Asymptotic analysis for covariance parameter estimation of Gaussian processes with functional inputs

Lucas Reding,* Andrés Felipe López-Lopera and François Bachoc

Univ. Polytechnique Hauts-de-France, INSA Hauts-de-France, CERAMATHS, F - 59313 Valenciennes, France
Email: lucas.reding@uphf.fr; andres.lopezlopera@uphf.fr
Institut de Mathématiques de Toulouse, Université Paul Sabatier, F-31062 Toulouse, France.
Email: francois.bachoc@math.univ-toulouse.fr

Abstract

We consider covariance parameter estimation for Gaussian processes with functional inputs. From an increasing-domain asymptotics perspective, we prove the asymptotic consistency and normality of the maximum likelihood estimator. We extend these theoretical guarantees to encompass scenarios accounting for approximation errors in the inputs, which allows robustness of practical implementations relying on conventional sampling methods or projections onto a functional basis. Loosely speaking, both consistency and normality hold when the approximation error becomes negligible, a condition that is often achieved as the number of samples or basis functions becomes large. These later asymptotic properties are illustrated through analytical examples, including one that covers the case of non-randomly perturbed grids, as well as several numerical illustrations.

1 Introduction

Gaussian processes (GPs) represent a widely used class of stochastic processes in geostatistics [31], machine learning [18, 27], computer experiments [8, 10], and engineering [12, 13]. A GP forms a non-parametric probabilistic framework wherein a prior belief, i.e. a distribution, is placed over a function that is either unknown or costly to evaluate. More precisely, the latter function is modeled as a realization of the (conditional) GP that interpolates some observation points. For instance, in computer experiments, if the latent function represents a computer code, then the observation points correspond to evaluations of this code [8]. In regression tasks, the predictive GP mean serves as an optimal linear unbiased predictor, and the predictive GP variance quantifies the expected square error associated with this predictor [27, 31].

The covariance function of the GP, also known as the kernel, plays an important role in defining correlations between values of the process at different input points. For instance, in geostatistics, this function seeks to quantify the similarity between data points, indicating that nearby spatial inputs are more likely to yield similar observation values. It is commonly assumed that the covariance function belongs to a parametric family of symmetric and positive semidefinite functions. This family is often defined based on prior beliefs (e.g. regularity, stationarity, periodicity) of the behavior of the function that is being modeled [16, 25]. Therefore, the choice of a covariance function boils down to the estimation of the associated parameters of the parametric family, namely the covariance parameters. The two main approaches for covariance parameter estimation are based on maximum likelihood (ML) or cross-validation [7, 4, 31], with the former being the most common one in both theoretical and practical studies.

---

*Corresponding author

Asymptotic guarantees, such as the consistency and normality, of ML estimators for GPs indexed by vectors in $\mathbb{R}^d$, with $d \geq 1$, have been further investigated in [1, 14, 19, 31, 35]. These guarantees are primarily explored within two frameworks: increasing-domain asymptotics and fixed-domain asymptotics. Within increasing-domain asymptotics, the smallest distance between two inputs is lower bounded. In contrast, fixed-domain asymptotics involves a sequence (or a triangular array) of data points that becomes densely distributed in a fixed, bounded subset of $\mathbb{R}^d$. While results from increasing-domain asymptotics are broadly applicable to various covariance functions, results from fixed-domain asymptotics require specific derivation on a case-by-case basis. This implies that for fixed-domain asymptotics, the analysis must be customized to the characteristics of each covariance function individually, limiting the generalization of the results across a broader range of functions.

Recent works have extended the implementation of GPs to non-Euclidean inputs, such as distributions [2, 5, 3, 6, 32], functions [9, 21, 23], categorical or integer-valued variables [15, 29], and graphs [11, 36]. There, adapted similarity distances have been considered to construct valid covariance functions. For instance, for distribution and functional inputs, similarity distances based on the Wasserstein distance [3] and the L_2-norm [21, 23] (respectively) have been considered. The references [2, 6] address multidimensional distributions via optimal transport and Hilbertian embedding. In particular, [6] uses the Wasserstein barycenters to define general covariance functions using optimal transportation maps. Despite significant numerical progress in these models, establishing theoretical guarantees certifying their robustness remains challenging. Among the few available theoretical guarantees, for the case of 1D distribution inputs, the consistency and normality of the ML estimator under increasing-domain asymptotics have been showed in [3].

Building on the promising numerical outcomes demonstrated by GPs with functional inputs for tasks such as surrogate modeling [9, 21, 23], sampling and sensitivity analysis [24], our focus in this paper is to establish asymptotic guarantees for this class of models. More precisely, we focus on the consistency and normality of the ML estimator from an increasing-domain asymptotic perspective. In a general context, our analysis considers stationary covariance functions on some Hilbert space. We first prove both the asymptotic consistency and normality of the ML estimator when the functional form of the inputs is assumed to be fully known. However, since this latter assumption may not be realistic in numerical implementations due to discretization or the availability of only partial observations [26], we also extend the previous asymptotic results to scenarios where errors (e.g. resulting from sampling or linear projections) are considered. These additional guarantees enable the certification of the robustness of numerical implementations of the GP model. Loosely speaking, both the consistency and normality of the ML estimator hold when the approximation error becomes negligible, a condition that is often achieved as the number of samples or basis functions tends to be infinite.

Throughout this paper, we provide theoretical and numerical examples to illustrate the proposed conditions and theorems. More precisely, we examine an example that covers the case of nonrandomly perturbed grids, showcasing that the sufficient conditions outlined in this work can be easily met. Additionally, we illustrate the resulting theorems with various analytical and numerical examples. The Python codes related to the numerical implementations are available in the Github repository: `https://github.com/anfelopera/fGPs`.

This paper is structured as follows. Sections 2 recalls essential definitions and properties related to GPs with functional inputs, as well as the ML estimator. In Sections 3 and 4, the focus shifts to introducing the necessary conditions, lemmas and theorems establishing the asymptotic consistency and normality of the ML estimator. There, we neglect the approximation errors over the functional inputs. Section 5 extends the previous asymptotic results to account for the approximation errors. Section 6 provides analytical examples and numerical illustrations. Finally, Section 7 summarizes the conclusions drawn from our analysis and outlines potential avenues for future research.
Variations of the outlined conditions for isotropic covariance functions, and the proofs of the lemmas are provided in Appendices 7, 7 and 7.

