^2. \quad (2)$$

This problem can be rewritten as a constrained GP model so that the solution can be interpreted as a Bayesian estimation [Grammont et al., 2024]. Two important examples in practice are when  $C$  is composed of bounded or componentwise monotonic functions. We refer to [Bellec, 2018, Durot, 2002, Cousin et al., 2016, Durot and Lopuhaä, 2018, Groeneboom and Jongbloed, 2014, Golchi et al., 2015, Groeneboom et al., 2001, Hornung, 1978, Lin and Dunson, 2014, López-Lopera et al., 2018, López-Lopera et al., 2019, López-Lopera et al., 2020, Maatouk and Bay, 2017, Riihimäki and Vehtari, 2010, Zhou et al., 2019] for consideration of these constraints with GPs and more generally in statistics.

Unlike the unconstrained smoothing problem in (1), there is no explicit expression for the solution in the constrained case, thus a numerical approximation of  $\hat{u}$  is required. For clarity, we restrict our study to the one-dimensional setting  $\Omega = [0, 1]$ . Nevertheless, as explained in Remark 5.1, the techniques we develop can be extended to the general  $d$ -dimensional case, albeit with more cumbersome notations.

A fruitful approach for numerical approximation is to consider piecewise linear finite-dimensional kernels, RKHSs and GPs [Bachoc et al., 2019, Cousin et al., 2016, Grammont et al., 2024, López-Lopera et al., 2018, Maatouk and Bay, 2017, Zhou et al., 2019]. In particular, we consider the approximate solution  $\hat{u}_N$  in the RKHS defined by  $K_N$ , the covariance function of a finite-dimensional GP  $Y_N$  approximating the GP  $Y$ . Here,  $N \in \mathbb{N}$  is the number of knots defining the piecewise linear approximation. The function  $\hat{u}_N$  is then the solution to a **constrained discretized smoothing problem**, and also the maximum a posteriori (MAP) of the posterior distribution of the constrained finite-dimensional GP  $Y_N$ . In [Grammont et al., 2024], it is shown that  $\hat{u}_N$  converges to  $\hat{u}$  as  $N \rightarrow \infty$  for fixed data  $(x_i, y_i)_{i=1}^n$ , however, no error bounds are provided for this convergence. More generally, to our knowledge, no general error bounds have been provided for numerical approximations of the constrained optimal smoothing problem or for the equivalent formulation with the MAP. This is the ambitious aim of this paper.

An error estimation is always highly dependent on regularity, which is related to the function space in which the exact solution is sought. In this paper, this function space is determined by the kernel  $K$  that we assume to be  $\beta$ -Hölder,  $0 < \beta \leq 1$ . We note that Hölder-continuity is a very standard regularity measure in statistics and machine learning for functions that are not necessarily differentiable [see, e.g., Locatelli and Carpentier, 2018]. In addition, our error bound construction would not benefit from a stronger regularity than Lipschitzness (1-Hölder continuity), because of piecewise linearity (see Remark 5.2).

To provide as much generality as possible, we allow for non-equispaced knots defining the finite-dimensional approximation, and we even allow the sequence of knots not to be dense in the input space. Non-equispaced knots enable higher accuracy for a given computational budget  $N$  [Bachoc et al., 2022, López-Lopera et al., 2022] and can be selected automatically by the MaxMod algorithm introduced in [Bachoc et al., 2022]. Furthermore, the convergence proof of MaxMod includes an intermediary step analyzing convergences for non-dense knots. This justifies our consideration of non-dense knots when providing error bounds.

To account for non-dense and non-equispaced knots, we measure the asymptotic density of the  $N$  knots by the specific grid size  $\delta_N$  defined in (8). Ultimately, the decay rate of our error bound depends on  $\beta$  (i.e. the regularity of the GP), the grid size  $\delta_N$ , and a third factor we have highlighted: the distance of the kernel interpolant of the approximate solution to the set of constraints  $C_F$ . In the following, this third factor is denoted as  $\alpha_N$  (see definition in (28)). Furthermore, for non-dense knots, the limit function to  $\hat{u}_N$  is not  $\hat{u}$  as in (2), but the function  $\hat{u}_F$  defined in (5), which depends on the closure set  $F$  of the sequence of knots. This definition relies on the notion of multiaffine extension introduced in [Bachoc et al., 2022]. Naturally, when  $F = [0, 1]$ , we have  $\hat{u}_F = \hat{u}$ .

Our final and general error bounds are provided in Theorems 5.1 and 5.2. In Theorem 5.1, we

consider the case where  $\alpha_N = 0$ , meaning the kernel interpolant of the approximate solution satisfies the constraints  $C_F$ . Here, the error bound is of order  $\mathcal{O}(\delta_N^{\beta/4})$ . In Theorem 5.2, we address the most general case, where this kernel interpolant does not necessarily meet the constraints. In this scenario, the error bound also depends on  $\alpha_N$ . We demonstrate that  $\alpha_N$  approaches zero (see Remark 5.3), though we leave open the problem of quantifying its decay rate.

In numerical experiments, we illustrate and validate our theoretical results through various synthetic examples that account for different types of inequality constraints (e.g., boundedness and monotonicity) and regularity assumptions (i.e. smoothness of the kernel). We examine both scenarios where the knots are dense and where they are not. Numerically, we confirm the convergence as  $N \rightarrow \infty$ , observing a faster convergence for larger regularity  $\beta$ .

The paper is organized as follows. Section 2 presents the constrained optimal smoothing problem, introducing  $\hat{u}_F$ , our regularity indicators and the multi-affine extension, and stating various of their properties. Section 3 discusses the numerical approximation of constrained optimal smoothing, denoted as  $\hat{u}_N$ , and states its existence and unicity. Section 4 focuses on the quantitative properties of the set of approximants involved in  $\hat{u}_N$ , which are essential for the final bounds. Sections 5 and 6 provide the error bounds (with Theorems 5.1 and 5.2) and the numerical experiments, respectively. Section 7 concludes the paper.

Several proofs are included in the main text to elucidate the construction of the error bounds. Additional proofs, primarily technical or containing pre-existing concepts for completeness, are in Appendix A.

## 2 Framework for the constrained optimal smoothing problem

In this section, we present the tools needed to pose the constrained optimal smoothing problem properly. Then we prove, with classical arguments, that the problem has a unique solution with a regularity that we outline.

To ensure that our approach is easiest to understand, we have chosen to focus on the one-dimensional case. Remark 5.1 discusses the extension to the multi-dimensional case. Hence, in this paper,  $E$  is the set of real valued continuous functions on  $\Omega = [0, 1]$ . Let  $F$  be a compact subset of  $[0, 1]$  containing 0 and 1.  $F$  will be the closure of the set of knots required in the discretization process of the approximation (see Section 3). As discussed in Section 1, allowing for a general  $F$  that does not coincide with the entire set  $\Omega$  is an intermediary step in the convergence proof of MaxMod in [Bachoc et al., 2022]. This justifies our interest here to allow for a general  $F$ .

We define  $E_F$  as the set of real-valued continuous functions restricted to  $F$ , endowed with the supremum norm:

$$\|u\|_\infty = \max_{t \in F} |u(t)|,$$

for  $u \in E_F$ . Let  $\mathcal{H}_F$  and  $\mathcal{H}$  be the RKHSs associated to the centered GPs  $(Y(x))_{x \in F}$  and  $(Y(x))_{x \in [0,1]}$  (respectively), both having covariance function  $K$ . The Hilbert scalar product and norm for  $\mathcal{H}$  are written as  $\langle \cdot, \cdot \rangle_{\mathcal{H}}$  and  $\|\cdot\|_{\mathcal{H}}$ . Similarly, for  $\mathcal{H}_F$  they are written as  $\langle \cdot, \cdot \rangle_{\mathcal{H}_F}$  and  $\|\cdot\|_{\mathcal{H}_F}$ . Let us notice that  $\mathcal{H}_F = \mathcal{H}|_F$ , where

$$\mathcal{H}|_F = \{g : F \rightarrow \mathbb{R}; \exists f \in \mathcal{H} \text{ s.t. } \forall x \in F, g(x) = f(x)\},$$

see [Berlinet and Thomas-Agnan, 2011, Theorem 6]. Both  $\mathcal{H}$  and  $\mathcal{H}_F$  rely on the reproducing kernel  $K$ , which is always symmetric and positive semi-definite.

The regularity of  $K$  has a major influence on the error bounds. Here, regularity is measured by Hölder-continuity (recall the corresponding discussion in Section 1 and Remark 5.2). Let  $\beta$  be a real number such that

$$0 < \beta \leq 1.$$

**Definition 2.1.** *A function  $f$  is  $\beta$ -Hölder continuous on  $[0, 1]$  if there exists a constant  $c_f > 0$  such that,  $\forall s, t \in [0, 1]$ ,*

$$|f(s) - f(t)| \leq c_f |s - t|^\beta.$$

**Condition 2.1.** *The reproducing kernel  $K$  is  $\beta$ -Hölder continuous with respect to both its inputs. That is, there exists a constant  $c_K > 0$  such that*

$$\forall u, s, t \in [0, 1], \quad |K(u, s) - K(u, t)| \leq c_K |s - t|^\beta.$$

In the following, we will define quantities which are capable of reflecting the regularity of the problem and which will be useful in setting the error bounds. Let us define the modulus of continuity of a function defined and continuous on a compact subset  $S$  of  $[0, 1]$ :

$$M_f(\delta) = \begin{cases} \sup_{s, t \in S, |s-t| \leq \delta} |f(s) - f(t)| & \text{if } \delta \leq 1, \\ M_f(1) & \text{if } \delta \geq 1. \end{cases} \quad (3)$$

Let us define the following quantity intended to be an indicator of regularity:

$$\Psi_f(\delta) = \sup_{t \geq 1} \frac{M_f(t\delta)}{t}. \quad (4)$$

The indicator  $\Psi_f$  is adapted to the hat functions considered in Section 3 for the numerical approximation, and is then involved in the proofs of our main error bounds. The following proposition will clarify its order of magnitude according to the regularity of  $f$ .

**Proposition 2.1.** *Let  $S$  be a compact subset of  $[0, 1]$ . If  $f$  is continuous on  $S$ ,*

$$\lim_{\delta \rightarrow 0} \Psi_f(\delta) = 0.$$

*If  $f$  is  $\beta$ -Hölder continuous on  $S$ ,*

$$\Psi_f(\delta) \leq c_f \delta^\beta.$$

*Proof.* The proof is postponed to Appendix A. □

Then, we define the multi-affine extension announced in Section 1. This extension will allow us to define the constrained optimal smoothing problem for functions defined on  $F$  rather than on  $[0, 1]$ . Then, this extension will allow us to define the limit function  $\hat{u}_F$  discussed in Section 1, to which this paper provides upper bounds.

**Definition 2.2.** *For  $t \in [0, 1]$ ,*

- *If  $t \in F$ , then define  $t^- = t^+ = t$  and  $w_-(t) = w_+(t) = 1/2$ .*
- *If  $t \notin F$ , then define  $t^- = \sup\{x, x \in F, x \leq t\}$ ,  $t^+ = \inf\{x, x \in F, x \geq t\}$ , and*

$$\begin{aligned} w_-(t) &= (t^+ - t)/(t^+ - t^-), \\ w_+(t) &= (t - t^-)/(t^+ - t^-). \end{aligned}$$

*Define the operator  $P : E_F \rightarrow E$  as follows. For all  $u \in E_F$ ,*

$$P(u)(t) = u(t^-)w_-(t) + u(t^+)w_+(t),$$

*and call  $P(u) \in E$  the multi-affine extension of  $u$ .*

We remark that, in Definition 2.2, the multi-affine extension could also be called the affine extension. Nevertheless, Definition 2.2 is extended to the general multi-dimensional case in [Bachoc et al., 2022], where the name multi-affine extension is appropriate. Hence, for clarity, we will refer to  $P(u)$  in Definition 2.2 as a multi-affine extension also in the one-dimensional exposition of this paper.

**Proposition 2.2.**  *$P(u)$  is the unique function in  $E$  equal to  $u$  on  $F$  and affine on the intervals of  $[0, 1] \setminus F$ . Moreover the map  $u \rightarrow P(u)$  is linear and 1-Lipschitz from  $E_F$  to  $E$  equipped with the supremum norm. In particular, it preserves uniform convergence.*

*Proof.* See [Bachoc et al., 2022, Proposition 4.4].  $\square$

We recall the constraint set  $C$ , which is assumed to be closed and convex in  $E$ . Let us then define the set of functions on  $F$  which multi-affine extensions are in  $C$ .

$$C_F = \{f \in E_F : P(f) \in C\}.$$

The next condition guarantees the compatibility of the constraints with the kernel  $K$  and the set  $F$ .