2 Gaussian processes with functional inputs

2.1 Generalities

Before introducing the GP framework that will be considered in this paper, we first discuss some generalities related to the functional context. Throughout the following sections, we consider a separable Hilbert space $\mathcal{H}$, equipped with an inner product denoted $\langle \cdot, \cdot \rangle$. Here, $\mathcal{H}$ does not need to be real nor of finite dimension. We define the norm derived from the inner product as $\| \cdot \|$. In other words, for any $f \in \mathcal{H}$, we have

$$\|f\| = \sqrt{\langle f, f \rangle} < \infty.$$

We denote, for any $k \geq 1$, $\langle \cdot, \cdot \rangle_{\mathbb{R}^k}$, the standard Euclidean inner product on $\mathbb{R}^k$, and $L_k^2([0,1])$ the space of $\mathbb{R}^k$-valued, square integrable functions defined on $[0,1]$. Note that if $k = 1$, we will simply write $L_1^2([0,1]) := L^2([0,1])$. In the following, although final applications in this paper consider $\mathcal{H} = L^2([0,1])$, the asymptotic analysis is valid for any Hilbert space $\mathcal{H}$. Next, we provide some examples of Hilbert spaces together with some corresponding inner products.

<u>Euclidean:</u> $\mathcal{H} = \mathbb{R}^d$ with $d \geq 1$ and endowed with the standard Euclidean inner product. This corresponds to standard GPs indexed by vectors [27].

<u>Univariate functional with multidimensional parameter:</u> $\mathcal{H} = \mathbb{R}^d \times L^2([0,1])$ with $d \geq 1$ and endowed with the inner product

$$\forall (x,f), (x',g) \in \mathcal{H}, \qquad \langle (x,f), (x',g) \rangle = \langle x, x' \rangle_{\mathbb{R}^d} + \int_0^1 f(t)g(t)\mathrm{d}t.$$

<u>Multivariate functional with multidimensional parameter:</u> $\mathcal{H} = \mathbb{R}^d \times L_n^2([0,1])$ with $n, d \geq 1$ and endowed with the inner product

$$\forall (x,f), (x',g) \in \mathcal{H}, \qquad \langle (x,f), (x',g) \rangle = \langle x, x' \rangle_{\mathbb{R}^d} + \int_0^1 \langle f(t), g(t) \rangle_{\mathbb{R}^n} \mathrm{d}t.$$

The previous two examples can, for instance, model a computer code with d scalar inputs and one scalar-valued or vector-valued (respectively) functional input [9].

<u>Complex functional with multidimensional parameter:</u> $\mathcal{H} = \mathbb{C}^d \times L^2([0,1]; \mathbb{C})$ with $d \geq 1$, and $L^2([0,1]; \mathbb{C})$ the space of $\mathbb{C}$-valued, square integrable functions defined on $[0,1]$. Denote by $(\cdot | \cdot)_{\mathbb{C}^d}$ a complex inner product on $\mathbb{C}^d$. Endowing $\mathcal{H}$ with the following inner product

$$\forall (x,f), (x',g) \in \mathcal{H}, \qquad \langle (x,f), (x',g) \rangle = (x | x')_{\mathbb{C}^d} + \int_0^1 f(t)\overline{g(t)}\mathrm{d}t,$$

yields an example of a complex, separable Hilbert space.

2.2 Gaussian processes

Let us consider a GP Y indexed by $\mathcal{H}$ and defined over an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with zero-mean function and unknown covariance function (kernel) K_0. We assume that K_0 belongs to a parametric family of the form

$$\{K_\theta; \theta \in \Theta\}, \tag{1}$$

with $\Theta \subset \mathbb{R}^p$ compact and θ in the interior of Θ. The parameter $\theta = (\theta_1, \ldots, \theta_p)$ is referred to as the covariance parameter of K_θ. In particular, we denote $K_0 := K_{\theta_0}$ for some unknown covariance parameter $\theta_0 \in \Theta$. For instance, it is customary to consider the squared exponential (SE) covariance structure (see, e.g., [3, 27]):

$$\left\{ K_\theta = \theta_1 e^{-\frac{\|f-g\|^2}{2\theta_2^2}}, \theta = (\theta_1, \theta_2) \in C \times C' \right\}, \tag{2}$$

for $f, g \in \mathcal{H}$ and C, C' compact sets in $(0, \infty)$. For the kernel structure in (2), (θ_1, θ_2) corresponds to the variance and the length-scale parameters (respectively) of K_θ. While θ_1 can be seen as a scale parameter of the output, θ_2 can be viewed as a scale parameter for the input variable.

The parameters θ are typically estimated from a training dataset which consists of tuples of inputs and outputs. In our case, this dataset is given by $(f_i, y_i)_{i=1,\ldots,n}$, where f_i can represent a deterministic function (i.e. $f_i \in \mathcal{H}$ with $\mathcal{H} = L^2([0,1])$) and y_i represents the corresponding output when considering f_i. For instance, in a coastal flooding context, the outputs may correspond to the outcome of a computer code modeling the water height of a flood event as a function of the tide, with the latter being a time-varying function [9, 21]. The underlying mapping function relating f_i and y_i can be then modeled as a realization of a GP Y such that $Y(f_i) = y_i$ for all $i = 1,\ldots,n$. Due to the properties of Gaussian distributions, the conditional process Y that interpolates the observations is also GP-distributed [27, 31]. In this paper, we are interested in the case where the underlying mapping function indeed comes from a GP Y with covariance function K_{θ_0}.

Covariance parameter estimation via maximum likelihood. Here, we focus on ML estimation (MLE), a widely used method that has received great attention from a theoretical point of view [4, 14, 31]. MLE involves maximizing the Gaussian likelihood, which is equivalent to minimizing the modified negative log Gaussian likelihood:

$$\mathcal{L}_n(\theta) = \frac{1}{n} \ln |R_\theta| + \frac{1}{n} y^\top R_\theta^{-1} y, \tag{3}$$

with $y = [y_1, \ldots, y_n]^\top$, $R_\theta = R(\theta) := (K_\theta(f_i, f_j))_{1 \le i,j \le n}$ and $|R_\theta|$ the determinant of R_θ. Then, the ML estimator is $\widehat{\theta}_n \in \mathrm{argmin}_{\theta \in \Theta} \mathcal{L}_n(\theta)$.

Gaussian process regression. In regression tasks, the value $Y(f)$, for any input $f \in \mathcal{H}$, can be predicted by plugging $\widehat{\theta}_n$ into the conditional expectation expression (also known as the posterior mean) for GPs [27, 31]. More precisely, denoting $\widehat{Y}_{\theta_0}(f)$ for the conditional expectation of $Y(f)$ given $Y(f_1) = y_1, \ldots, Y(f_n) = y_n$, then we have

$$\widehat{Y}_{\theta_0}(f) := \mathbb{E}(Y(f)|Y(f_1) = y_1, \ldots, Y(f_n) = y_n) = r_{\theta_0}^\top(f) R_{\theta_0}^{-1} y,$$

with $r_\theta(f) = [K_\theta(f, f_1), \ldots, K_\theta(f, f_n)]^\top$ and $R_\theta = (K_\theta(f_i, f_j))_{1 \le i,j \le n}$ for any $\theta \in \Theta$. Given that the parameter θ_0 is unknown in practice, the GP prediction at f is thus given by

$$\widehat{Y}_{\widehat{\theta}_n}(f) := r_{\widehat{\theta}_n}^\top(f) R_{\widehat{\theta}_n}^{-1} y.$$

Furthermore, the expected squared error of this estimate is obtained by computing the conditional variance at f

$$\mathrm{Var}(Y(f)|Y(f_1) = y_1, \ldots, Y(f_n) = y_n) = K_{\theta_0}(f, f) - r_{\theta_0}^\top(f) R_{\theta_0}^{-1} r_{\theta_0}(f).$$

and replacing the unknown parameter θ_0 by its estimate $\widehat{\theta}_n$, yielding the following estimator

$$K_{\widehat{\theta}_n}(f, f) - r_{\widehat{\theta}_n}^\top(f) R_{\widehat{\theta}_n}^{-1} r_{\widehat{\theta}_n}(f).$$

3 Sufficient conditions

In the increasing-domain asymptotic analysis, we need to consider a deterministic triangular array $\{f_1^n, \ldots, f_n^n\}_{n \in \mathbb{N}^*}$ of elements in $\mathcal{H}$. We require the input sequences to satisfy the following separability hypothesis.

Condition 3.1 *Suppose that there exists $\Delta > 0$ such that for all $n \in \mathbb{N}^*$ and for all $1 \le i \ne j \le n$, we have $\|f_i^n - f_j^n\| \ge 2\Delta$.*

When $\mathcal{H} = \mathbb{R}^d, d \ge 1$, Condition 3.1 is standard in the increasing-domain asymptotics literature. In the following, to ease the notation, we will simply write $f_i = f_i^n$ for all $1 \le i \le n$ and whenever n is fixed.

Condition 3.2 imposes a form of uniform control over the projections of the input elements over a fixed orthonormal basis. More precisely, we require that the projections whose orders exceed a certain value are uniformly (with respect to the input elements) bounded in such a way that the series of the squared bounds is convergent. This condition will be needed to use approximation arguments and derive both the asymptotic consistency as well as the asymptotic normality of the estimator.

Condition 3.2 *For $n \in \mathbb{N}^*$ and $1 \leq i \leq n$, we denote*

$$f_{i,d} = \sum_{j=1}^{d} \langle f_i, e_j \rangle e_j,$$

with $(e_j)_{j \in \mathbb{N}^*}$ an orthonormal basis of $\mathcal{H}$. We assume there is a J_0 and a sequence $(s_j)_{j \geq J_0}$ of non-negative numbers such that for all $1 \leq i \leq n$, $|\langle f_i, e_j \rangle| \leq s_j$, for $j \geq J_0$, with $\sum_{j=J_0}^{\infty} s_j^2 < \infty$.

Note that Condition 3.2 with $J_0 = 1$ and Condition 3.1 are incompatible. Indeed, if they hold, there exists d such that $|\|f_{i,d} - f_{j,d}\| - \|f_i - f_j\|| \leq 2\sqrt{\sum_{k=d+1}^{\infty} s_k^2} \leq \Delta$. Hence, for $i \neq j$, $\|f_{i,d} - f_{j,d}\| \geq \Delta$. And, since $J_0 = 1$, then

$$\|f_{i,d}\|^2 \leq \|f_i\|^2 \leq \sum_{k=1}^{\infty} s_k^2.$$

Identifying the projections $f_{i,d}$ with their coordinates in $\mathbb{R}^d$ through the canonical isometry τ_d, we find that there exists a ball in $\mathbb{R}^d$ which contains an infinite number of $\tau_d(f_{i,d})$ with $\|\tau_d(f_{i,d}) - \tau_d(f_{j,d})\| = \|f_{i,d} - f_{j,d}\| \geq \Delta$ for $i \neq j$, which is a contradiction.

To derive the consistency of the estimator, we will use approximation arguments. Thus, Lemma 3.1 will play an important role as it guarantees that, under the hypothesis of Theorem 4.6, there exists an approximation order large enough to ensure that the approximated input data is well-behaved.

Lemma 3.1 *Suppose that Conditions 3.1 and 3.2 are satisfied, then there exists $d_0 \geq J_0$ such that for all $1 \leq i < j \leq n$, $\|f_{j,d_0} - f_{i,d_0}\| \geq \Delta$ and $\sum_{k=d_0+1}^{\infty} s_k^2 < \Delta^2/16$.*

Proof of Lemma 3.1. Let $1 \leq i \neq j \leq n$ and $d \in \mathbb{N}^*$, then using the triangle inequality as well as Parseval's identity, we get

$$\|f_j - f_i\| \leq \|f_j - f_{j,d}\| + \|f_{i,d} - f_i\| + \|f_{j,d} - f_{i,d}\|$$

$$= \left(\sum_{k=d+1}^{\infty} |\langle f_i, e_k \rangle|^2 \right)^{\frac{1}{2}} + \left(\sum_{k=d+1}^{\infty} |\langle f_j, e_k \rangle|^2 \right)^{\frac{1}{2}} + \|f_{j,d} - f_{i,d}\|.$$

Thus, using Conditions 3.1 and 3.2, we have $\liminf_{d \to \infty} \|f_{j,d} - f_{i,d}\| > \Delta$. ∎

Henceforth, whenever both Conditions 3.1 and 3.2 are verified, we will denote by d_0 the integer given in Lemma 3.1 and which guarantees that its conclusions hold.

In everything that follows, we will write $a \lhd b$ whenever there exists a positive constant C which does not depend on a, or b, but may depend on some underlying fixed parameter, such that $a \leq Cb$.

Condition 3.3 *The parametric set of covariance functions given by (1) is supposed to be stationary, that is we can write the kernels K_θ as*

$$\forall \theta \in \Theta, \forall f, g \in \mathcal{H}, \ K_\theta(f, g) = K_\theta(f - g).$$

We also suppose that both Conditions 3.1 and 3.2 are satisfied, and that there exists a fixed $\gamma_0 > d_0$, such that for all $f \in \mathcal{H}$ we have

$$\sup_{\theta \in \Theta} |K_\theta(f)| \lhd \frac{1}{1 + \|f\|^{\gamma_0}}.$$

Condition 3.3 requires the underlying GP to be stationary. Note that a special case of a stationary process is an isotropic process, where for every isometry τ of $\mathcal{H}$, it holds for any $f, g \in \mathcal{H}$

$$K_0(f, g) = K_0(\|f - g\|) = \text{cov}(Y(f), Y(g)) = \text{cov}(Y(\tau(f)), Y(\tau(g))). \tag{4}$$

An example of an isotropic covariance function is the SE kernel in (2).