**Condition 2.2.**  $\mathcal{H}_F \cap C_F \neq \emptyset$ .

Condition 2.2 means that there is a function in the RKHS  $\mathcal{H}_F$  which multi-affine extension satisfies the constraints. It is clear that our definition of  $\hat{u}_F$  in (5) below needs this condition, where  $\hat{u}_F$  is discussed in Section 1. For the standard cases where  $C$  is the a set of bounded, monotonous or convex functions, Condition 2.2 holds for many standard classes of kernels, see [Bachoc et al., 2022] and references therein.

Since  $P$  is linear and  $C$  is convex,  $C_F$  is a convex set of  $E_F$ . We consider the constrained optimal smoothing:

$$\hat{u}_F = \arg \min_{v \in \mathcal{H}_F \cap C_F} J_F(v), \quad (5)$$

where, for  $v \in \mathcal{H}_F$ ,

$$J_F(v) = \|v\|_{\mathcal{H}_F}^2 + \frac{1}{\tau} \sum_{i=1}^n (P(v(x_i)) - y_i)^2. \quad (6)$$

The problem (5) is the extension of the problem (2) announced in Section 1, taking into account that the closure set  $F$  is allowed to be different from  $[0, 1]$ . Next, we introduce the standard notion of strong convexity which we will use in this paper to show the existence and unicity of  $\hat{u}_F$  and also to obtain our error bounds.

**Definition 2.3.** For a Hilbert space  $V$  and a function  $g : V \rightarrow \mathbb{R}$ , we say that  $g$  is strongly convex with parameter  $m$  if and only if for all  $u, v \in V$  and  $t \in [0, 1]$ , we have

$$g(tu + (1-t)v) \leq tg(u) + (1-t)g(v) - \frac{m}{2}t(1-t)\|u - v\|_V^2.$$

**Proposition 2.3.** The function  $v \in \mathcal{H}_F \mapsto J_F(v)$  is strongly convex with parameter  $m = 2$ .

*Proof.* It is well-known that the function  $v \mapsto \|v\|_{\mathcal{H}_F}^2$  is strongly convex with parameter  $m = 2$ . Furthermore, the function

$$v \mapsto \frac{1}{\tau} \sum_{i=1}^n (P(v(x_i)) - y_i)^2,$$

is convex as the composition of the affine function  $v \mapsto \frac{1}{\sqrt{\tau}}(Pv(x_i) - y_i)_{i=1, \dots, n}$  by the squared Euclidean norm which is convex. Hence,  $J_F$ , which is the sum of the two functions, is strongly convex with parameter  $m = 2$ .  $\square$

The next lemma addresses the optimization space  $\mathcal{H}_F \cap C_F$  in (5). To prove this lemma and throughout this paper, we will exploit the reproducing property [Berlinet and Thomas-Agnan, 2011] in the RKHSs  $\mathcal{H}$  and  $\mathcal{H}_F$ : for any  $f \in \mathcal{H}$  and  $x \in [0, 1]$ , we have  $f(x) = \langle f, K(x, \cdot) \rangle_{\mathcal{H}}$ . Similarly for  $f \in \mathcal{H}_F$  and  $x \in F$ , we have  $f(x) = \langle f, K(x, \cdot) \rangle_{\mathcal{H}_F}$ .

**Lemma 2.1.** The set  $\mathcal{H}_F \cap C_F$  is a closed subset of  $\mathcal{H}_F$  w.r.t.  $\|\cdot\|_{\mathcal{H}_F}$ .

*Proof.* We have  $\mathcal{H}_F \cap C_F = \{v \in \mathcal{H}_F : Pv \in C\}$ . Let us fix  $v \in \mathcal{H}_F$  and consider a sequence  $v_n \in \mathcal{H}_F \cap C_F$  such that  $v_n \rightarrow v$  for  $\|\cdot\|_{\mathcal{H}_F}$ . By the reproducing property,

$$\|v_n - v\|_{\infty} \leq \|v_n - v\|_{\mathcal{H}_F} \sup_{t \in F} \sqrt{K(t, t)}.$$

As  $K$  is continuous, we deduce that  $\|v_n - v\|_{\infty} \rightarrow 0$ . Now  $P$  is 1-Lipschitz with respect to  $\|\cdot\|_{\infty}$  and thus continuous. Hence  $Pv_n \rightarrow Pv$  for  $\|\cdot\|_{\infty}$ . As  $C$  is a closed subset of  $E$ , then  $Pv \in C$ . This proves that  $v \in C_F$  and thus  $v \in \mathcal{H}_F \cap C_F$ .  $\square$

Finally, we can state the existence and unicity of  $\hat{u}_F$ , and additionally quantify its regularity.

**Theorem 2.1.** *The constrained optimal smoothing problem in (5) has a unique solution  $\hat{u}_F$ . Moreover  $\hat{u}_F$  is  $\frac{\beta}{2}$ -Hölder continuous with constant  $\sqrt{2c_K}\|\hat{u}_F\|_{\mathcal{H}_F}$ , recalling  $c_K$  from Condition 2.1.*

*Proof.* As  $J_F(v) \geq \|v\|_{H_F}^2$ , we have  $\lim_{\|v\| \rightarrow \infty} J_F(v) = \infty$ . Furthermore, as  $C_F$  is a convex set of  $E_F$ , as  $\mathcal{H}_F \cap C_F \neq \emptyset$  (Condition 2.2) and from Lemma 2.1,  $\mathcal{H}_F \cap C_F$  is a non empty closed convex set of  $E_F$ . By Proposition 2.3,  $J_F$  is strictly convex. Then, as  $J_F$  is clearly continuous on  $\mathcal{H}_F$ , by [Hiriart-Urruty, 2012, Theorem 1.18], (5) has a unique solution and, as  $K$  is  $\beta$ -Hölder continuous, we have:

$$\begin{aligned} |\hat{u}_F(t) - \hat{u}_F(s)| &= |\langle \hat{u}_F(\cdot), K(\cdot, t) - K(\cdot, s) \rangle_{\mathcal{H}_F}| \\ &\leq \|\hat{u}_F\|_{\mathcal{H}_F} \|K(\cdot, t) - K(\cdot, s)\|_{\mathcal{H}_F} \\ &\leq \|\hat{u}_F\|_{\mathcal{H}_F} (K(t, t) - K(t, s) + K(s, s) - K(s, t))^{1/2} \\ &\leq \|\hat{u}_F\|_{\mathcal{H}_F} \sqrt{2c_K} |s - t|^{\beta/2}. \end{aligned} \quad \square$$

### 3 The Discrete constrained optimal smoothing Problem

We now introduce the finite-dimensional spaces where the approximate to  $\hat{u}_F$  is sought. First, we deal with the notion of grid size compatible with the closure set  $F$  of the knots discussed in Sections 1 and 2. Then, we introduce the hat basis functions given by the knots and yielding the piecewise linear approximation  $\hat{u}_{N,F}$  discussed in Section 1.

We consider a sequence of nested subdivisions, i.e. sets of knots  $S_N \subset S_{N+1}$  with

$$S_N : \quad 0 = t_1 < \dots < t_N = 1.$$

As explained in Section 2,  $F$  is the closure of  $\bigcup_{N \geq 1} S_N$ , i.e.

$$F = \overline{\bigcup_{N \geq 1} S_N}.$$

Hence  $F$  is a compact set of  $[0, 1]$  containing 0 and 1. The nodes of  $S_N$  allow the construction of the finite-dimensional approximation spaces, as we will detail below.

Before tackling this approximation space, we first define the grid size of  $S_N$ . As the set  $F$  is allowed to have holes, it is insufficient to define the grid size simply as  $\max_{i=1}^{N-1} |t_{i+1} - t_i|$ . Indeed, when  $F \neq [0, 1]$ , this quantity will not tend to zero as  $N \rightarrow \infty$ . To overcome this issue, we need a more nuanced definition of the grid size. For  $t \in F$ , we define two nodes  $t_{S_N}^- \in S_N$  and  $t_{S_N}^+ \in S_N$ , which are, respectively, the closest grid points of  $S_N$  to  $t$  on its left and right. In other words, if  $t \in S_N$ , then  $t_{S_N}^- = t_{S_N}^+ = t$ , otherwise

$$t_{S_N}^- = \max\{x, x \in S_N, x \leq t\}, \quad t_{S_N}^+ = \min\{x, x \in S_N, x \geq t\}. \quad (7)$$

Thus, we can define the grid size as

$$\delta_N = \sup_{t \in F} \min(|t - t_{S_N}^-|, |t - t_{S_N}^+|). \quad (8)$$

Note that since  $F$  is the closure of  $S_N$ , we have  $\delta_N \rightarrow 0$  as  $N \rightarrow \infty$ , which is the benefit of the definition in (8).

We now proceed to formally define the piecewise linear approximation. We define the set of hat functions  $(\phi_1, \dots, \phi_N)$ , similarly as in [Bachoc et al., 2022],

$$\phi_i(t) = \begin{cases} \frac{t - t_{i-1}}{t_i - t_{i-1}} & \text{if } i \geq 1 \text{ and } t \in [t_{i-1}, t_i], \\ \frac{t_{i+1} - t}{t_{i+1} - t_i} & \text{if } i \leq n - 1 \text{ and } t \in [t_i, t_{i+1}], \\ 0 & \text{otherwise,} \end{cases}$$

with the convention that  $t_0 = -1$  and  $t_{N+1} = 2$ . Note that a linear combination of the  $\phi_i$ 's is a piecewise linear function.

Then, as in [Bay et al., 2016, 2017, Grammont et al., 2022], we place an RKHS structure on the linear combinations of the  $\phi_i$ 's, by defining a kernel  $K_N$ . This kernel can also be seen as an approximation of  $K$ , thus corresponding to a finite-dimensional approximation of a GP with kernel  $K$ . We let  $\mathcal{H}_{N,F}$  be the finite-dimensional subspace of  $E_F$  defined by

$$\mathcal{H}_{N,F} = \text{Span}\{\phi_i|_F, i = 1, \dots, N\}.$$

Here, for a function  $g : [0, 1] \rightarrow \mathbb{R}$ , we let  $g|_F$  be the restriction of  $g$  to  $F$ , that is the function  $F \rightarrow \mathbb{R}$  defined by, for  $x \in F$ ,  $g|_F(x) = g(x)$ . On  $\mathcal{H}_{N,F}$ , we now construct the kernel  $K_N$  to obtain an RKHS. For this we first need to define the matrix

$$\Gamma_N = (K(t_i, t_j))_{1 \leq i, j \leq N},$$

and we assume Condition 3.1.

**Condition 3.1.**  $\Gamma_N$  is invertible.

This condition is verified for a strictly positive definite kernel, for instance the squared exponential kernel [see e.g. Karlin and Studden, 1966, Chapter 1, §3, Example 5].

Then, let us define the following inner product on  $\mathcal{H}_{N,F}$  :

$$\langle u, v \rangle_N = c_u^\top \Gamma_N^{-1} c_v, \quad (9)$$

where  $c_u = [u(t_1), \dots, u(t_N)]^\top$  and  $c_v = [v(t_1), \dots, v(t_N)]^\top$ . Write  $\|\cdot\|_N$  for the corresponding norm. We can now state the RKHS structure of  $\mathcal{H}_{N,F}$ .

**Proposition 3.1.** [Bay et al. [2017], Theorem 1] *The space  $\mathcal{H}_{N,F}$ , equipped with the scalar product (9), is an RKHS with kernel  $K_N$  given by*

$$\forall x, x' \in F, \quad K_N(x, x') = \sum_{i,j=1}^N K(t_i, t_j) \phi_i(x) \phi_j(x').$$

*Proof.* See the proof in [Bay et al., 2017, Theorem 1]. □

These tools allow us to define the discrete problem which will approximate the original problem (5). We can approximate (6) by the following functional, for  $u_N \in \mathcal{H}_{N,F}$ ,

$$J_{N,F}(u_N) = \|u_N\|_N^2 + \frac{1}{\tau} \sum_{i=1}^n (P(u_N)(x_i) - y_i)^2. \quad (10)$$

Hence we approximate  $\hat{u}_F$  by  $\hat{u}_{N,F}$  solution of

$$\hat{u}_{N,F} = \arg \min_{v_N \in \mathcal{H}_{N,F} \cap C_F} J_{N,F}(v_N). \quad (11)$$

Note that when  $F = [0, 1]$ ,  $\hat{u}_{N,F}$  corresponds to the approximant  $\hat{u}_N$  that was discussed in Section 1.



**Proposition 3.2.**  $J_{N,F}$  is Fréchet differentiable, strongly convex with parameter 2, and

$$\lim_{\|v_N\|_N \rightarrow \infty} J_{N,F}(v_N) = \infty.$$

*Proof.* It is obvious that  $J_N$  is Fréchet differentiable. With the same arguments as in the proof of Proposition 2.3,  $J_{N,F}$  is strongly convex with parameter 2. Finally, as  $J_{N,F}(v_N) \geq \|v_N\|_N^2$  then  $\lim_{\|v_N\|_N \rightarrow \infty} J_{N,F}(v_N) = \infty$ .  $\square$

Then, we define  $\pi_N$  as the piecewise affine interpolation associated to the subdivision  $S_N$ , defined from  $E_F$  onto  $\mathcal{H}_{N,F}$  by

$$\forall f \in E_F, \quad \pi_N(f) = \sum_{j=1}^N f(t_j) \phi_{j|F}.$$

Notice that  $\pi_N$  is a projection in the sense that  $\pi_N \circ \pi_N = \pi_N$ . The following proposition provides a first approximation property for this projection. However, it does not offer a quantitative measure of the approximation's quality. This will be addressed in Section 4.

**Proposition 3.3.** For any  $f \in \mathcal{H}_F$ , we have

$$\|\pi_N(f)\|_N \leq \|f\|_{\mathcal{H}_F},$$

$$\pi_N(f) \xrightarrow{N \rightarrow \infty} f \quad \text{in } E_F.$$

Moreover  $\mathcal{H}_F$  is characterized by

$$\mathcal{H}_F = \left\{ f \in E_F : \sup_N \|\pi_N(f)\|_N < \infty \right\},$$

and, for all  $f \in \mathcal{H}_F$ , by

$$\|f\|_{\mathcal{H}_F}^2 = \lim_{N \rightarrow \infty} \|\pi_N(f)\|_N^2.$$

*Proof.* See [Bay et al., 2016, Theorem 3.1].  $\square$

In Proposition 3.3, and throughout the paper, recall that the convergence in  $E_F$  is defined with the uniform norm. We make the following assumption:

**Condition 3.2.**  $\pi_N(C_F) \subset C_F$ .

It can be shown that this assumption holds for boundedness, monotonicity and convexity constraints, using reasoning similar to [Bachoc et al., 2022]. Finally, we can state the existence and unicity of the approximate solution  $\hat{u}_{N,F}$ .

**Theorem 3.1.** Under Conditions 2.1 to 3.2, (11) has a unique solution  $\hat{u}_{N,F}$ .

*Proof.* From Condition 2.2, we can take  $g \in \mathcal{H}_F \cap C_F$ . Then, thanks to Condition 3.2,  $\pi_N(g) \in \mathcal{H}_{N,F} \cap C_F$ , so that  $\mathcal{H}_{N,F} \cap C_F$  is nonempty. From the same arguments as in Lemma 2.1, it is a closed convex subset of  $\mathcal{H}_{N,F}$ . Similarly as in the proof of Theorem 2.1, we have the conclusion.  $\square$

## 4 Quantitative properties of the set of approximants $\mathcal{H}_{N,F}$

In this section, we aim to quantitatively assess whether the class of approximants we have chosen is suitable. The first indicator is based on the error resulting from the finite-dimensional approximation:

$$F_N(f) = \|\pi_N(f) - f\|_\infty, \tag{12}$$

for  $f$  in  $E_F$  or  $\mathcal{H}_F$ .

As  $\mathcal{H}_{N,F}$  is intended to approximate  $\mathcal{H}_F$ , a second indicator evaluating the quality of the RKHS approximation plays a key role in the error bound. One way to achieve this is through their kernels:

$$G_N = \sup_{t \in F} \|\rho_N(K_N(\cdot, t)) - K(\cdot, t)\|_{\mathcal{H}_F}^2. \quad (13)$$

Here, since  $K$  and  $K_N$  belong to different spaces, we extend  $K_N$  using the operator  $\rho_N$ . Thus,  $\rho_N$  is the extension operator from the approximating space  $\mathcal{H}_{N,F}$  to the infinite-dimensional space  $\mathcal{H}_F$  defined as follows:

$$\forall v_N \in \mathcal{H}_{N,F}, \quad \rho_N(v_N) := \sum_{i=1}^N \lambda_i K(\cdot, t_i), \quad (14)$$

where  $\Lambda = (\lambda_1, \dots, \lambda_N)^\top$  solves  $\Gamma_N \Lambda = c_{v_N}$ , recalling the definition of  $c_{v_N}$  after (9). The vector  $\Lambda$  is defined such that the operator  $\rho_N$  is an isometry between  $\mathcal{H}_{N,F}$  and  $\mathcal{H}_F$ , i.e.,  $\forall v_N \in \mathcal{H}_{N,F}$ , we can check that

$$\|\rho_N(v_N)\|_{\mathcal{H}_F}^2 = \|v_N\|_N^2. \quad (15)$$

**Remark 4.1.** *It is possible to define  $\rho_N$  on  $E_F$ . In fact,  $\rho_N(f)$  is denoted by  $I_N(f)$  in [Karvonen, 2022], and is often called the kernel interpolant because it corresponds to the unique function in the span of  $K(\cdot, t_i)$  that interpolates  $f$  at the nodes  $t_i$ , where  $t_i \in S_N$ . The difference is that [Karvonen, 2022] considers interpolation at the observation points, whereas here, it is at the knots. While [Karvonen, 2022] aimed to investigate the accuracy of the kernel interpolant, it serves here as an intermediary that allows us to compare the exact solution with the approximate one.*

We now study both quality indicators in (12) and (13). For  $F_N$ , it is useful to give the simplest explicit formula to evaluate  $\pi_N(f)(t)$ , for  $t \in F$  and  $f \in E_F$ . For this, we define  $w_{N_-}(t)$  and  $w_{N_+}(t)$  as follows. Recall the definition of  $t_{S_N}^-$  and  $t_{S_N}^+$  in (7). If  $t \in S_N$ , then  $w_{N_-}(t) = w_{N_+}(t) = 1/2$ , otherwise

$$w_{N_-}(t) = \frac{t_{S_N}^+ - t}{t_{S_N}^+ - t_{S_N}^-}, \quad w_{N_+}(t) = \frac{t - t_{S_N}^-}{t_{S_N}^+ - t_{S_N}^-}.$$

This yields

$$\pi_N(f)(t) = f(t_{S_N}^-)w_{N_-}(t) + f(t_{S_N}^+)w_{N_+}(t). \quad (16)$$

We then have the following proposition.

**Proposition 4.1.** *Recall  $\Psi_f$  from (4) and  $\delta_N$  from (8). If  $f$  is in  $E$  or  $E_F$ , then*

$$F_N(f) \leq 2\Psi_f(\delta_N). \quad (17)$$

*Proof.* As  $w_{N_-}(t) + w_{N_+}(t) = 1$ , (16) implies

$$\pi_N(f)(t) - f(t) = (f(t_{S_N}^-) - f(t))w_{N_-}(t) + (f(t_{S_N}^+) - f(t))w_{N_+}(t).$$

Let us suppose without loss of generality that  $t - t_{S_N}^- \leq t_{S_N}^+ - t$ . As  $t - t_{S_N}^- \leq \delta_N$ ,

$$t_{S_N}^+ - t = \frac{(t_{S_N}^+ - t)(t - t_{S_N}^-)}{t - t_{S_N}^-} \leq \frac{(t_{S_N}^+ - t)}{t - t_{S_N}^-} \delta_N \leq x_N \delta_N, \quad (18)$$

where  $x_N = \frac{t_{S_N}^+ - t_{S_N}^-}{t - t_{S_N}^-} \geq 1$ , so that

$$\begin{aligned}
|\pi_N(f)(t) - f(t)| &\leq |f(t_{S_N}^-) - f(t)| \frac{t_{S_N}^+ - t}{t_{S_N}^+ - t_{S_N}^-} + |f(t_{S_N}^+) - f(t)| \frac{t - t_{S_N}^-}{t_{S_N}^+ - t_{S_N}^-} \\
&\leq |f(t_{S_N}^-) - f(t)| + |f(t_{S_N}^+) - f(t)| \frac{t - t_{S_N}^-}{t_{S_N}^+ - t_{S_N}^-} \\
&\leq M_f(\delta_N) + \frac{|f(t_{S_N}^+) - f(t)|}{\frac{t_{S_N}^+ - t_{S_N}^-}{t - t_{S_N}^-}} \quad (\text{recall } M_f(\delta) \text{ from (3)}) \\
&\leq M_f(\delta_N) + \frac{M_f(t_{S_N}^+ - t)}{\frac{t_{S_N}^+ - t_{S_N}^-}{t - t_{S_N}^-}} \\
&\leq M_f(\delta_N) + \frac{M_f(\delta_N x_N)}{x_N},
\end{aligned}$$

where we can conclude that  $|\pi_N(f)(t) - f(t)| \leq M_f(\delta_N) + \Psi_f(\delta_N) \leq 2\Psi_f(\delta_N)$ .  $\square$

For  $G_N$ , we have the following proposition.

**Proposition 4.2.** *We have*

$$G_N \xrightarrow[N \rightarrow \infty]{} 0. \quad (19)$$

Furthermore, if  $K$  satisfies Condition 2.1, then

$$G_N \leq d_2 \delta_N^\beta, \quad (20)$$

where  $d_2 := 6c_K$ .

*Proof.* We have  $\rho_N(K_N(\cdot, t)) = \sum_{i=1}^N \lambda_i(t) K(\cdot, t_i)$ , where  $\Gamma_N \Lambda(t) = c_{K_N(\cdot, t)}$ , and

$$K_N(\cdot, t) = \sum_{i=1}^N \left( \sum_{j=1}^N K(t_i, t_j) \phi_{j|F}(t) \right) \phi_{i|F}.$$

Hence  $\Gamma_N \Lambda(t) = \Gamma_N \phi(t)$ , where  $\phi(t) = (\phi_{1|F}(t), \dots, \phi_{N|F}(t))^\top$ . Then,  $\Lambda(t) = \phi(t)$  and

$$\rho_N(K_N(\cdot, t)) = \sum_{i=1}^N \phi_{i|F}(t) K(\cdot, t_i).$$

Applying the reproducing property of  $K$  and  $K_N$ , we obtain

$$\begin{aligned}
\langle K(\cdot, t_i), K(\cdot, t) \rangle_{\mathcal{H}_F} &= K(t, t_i), \\
\langle K_N(\cdot, t), K_N(\cdot, t) \rangle_N &= K_N(t, t).
\end{aligned}$$

As  $\rho_N$  is isometric (see (15)), we have

$$\begin{aligned}
\|\rho_N(K_N(\cdot, t)) - K(\cdot, t)\|_{\mathcal{H}_F}^2 &= \|\rho_N(K_N(\cdot, t))\|_{\mathcal{H}_F}^2 + \|K(\cdot, t)\|_{\mathcal{H}_F}^2 - 2\langle \rho_N(K_N(\cdot, t)), K(\cdot, t) \rangle_{\mathcal{H}_F} \\
&= \|K_N(\cdot, t)\|_N^2 + \|K(\cdot, t)\|_{\mathcal{H}_F}^2 - 2 \sum_{i=1}^N \phi_{i,F}(t) K(t, t_i) \\
&= K_N(t, t) + K(t, t) - 2 \sum_{i=1}^N \phi_{i,F}(t) K(t, t_i).
\end{aligned}$$

Setting  $K_t(\cdot) = K(\cdot, t)$ , if  $K$  satisfies Condition 2.1, then, according to Propositions 2.1 and 4.1,

$$\left| K_t(t) - \sum_{i=1}^N \phi_{i,F}(t) K_{t_i}(t) \right| = |K_t(t) - \pi_N(K_t(t))| \leq \|K_t - \pi_N(K_t)\|_\infty \leq 2c_K \delta_N^\beta.$$

We have, for  $t \in [t_{S_N}^-, t_{S_N}^+]$ ,

$$K_N(t, t) = K_{t_{S_N}^-}^-(t_{S_N}^-) w_{N_-}^2(t) + K_{t_{S_N}^+}^+(t_{S_N}^+) w_{N_+}^2(t) + 2K_{t_{S_N}^-}^-(t_{S_N}^+) w_{N_+}(t) w_{N_-}(t).$$