The inverse of the matrix R_θ also plays a key role in the definition of the ML estimator as can be seen in (3). As such, any convergence result regarding this estimator will require some form of control over R_θ^{-1} to avoid some approximation-related divergence phenomenon whenever $|R_\theta|$ is close to zero. Thus, Condition 3.4 can be understood as a stability condition.

Condition 3.4 *There exists a fixed $c > 0$ such that for any $\theta \in \Theta$, the sequence of matrices $R_\theta = (K_\theta(f_i, f_j))_{1 \le i,j \le n}$ satisfies*

$$\lambda_{\min}(R_\theta) \ge c,$$

where $\lambda_{\min}(R_\theta)$ denotes the smallest eigenvalue of R_θ. Thus $\lambda_{\max}(R_\theta^{-1}) \le 1/c$, where $\lambda_{\max}(R_\theta^{-1})$ denotes the largest eigenvalue of R_θ^{-1}.

For instance, in the case of a stationary covariance function given by $K_\theta(\omega) = \overline{K}_\theta(w) + \delta_\theta \mathbf{1}_{\{\omega=0\}}$ where $\overline{K}_\theta$ is a continuous covariance function and $\inf_{\theta \in \Theta} \delta_\theta > 0$, then Condition 3.4 is always satisfied.

A very common hypothesis within parametric estimation is that of the separability of the covariance parameters [1, 3]. Condition 3.5 imposes a form of global separability for the estimated parameter. Indeed, it can be interpreted as follows: given enough input data, it is always possible to distinguish between the true covariance parameter θ_0 and any other parameter θ. This condition is crucial to ensure that the studied problem is well-posed.

Condition 3.5 *For all $\alpha > 0$,*

$$\liminf_{n \to \infty} \inf_{\|\theta - \theta_0\| \ge \alpha} \frac{1}{n} \sum_{i,j=1}^n [K_\theta(f_i, f_j) - K_{\theta_0}(f_i, f_j)]^2 > 0.$$

Condition 3.6 can be interpreted as the asymptotic linear independence of the derivatives of the covariance function at θ_0. In particular, this condition will ensure that the asymptotic covariance matrix M_n given by (9) is well-defined, i.e. that it is indeed a covariance matrix.

Condition 3.6 *For all $(\lambda_1, \ldots, \lambda_p) \ne (0, \ldots, 0)$,*

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{i,j=1}^n \left(\sum_{k=1}^p \lambda_k \frac{\partial K_\theta}{\partial \theta_k} \bigg|_{\theta=\theta_0} (f_i, f_j) \right)^2 > 0.$$

Both Conditions 3.7 and 3.8 are standard regularity conditions within the ML estimation with Gaussian data. For these two conditions, we first need to assume that Conditions 3.1 and 3.2 are satisfied.

Condition 3.7 *For any $f \in \mathcal{H}$, $K_\theta(f)$ is continuously differentiable with respect to θ and, for some $\gamma_1 > d_0$ with d_0 as in Lemma 3.1, we have*

$$\sup_{\theta \in \Theta} \max_{i=1,\ldots,p} \left| \frac{\partial K_\theta}{\partial \theta_i}(f) \right| \lhd \frac{1}{1 + \|f\|^{\gamma_1}}.$$

Condition 3.8 *For any $f \in \mathcal{H}$, $K_\theta(f)$ is three times continuously differentiable with respect to θ and we have, for $q \in \{2, 3\}$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, and for some $\gamma_2 > d_0$ with d_0 as in Lemma 3.1,*

$$\sup_{\theta \in \Theta} \left| \frac{\partial^q K_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(f) \right| \lhd \frac{1}{1 + \|f\|^{\gamma_2}}.$$

The case of isotropic processes. All the conditions mentioned above hold for any stationary GP even if the process is not isotropic. However, when isotropy is indeed assumed, then we can relax some of the previous conditions to more tractable versions. The adaptations of those conditions to the isotropic case are summarized in Appendix 7.

We next give an example that covers the case of non-randomly perturbed grids in $\mathbb{N}^{d_0}$ with d_0 as in Lemma 3.1. Note that small perturbations of regular grids are often considered to improve estimation by creating pairs of observations that are less independent [1, 14].

Example 3.2 *Let* $(e_n)_{n \in \mathbb{N}}$ *be an orthonormal basis of* $\mathcal{H} = L^2([0,1])$ *and* $(x_i)_{i \in \mathbb{N}} \subset \mathbb{N}^{d_0}$ *be deterministic such that for all* $m \in \mathbb{N}$,

$$\{x_1, \ldots, x_{m^{d_0}}\} = \{1, \ldots, m\}^{d_0}.$$

We consider $\delta < \frac{1}{2}$ *fixed and we define*

$$f_i = \sum_{k=1}^{d_0} ((x_i)_k + \delta \xi_{i,k}) e_k + \sum_{k=d_0+1}^{\infty} \xi_{i,k} s_k e_k,$$

for $i \in \mathbb{N}$, *where* $(\xi_{i,k})_{i,k \in \mathbb{N}}$ *is deterministic in* $[-1,1]$. *In that case, Condition 3.1 is satisfied with* $\Delta \leq \frac{1-2\delta}{2}$. *Now, let* $\theta = (\theta_1, \theta_2)$ *with* Θ *compact in* $(0,\infty)^2$ *and* $K_\theta(f,g) = F_\theta(\|f - g\|) = \theta_1 e^{-\theta_2 \|f-g\|}$ *for all* $f, g \in L^2([0,1])$. *Then, Conditions 3.5, and 3.6 are also verified.*

Proof.
Let $n \in \mathbb{N}^*$ and $1 \leq i \neq j \leq n$. There exists $k \in \{1, \ldots, d_0\}$ such that $(x_i)_k \neq (x_j)_k$. Since, $\delta < \frac{1}{2}$, we have

$$\left\| \sum_{k=1}^{d_0} \left[(x_i - x_j)_k + \delta \xi_{i,j,k}^- \right] e_k + \sum_{k=d_0+1}^{\infty} s_k e_k \xi_{i,j,k}^- \right\| \geq 1 - 2\delta \geq 2\Delta, \tag{5}$$

with $\xi_{i,j,k}^- = \xi_{i,k} - \xi_{j,k}$ and Δ as mentioned in Example 3.2. In other words, Condition 3.1 is satisfied with $\Delta \leq \frac{1-2\delta}{2}$.