Also

$$\begin{aligned} K_{t_{S_N}^-}^-(t) w_{N_-}(t) &= K_{t_{S_N}^-}^-(t) w_{N_-}(t) [w_{N_-}(t) + w_{N_+}(t)], \\ K_{t_{S_N}^+}^+(t) w_{N_+}(t) &= K_{t_{S_N}^+}^+(t) w_{N_+}(t) [w_{N_-}(t) + w_{N_+}(t)]. \end{aligned}$$

Hence,

$$\begin{aligned} \left| K_N(t, t) - \sum_{i=1}^N \phi_{i,F}(t) K_{t_i}(t) \right| &= \left| K_N(t, t) - K_{t_{S_N}^-}^-(t) w_{N_-}(t) - K_{t_{S_N}^+}^+(t) w_{N_+}(t) \right| \\ &\leq \left| K_{t_{S_N}^-}^-(t_{S_N}^-) - K_{t_{S_N}^-}^-(t) \right| w_{N_-}^2(t) + \left| K_{t_{S_N}^+}^+(t_{S_N}^+) - K_{t_{S_N}^+}^+(t) \right| w_{N_+}^2(t) \\ &\quad + \left[ \left| K_{t_{S_N}^-}^-(t_{S_N}^+) - K_{t_{S_N}^-}^-(t) \right| + \left| K_{t_{S_N}^+}^+(t_{S_N}^-) - K_{t_{S_N}^+}^+(t) \right| \right] w_{N_+}(t) w_{N_-}(t). \end{aligned}$$

Let us now suppose, without loss of generality, that  $t - t_{S_N}^- \leq t_{S_N}^+ - t$ . As  $K_{t_{S_N}^+}^+$  and  $K_{t_{S_N}^-}^-$  are  $\beta$ -Hölder continuous, we have

$$\left| K_{t_{S_N}^-}^-(t_{S_N}^-) - K_{t_{S_N}^-}^-(t) \right| w_{N_-}^2(t) \leq c_K \delta_N^\beta.$$

Using (18), as  $t_{S_N}^+ - t \leq x_N \delta_N$ , where  $x_N = \frac{t_{S_N}^+ - t_{S_N}^-}{t - t_{S_N}^-} \geq 1$ , we obtain

$$\begin{aligned} \left| K_{t_{S_N}^+}^+(t_{S_N}^+) - K_{t_{S_N}^+}^+(t) \right| w_{N_+}^2(t) &\leq M_{K_{t_{S_N}^+}^+} (t_{S_N}^+ - t) \left( \frac{t - t_{S_N}^-}{t_{S_N}^+ - t_{S_N}^-} \right)^2 \\ &\leq \frac{M_{K_{t_{S_N}^+}^+} (\delta_N x_N)}{x_N^2} \\ &\leq \frac{M_{K_{t_{S_N}^+}^+} (\delta_N x_N)}{x_N} \\ &\leq \psi_{K_{t_{S_N}^+}^+} (\delta_N) \leq c_K \delta_N^\beta, \end{aligned}$$

$$\begin{aligned} \left| K_{t_{S_N}^-}^-(t_{S_N}^+) - K_{t_{S_N}^-}^-(t) \right| w_{N_+}(t) w_{N_-}(t) &\leq M_{K_{t_{S_N}^-}^-} (t_{S_N}^+ - t) \left( \frac{t - t_{S_N}^-}{t_{S_N}^+ - t_{S_N}^-} \right) \\ &\leq \frac{M_{K_{t_{S_N}^-}^-} (\delta_N x_N)}{x_N} \leq c_K \delta_N^\beta, \end{aligned}$$

$$\left| K_{t_{S_N}^+}^+(t_{S_N}^-) - K_{t_{S_N}^+}^+(t) \right| w_{N_+}(t) w_{N_-}(t) \leq c_K \delta_N^\beta.$$

Finally,

$$\left| K_N(t, t) - \sum_{i=1}^N \phi_{i,F}(t) K_{t_i}(t) \right| \leq 4c_K \delta_N^\beta,$$

so that  $\|\rho_N(K_N(\cdot, t)) - K(\cdot, t)\|_{\mathcal{H}_F}^2 \leq 6c_K \delta_N^\beta$ , which allows to conclude the proof of (20), under Condition 2.1. The proof of (19) uses the same arguments and that  $\lim_{\delta \rightarrow 0} \Psi_f(\delta) = 0$  from Proposition 2.1.  $\square$

## 5 Error Bound

We here analyze the error committed when approximating  $\widehat{u}_F$ , the solution of (5), by  $\widehat{u}_{N,F}$ , the solution of (11). We not only prove that  $\widehat{u}_{N,F} \rightarrow \widehat{u}_F$  when  $N \rightarrow \infty$ , but also give an error bound. To estimate the error  $\|\widehat{u}_{N,F} - \widehat{u}_F\|_\infty$ , we split it in two terms: the piecewise linear interpolation error  $\|\pi_N(\widehat{u}_F) - \widehat{u}_F\|_\infty$ , and the distance from the approximate solution to the projection of the exact solution, both on  $\mathcal{H}_{N,F}$ ,  $\|\pi_N(\widehat{u}_F) - \widehat{u}_{N,F}\|_\infty$ . This splitting is treated rigorously in the following proposition.

**Proposition 5.1.** *Under Conditions 2.1 to 3.2, there is a constant  $c$  such that*

$$\|\widehat{u}_{N,F} - \widehat{u}_F\|_\infty \leq c \|\pi_N(\widehat{u}_F) - \widehat{u}_{N,F}\|_N + d_1 \delta_N^{\beta/2}, \quad (21)$$

where  $d_1 := \sqrt{8c_K} \|\widehat{u}_F\|_{\mathcal{H}_F}$ , with  $c_K$  and  $\beta$  from Condition 2.1.

*Proof.* We have

$$\|\widehat{u}_{N,F} - \widehat{u}_F\|_\infty \leq \|\widehat{u}_{N,F} - \pi_N(\widehat{u}_F)\|_\infty + \|\pi_N(\widehat{u}_F) - \widehat{u}_F\|_\infty.$$

As  $\|\pi_N(\widehat{u}_F) - \widehat{u}_F\|_\infty = F_N(\widehat{u}_F)$ , according to the bound in (17),  $F_N(\widehat{u}_F) \leq 2\psi_{\widehat{u}_F}(\delta_N)$ . Theorem 2.1 states that  $\widehat{u}_F$  is  $\frac{\beta}{2}$ -Hölder continuous, so that applying Proposition 2.1, we obtain

$$\|\pi_N(\widehat{u}_F) - \widehat{u}_F\|_\infty \leq 2\sqrt{2c_K} \|\widehat{u}_F\|_{\mathcal{H}_F} \delta_N^{\beta/2} = d_1 \delta_N^{\beta/2}. \quad (22)$$

As  $\mathcal{H}_{N,F}$  is an Hilbertian space of  $E_F$ , from Lemma 5.1 in [Grammont et al., 2024], there exists a constant  $c$  such that,  $\forall h_N \in \mathcal{H}_{N,F}$ ,

$$\|h_N\|_\infty \leq c \|h_N\|_N. \quad (23)$$

□

Similarly as in (23) and up to increasing  $c$ , for all  $h \in \mathcal{H}_F$ ,

$$\|h\|_\infty \leq c \|h\|_{\mathcal{H}_F}. \quad (24)$$

It remains to address the second term  $\|\pi_N(\widehat{u}_F) - \widehat{u}_{N,F}\|_N$ , that requires a more delicate treatment, provided in the next propositions. In the following, for the sake of readability, some of the proofs will be presented in Appendix A.

We first show that the bound of  $\|\pi_N(\widehat{u}_F) - \widehat{u}_{N,F}\|_N$  relies on the characterization of the strong convexity for a differentiable function and the necessary condition of the first order for its minimum.

**Proposition 5.2.** *Under Conditions 2.1 to 3.2, we have*

$$\|\pi_N(\widehat{u}_F) - \widehat{u}_{N,F}\|_N^2 \leq J_{N,F}(\pi_N(\widehat{u}_F)) - J_{N,F}(\widehat{u}_{N,F}),$$

with  $J_{N,F}$  as in (10).

*Proof.* As  $J_{N,F}$  is differentiable from Proposition 3.2 with Fréchet derivative  $J'_{N,F}$ , then the strong convexity leads to,  $\forall v, u \in \mathcal{H}_{N,F}$ ,

$$J_{N,F}(v) - J_{N,F}(u) \geq \langle J'_{N,F}(u), v - u \rangle_N + \|v - u\|_N^2.$$

If  $u = \widehat{u}_{N,F}$  and  $v = \pi_N(\widehat{u}_F)$ , then

$$J_{N,F}(\pi_N(\widehat{u}_F)) - J_{N,F}(\widehat{u}_{N,F}) \geq \langle J'_{N,F}(\widehat{u}_{N,F}), \pi_N(\widehat{u}_F) - \widehat{u}_{N,F} \rangle_N + \|\pi_N(\widehat{u}_F) - \widehat{u}_{N,F}\|_N^2.$$

As  $\pi_N(\widehat{u}_F) \in \mathcal{H}_{N,F} \cap C_F$  thanks to Condition 3.2, and  $\widehat{u}_{N,F}$  solves (11), we have

$$\langle J'_{N,F}(\widehat{u}_{N,F}), \pi_N(\widehat{u}_F) - \widehat{u}_{N,F} \rangle_N \geq 0,$$

which allows to conclude the proof. □

The next proposition derives the bounds of  $J_{N,F}(\pi_N(\hat{u}_F))$  and  $J_{N,F}(\hat{u}_{N,F})$ .

**Proposition 5.3.** *Under Conditions 2.1 to 3.2, we have*

$$J_{N,F}(\pi_N(\hat{u}_F)) = -\hat{E}_N + J_F(\hat{u}_F) + \epsilon_N, \quad (25)$$

$$J_{N,F}(\hat{u}_{N,F}) = J_F(\rho_N(\hat{u}_{N,F})) + \eta_N, \quad (26)$$

where

$$\begin{aligned} \hat{E}_N &= \|\hat{u}_F\|_{\mathcal{H}_F}^2 - \|\pi_N(\hat{u}_F)\|_N^2, \\ |\epsilon_N| &\leq d_3 \delta_N^{\beta/2}, \quad d_3 = \frac{2nd_1}{\tau} \left( c \|\hat{u}_F\|_{\mathcal{H}_F} + \max_i |y_i| \right), \\ |\eta_N| &\leq d_4 \delta_N^{\beta/2}, \quad d_4 = \frac{2n\sqrt{d_2}}{\tau} \|\hat{u}_{N,F}\|_N \left( c \|\hat{u}_{N,F}\|_N + \max_i |y_i| \right). \end{aligned}$$

*Proof.* This proposition has no standalone value but it serves as an intermediary result required for subsequent derivations. Its proof relies solely on computations and the application of previously established results. The complete proof is provided in Appendix A.  $\square$

Now, from Propositions 5.2 and 5.3, we can establish the error bound of  $\|\pi_N(\hat{u}_F) - \hat{u}_{N,F}\|_N$ .

**Proposition 5.4.** *Under Conditions 2.1 to 3.2*

$$\|\pi_N(\hat{u}_F) - \hat{u}_{N,F}\|_N^2 \leq J_F(\hat{u}_F) - J_F(\rho_N(\hat{u}_{N,F})) + d_5 \delta_N^{\beta/2}, \quad (27)$$

where  $d_5 = d_3 + d_4$ .

*Proof.* From Proposition 3.3,  $\hat{E}_N \geq 0$ . Using Propositions 5.2 and 5.3,

$$\begin{aligned} \|\pi_N(\hat{u}_F) - \hat{u}_{N,F}\|_N^2 &\leq J_{N,F}(\pi_N(\hat{u}_F)) - J_{N,F}(\hat{u}_{N,F}) \\ &= -\hat{E}_N + J_F(\hat{u}_F) + \epsilon_N - J_F(\rho_N(\hat{u}_{N,F})) - \eta_N \\ &\leq J_F(\hat{u}_F) - J_F(\rho_N(\hat{u}_{N,F})) + d_3 \delta_N^{\beta/2} + d_4 \delta_N^{\beta/2}. \end{aligned}$$

Since  $d_5 = d_3 + d_4$ , we have the bound in (27).  $\square$

To complete the construction of an error bound that is easy to read and interpret, we need to add a third quantity to the first two ( $\beta$ , which measures regularity, and  $\delta_N$ , which measures the grid size). This final quantity required for the error bound is the distance, in  $\mathcal{H}_F$ , between  $\rho_N(\hat{u}_{N,F})$  and the constraints  $C_F$ . This is the distance of the kernel interpolant of  $\hat{u}_{N,F}$  to the set of constraints:

$$\alpha_N := d(\rho_N(\hat{u}_{N,F}), C_F) = \|P_C(\rho_N(\hat{u}_{N,F})) - \rho_N(\hat{u}_{N,F})\|_{\mathcal{H}_F}, \quad (28)$$

where  $P_C$  in  $\mathcal{H}_F$  is the orthogonal projection in  $\mathcal{H}_F$  onto the closed convex set  $C_F$ .

We first address the case where  $\rho_N(\hat{u}_{N,F}) \in C_F$ , i.e.  $\alpha_N = 0$ , which corresponds to Condition 5.1 next.

**Condition 5.1.** *For  $N$  large enough,  $\rho_N(\hat{u}_{N,F}) \in C_F$ .*

**Theorem 5.1.** *Under Conditions 2.1 to 5.1,*

$$\|\hat{u}_{N,F} - \hat{u}_F\|_\infty = \mathcal{O}(\delta_N^{\beta/4}). \quad (29)$$

*Proof.* From Proposition 5.1,  $\|\hat{u}_{N,F} - \hat{u}_F\|_\infty \leq c \|\pi_N(\hat{u}_F) - \hat{u}_{N,F}\|_N + d_1 \delta_N^{\beta/2}$ . If Condition 5.1 is verified, then  $\rho_N(\hat{u}_{N,F}) \in C_F$  so that  $J_F(\hat{u}_F) - J_F(\rho_N(\hat{u}_{N,F})) \leq 0$ . Hence, from Proposition 5.4,

$$\|\pi_N(\hat{u}_F) - \hat{u}_{N,F}\|_N^2 \leq d_5 \delta_N^{\beta/2},$$

so that

$$\|\hat{u}_{N,F} - \hat{u}_F\|_\infty \leq c \sqrt{d_5 \delta_N^{\beta/2}} + d_1 \delta_N^{\beta/2}.$$

$\square$

Intuitively, Condition 5.1 is expected to hold when  $\widehat{u}_{N,F}$  is significantly “inside” the constraint set, since  $\rho_N(\widehat{u}_{N,F})$  is expected to be close to  $\widehat{u}_{N,F}$  for large  $N$ . Nevertheless, when  $\widehat{u}_{N,F}$  is close to the boundary of the constraint set, then its kernel interpolant  $\rho_N(\widehat{u}_{N,F})$  could fall outside of this set. To give a very simple example, when the constraint set imposes functions to take values in  $[0, 1]$ , it is possible that the values of a function at the knots are in  $[0, 1]$  but very close to 0 or 1, so that its kernel interpolant function takes some values outside of  $[0, 1]$ , for some kernels  $K$ .

Hence, it is valuable to analyze the case where Condition 5.1 does not hold (i.e.  $\alpha_N \neq 0$ ). In this case, the convergence proof of the error bound becomes more challenging and relies on some weak compactness arguments as we show in Theorem 5.2 next.

**Theorem 5.2.** *Under Conditions 2.1 to 3.2,*

$$\widehat{u}_{N,F} \xrightarrow{N \rightarrow \infty} \widehat{u}_F \quad \text{in } E_F. \quad (30)$$

Moreover, with  $c$  as in Proposition 5.1,

$$\|\widehat{u}_{N,F} - \widehat{u}_F\|_\infty \leq c \sqrt{d_8 \alpha_N + d_5 \delta_N^{\beta/2} + d_1 \delta_N^{\beta/2}}, \quad (31)$$

where  $d_8$  is a constant which is defined in the proof.

*Proof.* The convergence in (30) is already proved in [Grammont et al., 2024]. The proof is also provided in Appendix A for a self-contained reading.

Let us now construct the error estimation involving  $\delta_N$  and  $\alpha_N$  (see definitions in (8) and (28), respectively). Setting  $\widehat{h}^N := \rho_N(\widehat{u}_{N,F})$ , we have

$$J_F(\widehat{u}_F) - J_F(\rho_N(\widehat{u}_{N,F})) = J_F(\widehat{u}_F) - J_F(P_C(\widehat{h}^N)) + J_F(P_C(\widehat{h}^N)) - J_F(\widehat{h}^N).$$

As  $P_C(\widehat{h}^N) \in \mathcal{H}_F \cap C_F$ , then  $J_F(\widehat{u}_F) - J_F(P_C(\widehat{h}^N)) \leq 0$ . This implies that the quantity  $J_F(\widehat{u}_F) - J_F(\rho_N(\widehat{u}_{N,F}))$  is bounded by  $J_F(P_C(\widehat{h}^N)) - J_F(\widehat{h}^N)$ , which is equal to

$$\|P_C(\widehat{h}^N)\|_{\mathcal{H}_F}^2 - \|\widehat{h}^N\|_{\mathcal{H}_F}^2 + \frac{1}{\tau} \sum_{i=1}^n \left( P(P_C(\widehat{h}^N))(x_i) - P(\widehat{h}^N)(x_i) \right) \left( P(P_C(\widehat{h}^N))(x_i) + P(\widehat{h}^N)(x_i) - 2y_i \right).$$

Recall that  $P(u) \in E$  the multi-affine extension of  $u$  (see Definition 2.2).

In the proof of (30) in [Grammont et al., 2024, Appendix, Proof of Proposition 5.8], it is shown that  $(\widehat{h}^N)_N$  is bounded in  $\mathcal{H}_F$ . As  $P_C$  is 1-Lipschitz, then

$$\|P_C(\widehat{h}^N) - P_C(\widehat{u}_F)\|_{\mathcal{H}_F} \leq \|\widehat{h}^N - \widehat{u}_F\|_{\mathcal{H}_F},$$

and the sequence  $(P_C(\widehat{h}^N))_N$  is also bounded in  $\mathcal{H}_F$ . According to (24), these two sequences are also bounded in  $E_F$ . This implies that, for all  $i \in \{1, \dots, n\}$ ,  $|P(P_C(\widehat{h}^N))(x_i) + P(\widehat{h}^N)(x_i) - 2y_i|$  is bounded (we let  $d_6$  be the bound), and so is  $\|P_C(\widehat{h}^N)\|_{\mathcal{H}_F} + \|\widehat{h}^N\|_{\mathcal{H}_F}$  (we let  $d_7$  be the bound). Moreover,

$$\begin{aligned} \left| P(P_C(\widehat{h}^N))(x_i) - P(\widehat{h}^N)(x_i) \right| &\leq \|P_C(\widehat{h}^N) - \widehat{h}^N\|_\infty \\ &\leq c \|P_C(\widehat{h}^N) - \widehat{h}^N\|_{\mathcal{H}_F} \\ &= c \alpha_N, \\ \|P_C(\widehat{h}^N)\|_{\mathcal{H}_F}^2 - \|\widehat{h}^N\|_{\mathcal{H}_F}^2 &= (\|P_C(\widehat{h}^N)\|_{\mathcal{H}_F} + \|\widehat{h}^N\|_{\mathcal{H}_F})(\|P_C(\widehat{h}^N)\|_{\mathcal{H}_F} - \|\widehat{h}^N\|_{\mathcal{H}_F}) \\ &\leq d_7 \alpha_N. \end{aligned}$$

Hence,

$$J_F(\widehat{u}_F) - J_F(\rho_N(\widehat{u}_{N,F})) \leq \left( d_7 + \frac{ncd_6}{\tau} \right) \alpha_N = d_8 \alpha_N.$$

Using the results in (21) and (27) from Propositions 5.1 and 5.4, we obtain the error bound in (31).  $\square$

**Remark 5.1.** *With a similar approach as provided here, it is possible to provide error bounds on the numerical approximation of the constrained smoothing problem in higher dimensions. In particular, the multi-affine extension is defined for general dimensions in Bachoc et al. [2022], and all its properties, including those related to the constraint sets of bounded, monotonic, and componentwise convex functions, are also established for general dimensions. Nevertheless, presenting detailed proofs, as we do here, in general dimension yields significantly more complex notations and cumbersome arguments. To maintain readability, we present our results and proofs in one dimension.*

**Remark 5.2.** *Our error bounds depend on the regularity parameter  $\beta \in (0, 1]$  for the kernel  $K$  (Condition 2.1), using the notion of Hölder-continuity, and with the rate  $\mathcal{O}(\delta_N^{\beta/4})$  in Theorem 5.1. It is natural to ask whether faster decay rates of the upper bounds could be achieved with stronger regularity assumptions, particularly by assuming derivatives of multiple orders. However, it is unclear if this additional regularity would be beneficial in our setting. This is because we rely on piecewise linear interpolation, which typically does not gain further benefits from regularity beyond Lipschitzness. As explained in [Bachoc et al., 2022], piecewise linear interpolation is crucial for numerically handling standard constraint sets (boundedness, monotonicity and convexity). Using an interpolation scheme that would benefit from regularity beyond Lipschitzness, for instance piecewise polynomial interpolation, is not suitable for numerically handling these constraint sets.*

**Remark 5.3.** *The distance of the kernel interpolant of  $\hat{u}_{N,F}$  from the set of constraints  $C_F$ , denoted as  $\alpha_N$  (see definition in (28)), tends to zero as  $N \rightarrow \infty$ . Indeed, recall the definition  $\hat{h}^N = \rho_N(\hat{u}_{N,F})$  from the proof of Theorem 5.2. As  $P_C(\hat{u}_F) = \hat{u}_F$ , we have*

$$\alpha_N \leq \|P_C(\hat{h}^N) - P_C(\hat{u}_F)\|_{\mathcal{H}_F} + \|\hat{u}_F - \hat{h}^N\|_{\mathcal{H}_F} \leq 2\|\hat{u}_F - \hat{h}^N\|_{\mathcal{H}_F}.$$

*In the proof of (30) in Theorem 5.2, it is proved that a subsequence of  $\hat{h}^N$  is weakly convergent to  $\hat{u}_F$  in  $\mathcal{H}_F$ . This implies that  $\hat{h}^{N_k}(t) \rightarrow \hat{u}_F(t)$  for all  $t \in F$ . By construction of the multi-affine extension  $P$  in Definition 2.2, it is also true for all  $t \in [0, 1]$ . Thus,*

$$\frac{1}{\tau} \sum_{i=1}^n \left( P(\hat{h}^{N_k})(x_i) - y_i \right)^2 \rightarrow \frac{1}{\tau} \sum_{i=1}^n \left( P(\hat{u}_F)(x_i) - y_i \right)^2.$$

*As  $J_F(\hat{h}^{N_k}) \rightarrow J_F(\hat{u}_F)$ , we have*

$$\|\hat{h}^{N_k}\|_{\mathcal{H}_F} \rightarrow \|\hat{u}_F\|_{\mathcal{H}_F}.$$

*This property, combined with the weak convergence, leads to the convergence of  $\hat{h}^{N_k}$  to  $\hat{u}_F$  in  $\mathcal{H}_F$ . As  $\hat{u}_F$  is the unique accumulation point, we have*

$$\|\hat{u}_F - \hat{h}^N\|_{\mathcal{H}_F} \rightarrow 0.$$

## 6 Numerical Experiments

In this section we aim to numerically illustrate Theorems 5.1 and 5.2. Let us recall that the approximate solution  $\hat{u}_{N,F}$  is also the MAP estimate of a GP approximation conditionally to noisy observations and the inequality constraints [Grammont et al., 2024]. Our numerical assessment relies on this property. Therefore, we consider constrained GPs with stationary Matérn kernels [Genton, 2001]:

$$K(x, x') = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \sqrt{2\nu} \frac{|x - x'|}{\ell} \right)^\nu H_\nu \left( \sqrt{2\nu} \frac{|x - x'|}{\ell} \right), \quad (32)$$

where  $x, x' \in [0, 1]$ ,  $\Gamma$  is the Gamma function,  $H_\nu$  is the modified Bessel function of the second kind of order  $\nu$ , and  $(\sigma^2, \ell) \in (0, \infty)^2$  are the variance and length-scale parameters, respectively. The parameter  $\nu \in (0, \infty)$  allows controlling the regularity of the GP. The larger  $\nu$ , the smoother the GP samples.