Before establishing that Condition 3.5, and 3.6 are also verified under the hypothesis of Example 3.2, we show a useful upper bound on the distance between two inputs. Consider $\theta = (\theta_1, \theta_2) \in \Theta$ and F_θ as described in Example 3.2. Let m be the largest integer such that $m^{d_0} \leq n$. Then m^{d_0}/n converges to 1 as $n \to \infty$ since d_0 is fixed. Then for all $i \in \{1, \ldots, m^{d_0}\}$, there is a corresponding j such that

$$\begin{aligned}
(x_i)_1 - (x_j)_1 &= \pm 1, \\
(x_i)_k - (x_j)_k &= 0, \quad k = 2, \ldots, d_0,
\end{aligned}$$

and thus

$$\left\| \sum_{k=1}^{d_0} \left[(x_i - x_j)_k + \delta \xi_{i,j,k}^- \right] e_k + \sum_{k=d_0+1}^{\infty} s_k e_k \xi_{i,j,k}^- \right\| \leq 1 + 2\delta(\sqrt{d_0 - 1} + 1) + 2R_{d_0}, \tag{6}$$

where $R_{d_0}^2 = \sum_{k=d_0+1}^{\infty} s_k^2$.
Now, we prove Conditions 3.5 and 3.6. Denote $G_{\theta,\theta'}(t) = F_\theta(t) - F_{\theta'}(t)$ and

$$D_{n,\delta} = \frac{1}{n} \sum_{i,j=1}^{n} G_{\theta,\theta'}^2 \left(\left\| \sum_{k=1}^{d_0} \left[(x_i - x_j)_k + \delta \xi_{i,j,k}^- \right] e_k + \sum_{k=d_0+1}^{\infty} s_k e_k \xi_{i,j,k}^- \right\| \right).$$

Suppose that $\theta_1 \neq \theta_1'$, then

$$D_{n,\delta} \geq \frac{1}{n} \sum_{i=1}^{n} G_{\theta,\theta'}^2(0) \geq \frac{1}{n} \sum_{i=1}^{n} (\theta_1 - \theta_1')^2 = (\theta_1 - \theta_1')^2,$$

7

and thus $\liminf_{n\to\infty} D_{n,\delta} > 0$. Now, without loss of generality, assume that $\theta_1 = \theta_1' = 1$ and $\theta_2 \neq \theta_2'$. According to the convexity of the function $\theta_2 \mapsto e^{-\theta_2 L}$, we have for any positive L,

$$\left| e^{-\theta_2 L} - e^{-\theta_2' L} \right| \geq L \left| \theta_2 - \theta_2' \right| e^{-\max(\theta_2,\theta_2')L}.$$

Thus, using (5) and (6), we get

$$D_{n,\delta} \geq \frac{(1-2\delta)^2}{n} \sum_{i=1}^{m^{d_0}} (\theta_2 - \theta_2')^2 e^{-2\max(\theta_2,\theta_2')[1+2\delta(\sqrt{d_0-1}+1)+2R_{d_0}]}.$$

We see that $\liminf_{n\to\infty} D_{n,\delta} > 0$. This concludes the proof for Condition 3.5.

To show that Condition 3.6 holds, we let $(\lambda_1, \lambda_2) \neq (0,0)$, $\theta_0 = (\theta_1, \theta_2)$. Note that $\left.\frac{\partial F_\theta(t)}{\partial \theta_1}\right|_{\theta=\theta_0} = e^{-\theta_2 t}$ as well as $\left.\frac{\partial F_\theta(t)}{\partial \theta_2}\right|_{\theta=\theta_0} = -\theta_1 t e^{-\theta_2 t}$ for all $t \geq 0$. Then,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i,j=1}^{n} \left[\sum_{k=1}^{p} \lambda_k \left.\frac{\partial K_\theta}{\partial \theta_k}\right|_{\theta=\theta_0} (f_i, f_j) \right]^2 = \liminf_{n\to\infty} \frac{1}{n} \sum_{i,j=1}^{n} \left([\lambda_1 - \lambda_2 \theta_1 g_{ij}] e^{-\theta_2 g_{ij}} \right)^2,$$

where $g_{ij} = \|f_i - f_j\|$. For $\lambda_1 \neq 0$,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i,j=1}^{n} \left([\lambda_1 - \lambda_2 \theta_1 g_{ij}] e^{-\theta_2 g_{ij}} \right)^2 \geq \liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \lambda_1^2 > 0.$$

If $\lambda_1 = 0$ and $\lambda_2 \neq 0$, then using (6), we obtain

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i,j=1}^{n} \lambda_2^2 \theta_1^2 g_{ij}^2 e^{-2\theta_2 g_{ij}} \geq \liminf_{n\to\infty} \frac{4}{n} \sum_{i=1}^{m^{d_0}} \lambda_2^2 \theta_1^2 \Delta^2 e^{-2\theta_2[1+2\delta(\sqrt{d_0-1}+1)+2R_{d_0}]} > 0.$$

This concludes the proof of Condition 3.6. ∎

4 Asymptotic results

In this section, we introduce the theorems related to the asymptotic properties of the ML estimator of a GP with functional inputs. For conciseness, the proofs of all the technical lemmas are summarized in Appendix 7.

4.1 Asymptotic consistency

We first focus on the asymptotic consistency of the ML estimator. Before presenting the main theorem, it is worth establishing the following technical lemmas to provide a detailed proof.

The following lemma, when combined with Conditions 3.1 and 3.2, enables control of the L^1-norm of the covariance matrix R_θ. This, coupled with an application of Gershgorin's circle theorem [20], allows for the control of the eigenvalues of R_θ.

Lemma 4.1 *For $g : \mathbb{R}^+ \to \mathbb{R}$ a non-increasing function and $(x_i)_{i \in \mathbb{N}^*}$ a sequence of real non-negative numbers, we have for any $n \in \mathbb{N}^*$,*

$$\sum_{i=1}^{n} g(x_i) \leq \sum_{k=1}^{\infty} \nu_k g(k-1),$$

with $\nu_k = \#\{i \ : \ x_i \in [k-1, k)\}$, $k \in \mathbb{N}^*$.

Lemma 4.2 *Assume that Conditions 3.1 to 3.3 hold. Then*

$$\sup_{\theta \in \Theta} \max_{j=1,\ldots,n} \sum_{i=1}^{n} |K_\theta(f_j, f_i)|$$

is bounded as $n \to \infty$.