Given the settings above, we sample twenty constrained GP replicates using the finite-dimensional approximation in [López-Lopera et al., 2018] assuming an equispaced grid of knots with  $N = 200$ . The



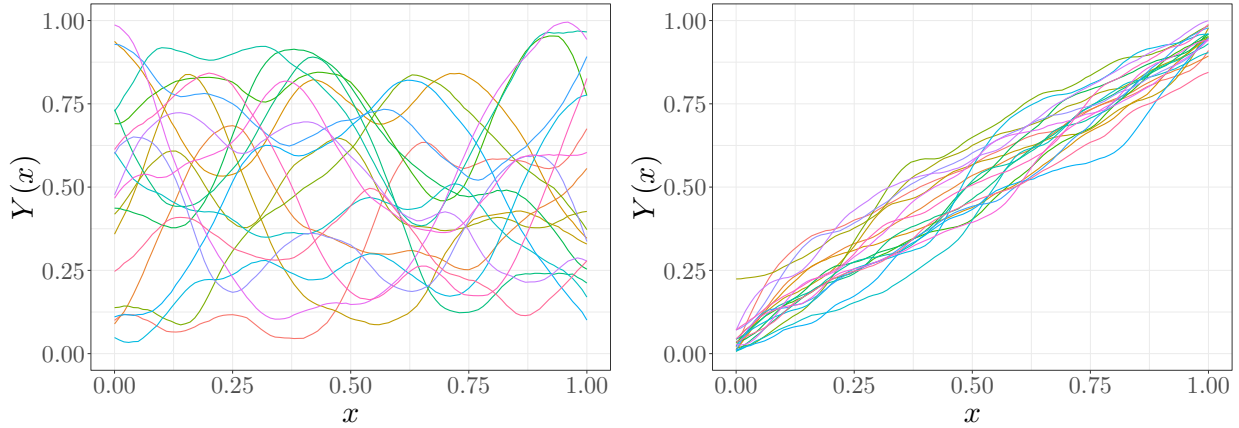


Figure 1: Random GP replicates under boundedness (left) and both boundedness and monotonicity constraints (right) used in the experiments in Section 6.1. As boundedness constraints, we consider  $0 \leq Y(x) \leq 1$ , for all  $x \in [0, 1]$ .

choice of  $N$  balances the need for better resolution of the piecewise approximation while considering the computational limitations inherent in Monte Carlo techniques. To introduce noise, we corrupt the samples by using independent centred Gaussian noises with noise variance  $\tau = 5 \times 10^{-2}$ .

We compute the MAP estimate  $\hat{u}_{N,F}$  for each random noisy replicate. This procedure results in twenty predictors that will be used to illustrate Theorems 5.1 and 5.2. For the predictor  $\hat{u}_F$ , which cannot be evaluated in practice, we approximate it using  $\hat{u}_{N,F}$  assuming an equispaced grid of knots with  $N = 10^3$ . As parameter estimation is not the focus here, we use the same covariance parameters and noise variance that have been set to generate the constrained GP replicates.

In the first part of our experiments (Section 6.1), we focus on the cases where the grid of the knots is either dense or not. Thus, we fix  $\nu = 5/2$  to have the same regularity conditions. In the second part (Section 6.2), we vary  $\nu$  seeking to test convergence for different values of  $\beta$  while keeping promoting a dense grid of knots. In our context, the link between  $\nu$  and  $\beta$  is given by  $\beta = \min(1, 2\nu)$  (see for instance Loh [2015]).

For dense grids, we conduct the MaxMod algorithm introduced by [Bachoc et al., 2022] and discussed in Section 1, with a minimal initial number of knots (i.e.  $N_0 = 2$ ), and a maximal budget  $N_{\max} = 250$ . This budget has been set aiming for a trade-off between computational time and numerical stability due to inversion of covariance matrices. Using the MaxMod algorithm will allow to verify convergence of the error bounds without involving equispaced grid of knots necessarily.

The implementation of the constrained GPs and the MaxMod algorithm are based on the R package `lineqGPR` [López-Lopera, 2022].

## 6.1 Error bounds with fixed regularity assumptions

In this experiment, we sample random GP replicates under monotonicity and boundedness constraints (see Figure 1). We impose the boundedness constraint  $0 \leq Y(x) \leq 1$ , for all  $x \in [0, 1]$ , and use a Matérn 5/2 kernel with  $\sigma^2 = 1$  and  $\ell = 0.4$ . We consider two cases where the grid of knots is dense or not. For dense grids, we apply the MaxMod algorithm to each random replicate. For non-dense grids, we restrict the addition of knots to the interval  $I = [0, 0.3] \cup [0.6, 1]$ . The refinement process is then conducted via rejection sampling with  $t \sim \text{Uniform}(0, 1)$ . In both cases, we set  $N_{\max} = 250$ .

Figure 2 presents boxplots of the error  $\|\hat{u}_{N,F} - \hat{u}_F\|_\infty$  and the grid size  $\delta_N$  (defined in (8)) for the twenty replicates. We must remark that the asymptotic error bounds of  $\|\hat{u}_{N,F} - \hat{u}_F\|_\infty$  cannot be displayed, as  $\alpha_N$  (defined in (28)) cannot be computed numerically. We observe that the error decreases as  $\delta_N$  decreases, which is consistent with Theorems 5.1 and 5.2 as the asymptotic error there bounds become smaller as  $\delta_N$  decreases. The boxplots show median error values smaller than  $10^{-3}$  once the maximal budget  $N_{\max}$  is reached, except for the example under boundedness constraints with non-dense grids of knots. This increase in error is due to abrupt changes in the MAP around

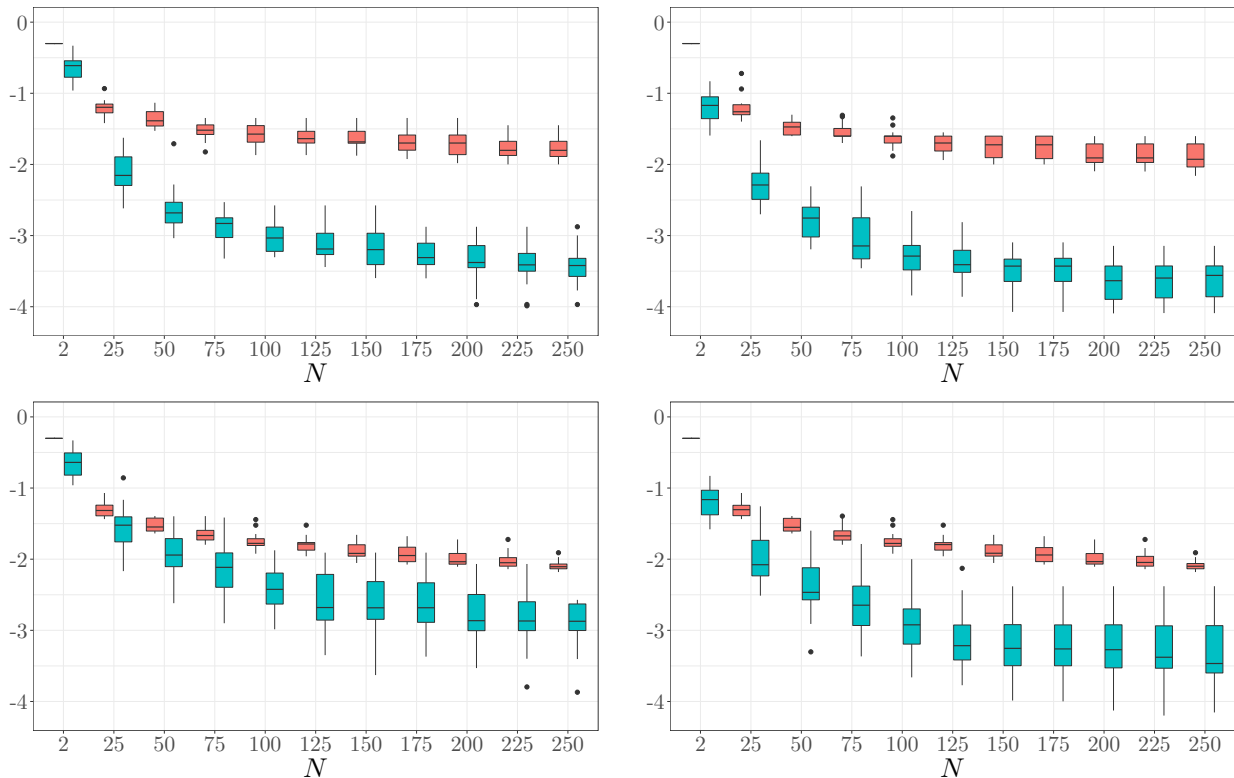


Figure 2: Boxplots of the error  $\log_{10} \|\widehat{u}_{N,F} - \widehat{u}_F\|_\infty$  (blue) and the grid size  $\delta_N$  (red) as a function of the number of knots  $N$ . The boxplots are computed for the twenty replicates in Figure 1, considering cases when the grid of knots is dense or not (top and bottom, respectively). Results are presented for examples under boundedness (left) or both boundedness and monotonicity (right) constraints.

$x = 0.3$  and  $x = 0.6$  (limits of the rejection interval). To achieve smaller error values, it is possible to repeat the experiments with a larger  $N_{\max}$  expecting adding knots close to the limits. In particular for this example, a median error value smaller than  $10^{-3}$  is achieved after  $N = 275$ .

## 6.2 Error bounds with different regularity assumptions

We now perform experiments with dense grids while varying  $\beta$ , the parameter related to the regularity of the kernel. To do so, we consider  $\nu = 1/4, 3/8, 1/2, 3/4, 5/2$  (respectively,  $\beta = 1/2, 3/4, 1, 1, 1$ ). Although  $\beta = 1$  for  $\nu = 1/2, 3/4, 5/2$ , we opt to conduct experiments with these values to observe the impact of the GP samples' smoothness order on the rate of error decrease. We follow the same procedure as described in Section 6.1 with the same GP parameters except for the length-parameter that we have increased to  $\ell = 0.8$  to control the variability of the samples (see Figure 3, left panels). This choice seeks to have visible convergence trends for  $N \leq N_{\max} = 250$ . As monotonicity is unlikely to be satisfied for Matérn kernels with  $\nu \leq 1/2$ , we focus here on boundedness constraints. To assess if narrower bounds may have an impact in the results, we suggest GP replicates satisfying  $0 \leq Y(x) \leq 0.5$ .

In line with Figure 2, Figure 3 shows a decreasing trend of the error  $\|\widehat{u}_{N,F} - \widehat{u}_F\|_\infty$  as  $\delta_N$  decreases independently of the value of  $\beta$ . In particular, we note that the error decreases faster as  $\beta$  increases, which is consistent with Theorems 5.1 and 5.2 as the asymptotic error bounds there become smaller as  $\beta$  increases. In addition, the rate of decrease is higher for smoother GP samples (i.e. for larger  $\nu$  values). This pattern has also been verified for the squared exponential kernel, i.e. when  $\nu \rightarrow \infty$ , in experiments unreported here.