Lemma 4.3 *Suppose that Conditions 3.1 to 3.3, and Condition 3.7 are all satisfied. Then, for $k = 1, \ldots, p$,*

$$\sup_{\theta \in \Theta} \max_{j=1,\ldots,n} \sum_{i=1}^{n} \left| \frac{\partial K_\theta}{\partial \theta_k}(f_j, f_i) \right|$$

is bounded as $n \to \infty$.

Note that only the first part of Condition 3.3, i.e. the stationary requirement, is needed for Lemma 4.3 to hold. However, we have decided to unify both parts as one condition for clarity purposes thereafter.

Lemma 4.4 *Under the same assumptions as in Lemma 4.3, we have that*

$$\sup_{\theta \in \Theta} \lambda_{\max}(R_\theta) \qquad and \qquad \sup_{\theta \in \Theta} \max_{i=1,\ldots,p} \lambda_{\max}\left(\frac{\partial R}{\partial \theta_i}(\theta) \right)$$

are bounded as $n \to \infty$.

Lemma 4.5 requires us to introduce a new item of notation. For any sequence of random variables $(X_n)_{n \in \mathbb{N}}$, we use the probabilistic Landau notation, that is

$$X_n = O_{\mathbb{P}}(1) \qquad \text{means} \qquad \lim_{M \to \infty} \limsup_{n \to \infty} \mathbb{P}(|X_n| > M) = 0.$$

Lemma 4.5 *Suppose that $M := \max_{i=1,\ldots,p} \sup_{\theta \in \Theta} \left| \frac{\partial \mathcal{L}_n}{\partial \theta_i}(\theta) \right| = O_{\mathbb{P}}(1)$ and that $\mathcal{L}_n(\theta) - \mathbb{E}\left[\mathcal{L}_n(\theta)\right] \xrightarrow[n \to \infty]{\mathbb{P}} 0$ for all $\theta \in \Theta$, then*

$$\sup_{\theta \in \Theta} |\mathcal{L}_n(\theta) - \mathbb{E}\left[\mathcal{L}_n(\theta)\right]| \xrightarrow[n \to \infty]{\mathbb{P}} 0.$$

We recall that the proofs of the lemmas outlined in this section are summarized in Appendix 7.

Having introduced the aforementioned lemmas, we can now establish Theorem 4.6 which gives the asymptotic consistency of the ML estimator.

Theorem 4.6 *Let $\widehat{\theta}_n \in \Theta$ be the maximum likelihood estimator obtained by minimizing the modified negative log Gaussian likelihood in (3). Under Conditions 3.1 to 3.5, as well as Condition 3.7, we have*

$$\widehat{\theta}_n \xrightarrow[n \to \infty]{\mathbb{P}} \theta_0.$$

Proof of Theorem 4.6. We start the proof in the same manner as in the proof of Proposition 3.1 in [1]. Indeed, we first aim to obtain the following uniform convergence

$$\sup_{\theta \in \Theta} |\mathcal{L}_n(\theta) - \mathbb{E}(\mathcal{L}_n(\theta))| \xrightarrow[n \to \infty]{\mathbb{P}} 0. \tag{7}$$

Let $n \in \mathbb{N}^*$ and recall that $\widehat{\theta}_n \in \operatorname{argmin}_{\theta \in \Theta} \mathcal{L}_n(\theta)$ with

$$\mathcal{L}_n(\theta) = \frac{1}{n} \ln |R_\theta| + \frac{1}{n} y^\top R_\theta^{-1} y.$$

Let $\theta \in \Theta$, it holds

$$\operatorname{Var}\left(\mathcal{L}_n(\theta)\right) = \frac{2}{n^2} \operatorname{tr}\left(R_\theta^{-1} R_{\theta_0} R_\theta^{-1} R_{\theta_0} \right).$$

9

However, according to Lemma 4.4 the quantity $\sup_{\theta \in \Theta} \lambda_{\max}(R_\theta)$ is bounded in n. Therefore, we have the convergence $\operatorname{Var}(\mathcal{L}_n(\theta)) \xrightarrow[n \to \infty]{} 0$ and so $\mathcal{L}_n(\theta) - \mathbb{E}(\mathcal{L}_n(\theta)) = o_{\mathbb{P}}(1)$. Moreover,

$$\sup_{\theta \in \Theta} \max_{i=1,\ldots,p} \lambda_{\max}\left(\frac{\partial R}{\partial \theta_i}(\theta)\right)$$

is also bounded in n. Thus, applying Lemma 4.5, we also have the uniform convergence in θ as given by (7).

In a second time, we shall establish that

$$\mathbb{E}(\mathcal{L}_n(\theta) - \mathcal{L}_n(\theta_0)) \rhd \|R_\theta - R_{\theta_0}\|_2^2, \tag{8}$$

where $\|R\|_2^2 := \frac{1}{n}\sum_{i,j=1}^n R_{i,j}^2$. To do so, we notice that since $\sup_\theta \lambda_{\max}(R_\theta)$ (resp. $\inf_\theta \lambda_{\min}(R_\theta)$) is uniformly bounded (resp. uniformly bounded away from 0) in n, then there exist $a, b > 0$ such that for all n and for all $\theta \in \Theta$, the spectrum of $R_\theta^{-1} R_{\theta_0}$ lies within $[a, b]$. Now, let $\theta \in \Theta$ and notice that

$$\mathbb{E}(\mathcal{L}_n(\theta)) = \frac{1}{n} \ln |R_\theta| + \frac{1}{n} \operatorname{tr}\left(R_\theta^{-1} R_{\theta_0}\right).$$

Denoting by $\lambda_1, \ldots, \lambda_n$ the eigenvalues of $R_\theta^{-1} R_{\theta_0}$, we get

$$\mathbb{E}(\mathcal{L}_n(\theta) - \mathcal{L}_n(\theta_0)) = \frac{1}{n}\left(\operatorname{tr}\left(R_\theta^{-1} R_{\theta_0}\right) - \ln |R_\theta^{-1} R_{\theta_0}| - n\right)$$

$$= \frac{1}{n} \sum_{i=1}^n (\lambda_i - \ln \lambda_i - 1).$$

Since, for all $t \in [a, b]$ we have $t - \ln t - 1 \rhd (t-1)^2$, then

$$\frac{1}{n}\sum_{i=1}^n (\lambda_i - \ln \lambda_i - 1) \rhd \frac{1}{n}\sum_{i=1}^n (\lambda_i - 1)^2$$

$$= \left\|I_n - R_\theta^{-1} R_{\theta_0}\right\|_2^2$$

$$= \left\|R_\theta^{-1}\left(R_\theta - R_{\theta_0}\right)\right\|_2^2.$$

However since $\sup_{\theta \in \Theta} \lambda_{\max}(R_\theta)$ is bounded in n, we have

$$\left\|R_\theta^{-1}\left(R_\theta - R_{\theta_0}\right)\right\|_2^2 \rhd \|R_\theta - R_{\theta_0}\|_2^2,$$

and thus, we can conclude that (8) is verified. From (7) and Condition 3.5, we have the convergence $\mathbb{P}(\|\widehat{\theta}_n - \theta_0\| > \epsilon) \xrightarrow[n \to \infty]{} 0$, for all $\epsilon > 0$, and so the conclusion Theorem 4.6 holds. ∎

4.2 Asymptotic normality

Similarly to Section 4.1, we first provide two technical lemmas that are necessary to establish the main theorem concerning asymptotic normality. We start by obtaining an extension of Lemma 4.3 to the partial derivatives of orders two and three.

Lemma 4.7 *Suppose that Conditions 3.1 to 3.3, and 3.8 are satisfied, then for all $q \in \{2, 3\}$ and for all $i_1, \ldots, i_q \in \{1, \ldots, p\}$,*

$$\sup_{\theta \in \Theta} \max_{j=1,\ldots,n} \sum_{i=1}^n \left| \frac{\partial^q K_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(f_j, f_i) \right|$$

is bounded as $n \to \infty$.

Using Lemma 4.7, we can extend Lemma 4.4 to the partial derivatives of orders two and three.

Lemma 4.8 *Assume that the same conditions as in Lemma 4.7 hold. Then, for all $q \in \{2,3\}$ and for all $i_1, \ldots, i_q \in \{1, \ldots, p\}$,*

$$\sup_{\theta \in \Theta} \lambda_{\max} \left(\frac{\partial^q R}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(\theta) \right)$$

is bounded as $n \to \infty$.

We can now introduce Theorem 4.9 which yields the asymptotic normality of the ML estimator.

Theorem 4.9 *For all $n \in \mathbb{N}$, let M_n be the $p \times p$ asymptotic covariance matrix defined by*

$$(M_n)_{i,j} = \frac{1}{2n} \operatorname{tr} \left(R_{\theta_0}^{-1} \frac{\partial R}{\partial \theta_i}(\theta_0) R_{\theta_0}^{-1} \frac{\partial R}{\partial \theta_j}(\theta_0) \right). \tag{9}$$

Suppose that Conditions 3.1 to 3.8 are all satisfied, then the maximum likelihood estimator $\widehat{\theta}_n$ is asymptotically normal:

$$\sqrt{n} M_n^{1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, I_p). \tag{10}$$

Furthermore,

$$0 < \liminf_{n \to \infty} \lambda_{\min}(M_n) \leq \limsup_{n \to \infty} \lambda_{\max}(M_n) < \infty. \tag{11}$$

Proof of Theorem 4.9. From Lemma 4.4 and Condition 3.4, we have for every $n \in \mathbb{N}$, $|(M_n)_{i,j}| \leq B$ for a fixed $B < \infty$. In addition, for any $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ such that $\sum_{i=1}^p \lambda_i^2 = 1$; from Condition 3.4, we get that

$$\begin{aligned}
\sum_{i,j=1}^p \lambda_i \lambda_j (M_n)_{i,j} &= \sum_{i,j=1}^p \lambda_i \lambda_j \frac{1}{2n} \operatorname{tr} \left(R_{\theta_0}^{-1} \frac{\partial R}{\partial \theta_i}(\theta_0) R_{\theta_0}^{-1} \frac{\partial R}{\partial \theta_j}(\theta_0) \right) \\
&= \frac{1}{2n} \operatorname{tr} \left(\sum_{i,j=1}^p \lambda_i R_{\theta_0}^{-1} \frac{\partial R}{\partial \theta_i}(\theta_0) R_{\theta_0}^{-1} \lambda_j \frac{\partial R}{\partial \theta_j}(\theta_0) \right) \\
&= \frac{1}{2n} \operatorname{tr} \left(R_{\theta_0}^{-1} \left[\sum_{i=1}^p \lambda_i \frac{\partial R}{\partial \theta_i}(\theta_0) \right] R_{\theta_0}^{-1} \left[\sum_{j=1}^p \lambda_j \frac{\partial R}{\partial \theta_j}(\theta_0) \right] \right) \\
&\triangleright \left\| \sum_{i=1}^p \lambda_i \frac{\partial R}{\partial \theta_i}(\theta_0) \right\|_2^2,
\end{aligned}$$

where we recall that $\|R\|_2^2 := \frac{1}{n} \sum_{i,j=1}^n R_{i,j}^2$. Hence, from Condition 3.6, we obtain

$$\liminf_{n \to \infty} \lambda_{\min}(M_n) > 0.$$

This establishes (11).

We move on to the proof of the asymptotic normality of the ML estimator. We start by assuming that

$$\sqrt{n} M_n^{1/2}(\widehat{\theta}_n - \theta_0) \; \cancel{\xrightarrow[n \to \infty]{\mathcal{L}}} \; \mathcal{N}(0, I_p). \tag{12}$$

Then there exists a bounded measurable function $g : \mathbb{R}^p \to \mathbb{R}$, $\xi > 0$, and a subsequence $\varphi(n)$ such that along $\varphi(n)$ we have

$$\left| \mathbb{E} \left(g \left(\sqrt{\varphi(n)} M_{\varphi(n)}^{1/2}(\widehat{\theta}_{\varphi(n)} - \theta_0) \right) \right) - \mathbb{E}(g(U)) \right| \geq \xi,$$

with $U \sim \mathcal{N}(0, I_p)$. In addition, by compactness and up to extracting another subsequence, we can assume that

$$M_{\varphi(n)} \xrightarrow[n \to \infty]{a.s.} M_\infty,$$

where M_∞ is a symmetric positive definite matrix.