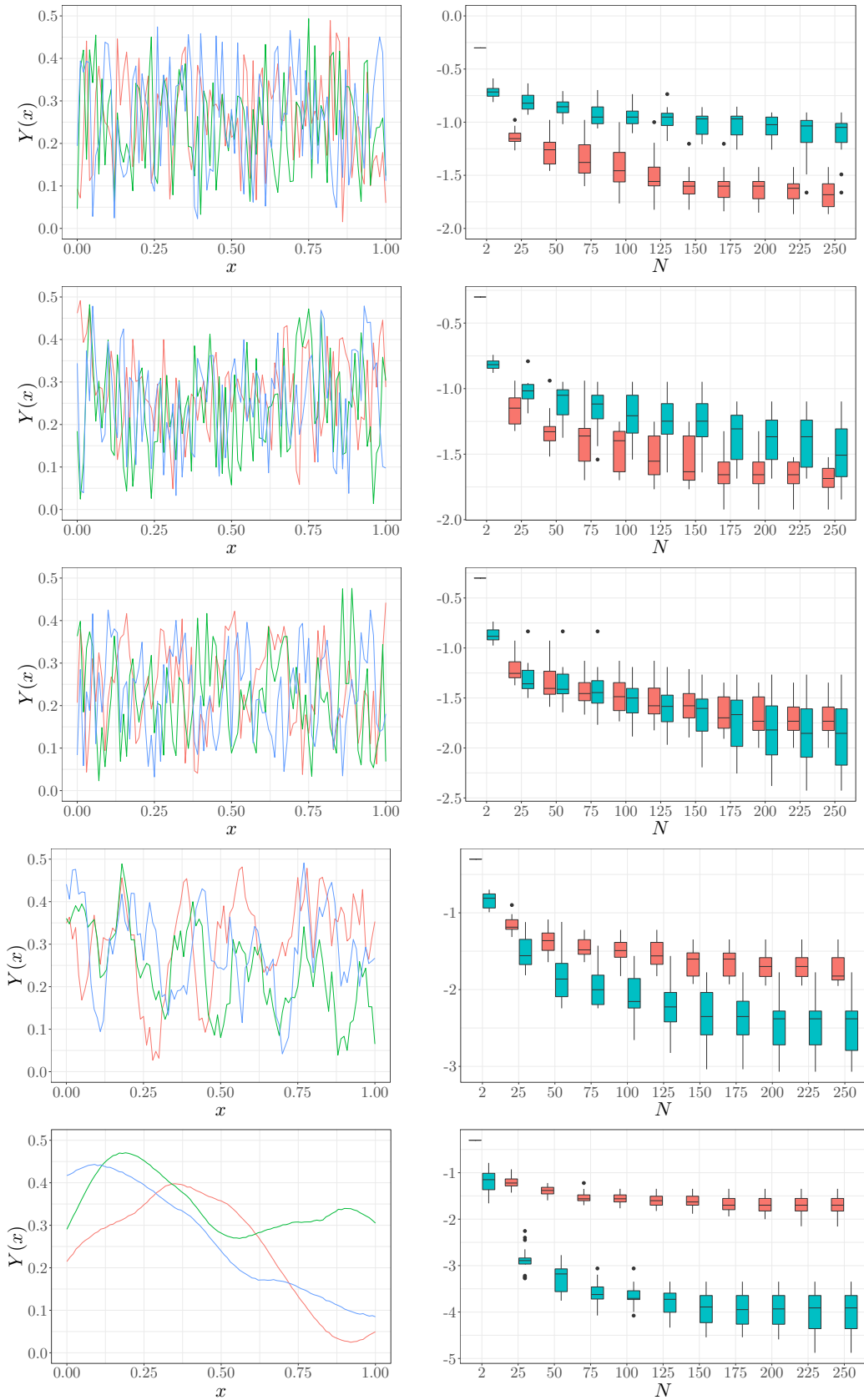


Figure 3: Right: boxplots of  $\log_{10} \|\hat{u}_{N,F} - \hat{u}_F\|_\infty$  (blue) and  $\delta_N$  (red) as a function of  $N$ . The boxplots are computed for twenty constrained GP replicates using a Matérn kernel with  $\nu = 1/4, 3/8, 1/2, 3/4, 5/2$  (from top to bottom). For a better visualization, we only display three of the twenty replicates (left).

## 7 Conclusions

We have introduced a general error bound (see Theorem 5.2) for the constrained optimal smoothing problem and for the equivalent formulation with the MAP estimator. We show that this bound depends on the grid size, the regularity of the kernel, and the distance from the kernel interpolant of the approximation to the set of constraints. In particular, if the kernel interpolant satisfies the constraints, then the latter distance is zero, and the error bound is given by  $\mathcal{O}(\delta_N^{\beta/4})$  (see Theorem 5.1). Convergence results are provided for non-equispaced and non-dense grids of knots, allowing certification of sequential schemes, such as the MaxMod algorithm in [Bachoc et al., 2022], introduced for the efficient allocation of knots. To the best of our knowledge, our theoretical results are the first to provide quantitative error bounds for numerical approximations of constrained GPs.

Our results are illustrated numerically through various synthetic examples that account for different types of inequality constraints (e.g., boundedness and monotonicity) and regularity assumptions (i.e. smoothness of the GP samples). Additionally, we examine scenarios with both dense and non-dense grids of knots. Our experiments show that the approximation error decreases as the grid size decreases, with a faster rate of decrease observed for smoother GP samples. This is in line with our theoretical analysis.

**Acknowledgments.** We thank Olivier Roustant (IMT, France) for his contribution to this work. Indeed, this study began thanks to his interest about the subject. Scientific exchanges with him enabled the improvement of the quality of this work. This work was supported by the projects GAP (ANR-21-CE40-0007) and GAME (ANR-23-CE46-0007) of the French National Research Agency (ANR).

## A Remaining proofs

In order to make the paper self-contained, we provide in this appendix the proofs that are primarily technical or containing pre-existing concepts from other works.

*Proof of Proposition 2.1.* Let  $f$  be continuous on  $S$ . Suppose that the property  $\lim_{\delta \rightarrow 0} \Psi_f(\delta) = 0$  is not true. Hence, there exists a sequence  $(\delta_n)_{n \in \mathbb{N}}$  such that  $\delta_n \rightarrow 0$  and a constant  $A > 0$  such that  $\Psi_f(\delta_n) > A$  for all  $n \in \mathbb{N}$ . Moreover, there exists a sequence  $(t_n)_{n \in \mathbb{N}}$  with  $t_n > 1$  such that

$$\frac{M_f(t_n \delta_n)}{t_n} > A. \quad (33)$$

If  $(t_n)_{n \in \mathbb{N}}$  has a bounded subsequence, then  $1 \leq t_{\phi(n)} \leq b$  and  $\frac{M_f(t_{\phi(n)} \delta_n)}{t_{\phi(n)}} \leq M_f(b \delta_n)$ . As  $f$  is continuous,  $M_f(b \delta_n)$  corresponds to a modulus of continuity, so  $\lim_{\delta_n \rightarrow 0} M_f(b \delta_n) = 0$ . This leads to a contradiction with (33). If  $\lim_{n \rightarrow \infty} t_n = \infty$ , then  $\frac{M_f(t_n \delta_n)}{t_n} \leq \frac{M_f(1)}{t_n}$ , and so  $\lim_{\delta_n \rightarrow 0} \frac{M_f(t_n \delta_n)}{t_n} = 0$ . Hence, there is a contradiction with (33).

We next analyze the case where  $f$  is  $\beta$ -Hölder continuous. Then  $M_f(\delta) \leq c_f \delta^\beta$ . Hence, for  $t \geq 1$  and  $t\delta \leq 1$ ,

$$\frac{M_f(t\delta)}{t} \leq c_f \delta^\beta t^{\beta-1} \leq c_f \delta^\beta,$$

and for  $t \geq 1$  and  $t\delta \geq 1$ ,

$$\frac{M_f(t\delta)}{t} = \frac{M_f(1)}{t} \leq \frac{c_f}{t} \leq c_f \delta.$$

As  $\delta \leq 1$  and  $0 < \beta \leq 1$ , we have that  $\Psi_f(\delta) \leq c_f \delta^\beta$ . □

*Proof of Proposition 5.3.* From (10), we have

$$\begin{aligned} J_{N,F}(\pi_N(\hat{u}_F)) &= \|\pi_N(\hat{u}_F)\|_N^2 + \frac{1}{\tau} \sum_{i=1}^n (P(\pi_N(\hat{u}_F))(x_i) - y_i)^2 \\ &= \|\pi_N(\hat{u}_F)\|_N^2 - \|\hat{u}_F\|_{\mathcal{H}_F}^2 + \|\hat{u}_F\|_{\mathcal{H}_F}^2 + \frac{1}{\tau} \sum_{i=1}^n (P(\pi_N(\hat{u}_F))(x_i) - y_i)^2. \end{aligned}$$

Using (6), then

$$\begin{aligned} J_{N,F}(\pi_N(\hat{u}_F)) &= \|\pi_N(\hat{u}_F)\|_N^2 - \|\hat{u}_F\|_{\mathcal{H}_F}^2 + J_F(\hat{u}_F) + \frac{1}{\tau} \sum_{i=1}^n (P(\pi_N(\hat{u}_F))(x_i) - y_i)^2 - (P(\hat{u}_F)(x_i) - y_i)^2 \\ &= \|\pi_N(\hat{u}_F)\|_N^2 - \|\hat{u}_F\|_{\mathcal{H}_F}^2 + J_F(\hat{u}_F) + \epsilon_N, \end{aligned}$$

with

$$\epsilon_N = \frac{1}{\tau} \sum_{i=1}^n [P(\pi_N(\hat{u}_F))(x_i) - P(\hat{u}_F)(x_i)] [P(\pi_N(\hat{u}_F))(x_i) + P(\hat{u}_F)(x_i) - 2y_i].$$

In the following, to simplify the notation, we will denote  $t_i^- = \max\{t, t \in S_N, t \leq x_i\}$ ,  $t_i^+ = \min\{t, t \in S_N, t \geq x_i\}$ ,  $w_- = w_{N_-}$  and  $w_+ = w_{N_+}$ .

From (22), we have the bound

$$\begin{aligned} |P\pi_N(\hat{u}_F)(x_i) - P\hat{u}_F(x_i)| &= [\pi_N(\hat{u}_F) - \hat{u}_F](t_i^-)w_-(x_i) + [\pi_N(\hat{u}_F) - \hat{u}_F](t_i^+)w_+(x_i) \\ &\leq \|\pi_N(\hat{u}_F) - \hat{u}_F\|_\infty \leq d_1\delta_N^{\beta/2}. \end{aligned}$$

Using Proposition 3.3, (23) and (24),

$$\begin{aligned} |P(\pi_N(\hat{u}_F))(x_i)| &= |\pi_N(\hat{u}_F)(t_i^-)w_-(x_i) + \pi_N(\hat{u}_F)(t_i^+)w_+(x_i)| \\ &\leq \|\pi_N(\hat{u}_F)\|_\infty \\ &\leq c\|\pi_N(\hat{u}_F)\|_N \\ &\leq c\|\hat{u}_F\|_{\mathcal{H}_F}, \\ |P(\hat{u}_F)(x_i)| &= |\hat{u}_F(t_i^-)w_-(x_i) + \hat{u}_F(t_i^+)w_+(x_i)| \\ &\leq \|\hat{u}_F\|_\infty \\ &\leq c\|\hat{u}_F\|_{\mathcal{H}_F}. \end{aligned}$$

Therefore,  $J_{N,F}(\pi_N(\hat{u}_F)) = -\hat{E}_N + J_F(\hat{u}_F) + \epsilon_N$ , with

$$|\epsilon_N| \leq d_3\delta_N^{\beta/2}, \quad \text{and} \quad d_3 = \frac{2nd_1}{\tau} \left( c\|\hat{u}_F\|_{\mathcal{H}_F} + \max_i |y_i| \right).$$

From the isometric property of  $\rho_N$  in (15), we have

$$\begin{aligned} J_{N,F}(\hat{u}_{N,F}) &= \|\hat{u}_{N,F}\|_N^2 + \frac{1}{\tau} \sum_{i=1}^n (P(\hat{u}_{N,F})(x_i) - y_i)^2 \\ &= \|\rho_N(\hat{u}_{N,F})\|_{\mathcal{H}_F}^2 + \frac{1}{\tau} \sum_{i=1}^n (P(\hat{u}_{N,F})(x_i) - y_i)^2 \\ &= J_F(\rho_N(\hat{u}_{N,F})) + \eta_N, \end{aligned}$$

with

$$\begin{aligned} \eta_N &= \frac{1}{\tau} \sum_{i=1}^n (P(\hat{u}_{N,F})(x_i) - y_i)^2 - (P(\rho_N(\hat{u}_{N,F}))(x_i) - y_i)^2 \\ &= \frac{1}{\tau} \sum_{i=1}^n [P(\hat{u}_{N,F})(x_i) - P(\rho_N(\hat{u}_{N,F}))(x_i)] [P(\hat{u}_{N,F})(x_i) + P(\rho_N(\hat{u}_{N,F}))(x_i) - 2y_i]. \end{aligned}$$

Thanks to (15), (23) and (24), we obtain

$$\begin{aligned} |P(\widehat{u}_{N,F}(x_i))| &\leq \|\widehat{u}_{N,F}\|_\infty \\ &\leq c\|\widehat{u}_{N,F}\|_N, \\ |P(\rho_N\widehat{u}_{N,F})(x_i)| &\leq \|\rho_N(\widehat{u}_{N,F})\|_\infty \\ &\leq c\|\rho_N(\widehat{u}_{N,F})\|_{\mathcal{H}_F} \\ &\leq c\|\widehat{u}_{N,F}\|_N, \end{aligned}$$

and hence,

$$P(\widehat{u}_{N,F} - \rho_N(\widehat{u}_{N,F}))(x_i) = (\widehat{u}_{N,F} - \rho_N(\widehat{u}_{N,F}))(t_i^-)w_-(x_i) + (\widehat{u}_{N,F} - \rho_N(\widehat{u}_{N,F}))(t_i^+)w_+(x_i).$$

For  $t \in F$ ,

$$\begin{aligned} (\widehat{u}_{N,F} - \rho_N(\widehat{u}_{N,F}))(t) &= \langle \widehat{u}_{N,F}, K_N(\cdot, t) \rangle_N - \langle \rho_N(\widehat{u}_{N,F}), K(\cdot, t) \rangle_{\mathcal{H}_F} \\ &= \langle \rho_N(\widehat{u}_{N,F}), \rho_N(K_N(\cdot, t)) \rangle_{\mathcal{H}_F} - \langle \rho_N(\widehat{u}_{N,F}), K(\cdot, t) \rangle_{\mathcal{H}_F} \\ &= \langle \rho_N(\widehat{u}_{N,F}), \rho_N(K_N(\cdot, t)) - K(\cdot, t) \rangle_{\mathcal{H}_F}, \\ |\widehat{u}_{N,F} - \rho_N(\widehat{u}_{N,F})(t)| &\leq \|\rho_N(\widehat{u}_{N,F})\|_{\mathcal{H}_F} \|\rho_N(K_N(\cdot, t)) - K(\cdot, t)\|_{\mathcal{H}_F} \\ &\leq \|\widehat{u}_{N,F}\|_N \sqrt{G_N}. \end{aligned}$$

From Proposition 4.2,  $J_{N,F}(\widehat{u}_{N,F}) = J_F(\rho_N(\widehat{u}_{N,F})) + \eta_N$  where

$$|\eta_N| \leq d_4 \delta_N^{\beta/2}, \quad d_4 = \frac{2n\sqrt{d_2}}{\tau} \|\widehat{u}_{N,F}\|_N \left( c\|\widehat{u}_{N,F}\|_N + \max_i |y_i| \right).$$

□

*Proof of (30) in Theorem 5.2.* Recall  $\widehat{h}^N = \rho_N(\widehat{u}_{N,F})$ . As  $\pi_N(\widehat{u}_F) \in \mathcal{H}_{N,F} \cap C_F$ , according to (25) and (26), and as  $\widehat{E}_N \geq 0$ ,

$$\|\widehat{h}^N\|_{\mathcal{H}_F}^2 \leq J_F(\widehat{h}^N) \leq J_{N,F}(\widehat{u}_{N,F}) + |\eta_N| \leq J_{N,F}(\pi_N\widehat{u}_F) + |\eta_N| \leq J_F(\widehat{u}_F) + |\epsilon_N| + |\eta_N|. \quad (34)$$

Hence, the sequence  $(\widehat{h}^N)_N$  is bounded in  $\mathcal{H}_F$  so that, by weak compactness in a Hilbert space, there exists a subsequence  $(\widehat{h}^{N_k})_k$  and  $h^* \in \mathcal{H}_F$  such that

$$\widehat{h}^{N_k} \xrightarrow[k \rightarrow \infty]{} h^* \in \mathcal{H}_F, \quad (\text{weak convergence}). \quad (35)$$

As  $\mathcal{H}_F$  is an RKHS with kernel  $K$ , for all  $t_i \in S_N$ ,  $K(\cdot, t_i) \in \mathcal{H}_F$  and

$$\langle \widehat{h}^{N_k}, K(\cdot, t_i) \rangle_{\mathcal{H}_F} = h^{N_k}(t_i) \xrightarrow[k \rightarrow \infty]{} \langle h^*, K(\cdot, t_i) \rangle_{\mathcal{H}_F} = h^*(t_i).$$

Therefore, for all  $N \geq 1$ ,  $\pi_N(\widehat{h}^{N_k}) \xrightarrow[k \rightarrow \infty]{} \pi_N(h^*)$  in the finite-dimensional space  $\mathcal{H}_N$ .

As  $S_N \subset S_{N+1}$ , as far as  $N_k \geq N$ ,  $\pi_N(\widehat{h}^{N_k}) = \pi_N(\rho_{N_k}(\widehat{u}_{N_k,F})) = \pi_N(\widehat{u}_{N_k,F})$ , so that

$$\pi_N(\widehat{u}_{N_k,F}) \xrightarrow[k \rightarrow \infty]{} \pi_N(h^*) \quad \text{in } \mathcal{H}_N.$$

As  $\mathcal{H}_N$  is an Hilbertian subspace of  $E_F$ ,

$$\pi_N(\widehat{u}_{N_k,F}) \xrightarrow[k \rightarrow \infty]{} \pi_N(h^*) \quad \text{in } E_F.$$

As  $\pi_N(\widehat{u}_{N_k,F}) \in C_F$  and  $C_F$  is closed in  $E_F$ , so that  $\forall N$ ,

$$\pi_N(h^*) \in C_F.$$

$C_F$  is closed in  $E_F$  and  $\pi_N(h^*) \xrightarrow{N \rightarrow \infty} h^*$  in  $E_F$ , then  $h^* \in C_F$  so that

$$J_F(\hat{u}_F) \leq J_F(h^*).$$

Then, as  $J_F$  is convex and lower semi continuous and  $\hat{h}^{N_k} \xrightarrow{k \rightarrow \infty} \hat{h}^* \in \mathcal{H}_F$  and thanks to (34), as  $\lim_{N \rightarrow \infty} \delta_N = 0$ , so that by Proposition 5.3,  $\lim_{N \rightarrow \infty} \epsilon_N = 0$  and  $\lim_{N \rightarrow \infty} \eta_N = 0$ ,

$$\begin{aligned} J_F(\hat{u}_F) &\leq J_F(h^*) \\ &\leq \liminf_k J_F(\hat{h}^{N_k}) \\ &\leq \liminf_k J_{N_k, F}(\hat{u}_{N_k, F}) \\ &\leq \liminf_k J_{N_k, F}(\pi_{N_k}(\hat{u}_F)) \\ &\leq \limsup_k J_{N_k, F}(\pi_{N_k}(\hat{u}_F)) \\ &\leq J_F(\hat{u}_F). \end{aligned}$$

Hence

$$\lim_{N \rightarrow \infty} J_F(\hat{h}^{N_k}) = \lim_{N \rightarrow \infty} J_{N_k, F}(\hat{u}_{N_k, F}) = \lim_{N \rightarrow \infty} J_{N_k, F}(\pi_{N_k}(\hat{u}_F)) = J_F(\hat{u}_F).$$

The real sequences  $(J_F(\hat{h}^N))$ ,  $(J_{N, F}(\hat{u}_{N, F}))$   $(J_{N, F}(\pi_N \hat{u}_F))$  are bounded and have a unique accumulation point  $J_F(\hat{u}_F)$  so that

$$\begin{aligned} J_F(\hat{h}^N) &\xrightarrow{N \rightarrow \infty} J_F(\hat{u}_F) \\ J_{N, F}(\hat{u}_{N, F}) &\xrightarrow{N \rightarrow \infty} J_F(\hat{u}_F) \\ J_{N, F}(\pi_N \hat{u}_F) &\xrightarrow{N \rightarrow \infty} J_F(\hat{u}_F). \end{aligned}$$

The following bounds

$$\begin{aligned} \|\hat{u}_{N, F} - \hat{u}_F\|_\infty &\leq c \|\pi_N(\hat{u}_F) - \hat{u}_{N, F}\|_N + \hat{F}_N, \\ \|\pi_N(\hat{u}_F) - \hat{u}_{N, F}\|_N^2 &\leq J_F(\hat{u}_F) - J_F(\hat{h}^N) + |\epsilon_N| + |\eta_N|, \end{aligned}$$

lead to (30) . □

## References

- F. Bachoc, A. Lagnoux, and A. F. López-Lopera. Maximum likelihood estimation for Gaussian processes under inequality constraints. *Electronic Journal of Statistics*, 13(2):2921–2969, 2019.
- F. Bachoc, A. F. López-Lopera, and O. Roustant. Sequential construction and dimension reduction of Gaussian processes under inequality constraints. *SIAM Journal on Mathematics of Data Science*, 4(2):772–800, 2022.
- X. Bay, L. Grammont, and H. Maatouk. Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation. *Electronic journal of statistics*, 10(1):1580–1595, 2016.
- X. Bay, L. Grammont, and H. Maatouk. A new method for interpolating in a convex subset of a Hilbert space. *Computational Optimization and Applications*, 68(1):95–120, 2017.
- P. Bellec. Sharp oracle inequalities for least squares estimators in shape restricted regression. *The Annals of Statistics*, 46(2):745–780, 2018.
- A. Berlinet and C. Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.

- A. Cousin, H. Maatouk, and D. Rullière. Kriging of financial term-structures. European Journal of Operational Research, 255(2):631–648, 2016. ISSN 0377-2217.
- C. Durot. Sharp asymptotics for isotonic regression. Probability theory and related fields, 122(2): 222–240, 2002.
- C. Durot and H. P. Lopuhaä. Limit theory in monotone function estimation. Statistical Science, 33 (4):547–567, 2018.
- M. G. Genton. Classes of kernels for machine learning: A statistics perspective. Journal of Machine Learning Research, 2:299–312, 2001.
- S. Golchi, D. R. Bingham, H. Chipman, and D. A. Campbell. Monotone emulation of computer experiments. SIAM/ASA Journal on Uncertainty Quantification, 3(1):370–392, 2015.
- L. Grammont, X. Bay, and H. Maatouk. Optimal smoothing and Gaussian processes with noisy data under constraints. hal-03625227, 2022, 2022.
- L. Grammont, H. Maatouk, and X. Bay. Equivalence between constrained optimal smoothing and Bayesian estimation. Journal of Nonparametric Statistics, pages 1–22, 2024.
- P. Groeneboom and G. Jongbloed. Nonparametric estimation under shape constraints, volume 38. Cambridge University Press, 2014.
- P. Groeneboom, G. Jongbloed, and J. Wellner. Estimation of a convex function: Characterizations and asymptotic theory. Annals of Statistics, 29(6):1653–1698, 2001.
- J.-B. Hiriart-Urruty. Bases, outils et principes pour l’analyse variationnelle, volume 70. Springer Science & Business Media, 2012.
- U. Hornung. Monotone spline-interpolation. In Numerische Methoden der Approximationstheorie, pages 172–191. Springer, 1978.
- S. Karlin and W. Studden. T-systems: with applications in analysis and statistics. Pure and Applied Mathematics, Interscience Publishers, New York, London, Sidney, 1966.
- T. Karvonen. Error bounds and the asymptotic setting in kernel-based approximation. Dolomites Research Notes on Approximation, 15(3):65–77, 2022.
- T. Karvonen. Asymptotic bounds for smoothness parameter estimates in Gaussian process interpolation. SIAM/ASA Journal on Uncertainty Quantification, 11(4):1225–1257, 2023.
- G. S. 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.
- L. Lin and D. B. Dunson. Bayesian monotone regression using Gaussian process projection. Biometrika, 101(2):303–317, 2014.
- A. Locatelli and A. Carpentier. Adaptivity to smoothness in X-armed bandits. In Proceedings of the 31st Conference On Learning Theory, volume 75, pages 1463–1492, 2018.
- W.-L. Loh. Estimating the smoothness of a Gaussian random field from irregularly spaced data via higher-order quadratic variations. The Annals of Statistics, 43(6):2766–2794, 2015.
- A. F. López-Lopera. lineqGPR: Gaussian process regression models with linear inequality constraints, 2022. URL <https://github.com/anfelopera/lineqGPR>. R package version 0.3.0.
- A. F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant. Finite-dimensional Gaussian approximation with linear inequality constraints. SIAM/ASA Journal on Uncertainty Quantification, 6(3): 1224–1255, 2018.



- A. F. López-Lopera, S. John, and N. Durrande. Gaussian process modulated Cox processes under linear inequality constraints. In International Conference on Artificial Intelligence and Statistics, pages 1997–2006, 2019.
- A. F. López-Lopera, F. Bachoc, N. Durrande, J. Rohmer, D. Idier, and O. Roustant. Approximating Gaussian process emulators with linear inequality constraints and noisy observations via MC and MCMC. In Monte Carlo and Quasi-Monte Carlo Methods, pages 363–381. Springer International Publishing, 2020.
- A. F. López-Lopera, F. Bachoc, and O. Roustant. High-dimensional additive Gaussian processes under monotonicity constraints. In Advances in Neural Information Processing Systems, volume 35, pages 8041–8053. Curran Associates, Inc., 2022.
- H. Maatouk and X. Bay. Gaussian process emulators for computer experiments with inequality constraints. Mathematical Geosciences, 49(5):557–582, 2017.
- C. E. Rasmussen and C. K. I. Williams. Gaussian Processes for Machine Learning. The MIT Press, Cambridge, MA, 2005. ISBN 026218253X.
- J. Riihimäki and A. Vehtari. Gaussian processes with monotonicity information. In International Conference on Artificial Intelligence and Statistics, pages 645–652, 2010.
- M. Stein. Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in Statistics. Springer, New York, 1999. ISBN 9780387986296.
- S. Zhou, P. Giuliani, J. Piekarewicz, A. Bhattacharya, and D. Pati. Reexamining the proton-radius problem using constrained Gaussian processes. Physical Review C, 99:055202, 2019.