Now the remaining of the proof is similar to the proof of Proposition 3.2 in [1]. Recall that for all n and all θ, we have

$$\mathcal{L}_n(\theta) = \frac{1}{n} \ln |R_\theta| + \frac{1}{n} y^\top R_\theta^{-1} y,$$

then, for any $i \in \{1, \ldots, p\}$,

$$\frac{\partial \mathcal{L}_n}{\partial \theta_i}(\theta_0) = \frac{1}{n} \operatorname{tr} \left(R_{\theta_0}^{-1} \frac{\partial R}{\partial \theta_i}(\theta_0) \right) - \frac{1}{n} y^\top R_{\theta_0}^{-1} \frac{\partial R}{\partial \theta_i}(\theta_0) R_{\theta_0}^{-1} y.$$

Hence, as in Proposition Appendix D.9 in [1], we get that

$$\sqrt{n} \nabla \mathcal{L}_n(\theta_0) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, 4M_\infty).$$

For the sake of compactness and readability, we denote $\frac{\partial R_{\theta_0}}{\partial \theta_i} := \frac{\partial R}{\partial \theta_i}(\theta_0)$ and $\frac{\partial^2 R_{\theta_0}}{\partial \theta_i \partial \theta_j} = \frac{\partial^2 R}{\partial \theta_i \partial \theta_j}(\theta_0)$. Let $i, j \in \{1, \ldots, p\}$, then we have

$$\frac{\partial^2 \mathcal{L}_n}{\partial \theta_i \partial \theta_j}(\theta_0) = \frac{1}{n} \operatorname{tr} \left(-R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_i} R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_j} + R_{\theta_0}^{-1} \frac{\partial^2 R_{\theta_0}}{\partial \theta_i \partial \theta_j} \right)$$
$$+ \frac{1}{n} y^\top R_{\theta_0}^{-1} \left(\frac{\partial R_{\theta_0}}{\partial \theta_i} R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_j} + \frac{\partial R_{\theta_0}}{\partial \theta_j} R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_i} - \frac{\partial^2 R_{\theta_0}}{\partial \theta_i \partial \theta_j} \right) R_{\theta_0}^{-1} y.$$

Thus, it follows from Condition 3.4, Lemma 4.4 and Lemma 4.8, that

$$\operatorname{Var} \left(\frac{\partial^2 \mathcal{L}_n}{\partial \theta_i \partial \theta_j}(\theta_0) \right) \xrightarrow[n \to \infty]{} 0. \tag{13}$$

Moreover, we also have

$$\mathbb{E} \left(\frac{\partial^2 \mathcal{L}_n}{\partial \theta_i \partial \theta_j}(\theta_0) \right) = \frac{1}{n} \operatorname{tr} \left(-R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_i} R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_j} + R_{\theta_0}^{-1} \frac{\partial^2 R_{\theta_0}}{\partial \theta_i \partial \theta_j} \right)$$
$$+ \frac{1}{n} \operatorname{tr} \left(2 R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_i} R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_j} - R_{\theta_0}^{-1} \frac{\partial^2 R_{\theta_0}}{\partial \theta_i \partial \theta_j} \right)$$
$$= \frac{1}{n} \operatorname{tr} \left(R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_i} R_{\theta_0}^{-1} \frac{\partial R_{\theta_0}}{\partial \theta_j} \right)$$
$$= (2M_n)_{i,j}.$$

This yields, using (13), the following convergence in probability

$$\frac{\partial^2 \mathcal{L}_n}{\partial \theta_i \partial \theta_j}(\theta_0) \xrightarrow[n \to \infty]{\mathbb{P}} (2M_\infty)_{i,j}. \tag{14}$$

Computing the third-order derivatives of the modified negative log-likelihood, we obtain

$$\frac{\partial^3 \mathcal{L}_n}{\partial \theta_i \partial \theta_j \partial \theta_k}(\theta_0) = \frac{1}{n} \operatorname{tr} \left(\frac{\partial}{\partial \theta_k} \left(-R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_i} R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_j} + R_\theta^{-1} \frac{\partial^2 R_\theta}{\partial \theta_i \partial \theta_j} \right) \right) \Big|_{\theta = \theta_0}$$
$$+ \frac{1}{n} y^\top \left(\frac{\partial}{\partial \theta_k} \left[R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_i} R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_j} R_\theta^{-1} + R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_j} R_\theta^{-1} \frac{\partial R_\theta}{\partial \theta_i} R_\theta^{-1} \right. \right.$$
$$\left. \left. - R_\theta^{-1} \frac{\partial^2 R_\theta}{\partial \theta_i \partial \theta_j} R_\theta^{-1} \right]_{\theta = \theta_0} \right) y$$
$$= \frac{1}{n} \operatorname{tr}(A_{\theta_0}) + \frac{1}{n} y^\top B_{\theta_0} y,$$

where A_{θ_0} and B_{θ_0} are sums of products of the matrices $R_{\theta_0}^{-1}$ or $\frac{\partial R}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(\theta_0)$ with $q \in \{0, \ldots, 3\}$ and $i_1, \ldots, i_q \in \{1, \ldots, p\}$. Hence, from Condition 3.4 and from Lemmas 4.4, and 4.8, we have

$$\sup_{\theta \in \Theta} \left| \frac{\partial^3 \mathcal{L}_n}{\partial \theta_i \partial \theta_j \partial \theta_k}(\theta) \right| = O_{\mathbb{P}}(1).$$

12

Using (14) and applying Proposition D.10 in [1], we can show that

$$\sqrt{\varphi(n)}(\widehat{\theta}_{\varphi(n)} - \theta_0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, M_\infty^{-1}).$$

Since $M_{\varphi(n)} \xrightarrow[n\to\infty]{a.s.} M_\infty$, we have

$$\sqrt{\varphi(n)}M_{\varphi(n)}^{1/2}(\widehat{\theta}_{\varphi(n)} - \theta_0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, I_p),$$

This is a contradiction with (12); which concludes the proof. ■

5 Asymptotic results with noisy inputs

In practical scenarios, complete knowledge of the functional inputs is often unattainable since only an approximate version of the input data is accessible. This approximation can be acquired through conventional sampling methods or through more advanced techniques such as projecting the input functions onto a fixed functional basis [26]. It is noteworthy that, when presenting numerical examples of certain asymptotic results for functional data, authors often tend to implicitly approximate their functional data using a sufficiently large sample vector. However, solely considering non-negligible approximation errors may render asymptotic results impractical. Thus, it is worthwhile to adapt the theorems in Section 4 to ensure that the numerical approximation of the inputs does not affect the convergence results.

To the best of our knowledge, this paper contains the first examination of the convergence of the ML estimator for approximated functional inputs within the GP framework. On a related topic for vector inputs, [14] examines the impact of covariance approximations. This more in-depth analysis of the asymptotic results for the approximated data has revealed, in certain cases, some non-trivial conditions for the convergence of the estimator to occur.

The proofs of intermediate lemmas are outlined in Appendix 7

5.1 Notation and sufficient conditions

Throughout this section, we use the notation introduced in Section 4. Given our focus on the asymptotic properties of approximated data, we consider some approximations of the input data $(f_i)_{i=1,\dots,n}$. For any $i \in \mathbb{N}$ and any $N \in \mathbb{N}$, we denote by f_i^N the approximated version of f_i. Similar to the case of noise-free inputs, our interest lies in studying the modified negative log Gaussian likelihood associated with the approximated data:

$$\mathcal{L}_{n,N}(\theta) = \frac{1}{n}\ln|R_{N,\theta}| + \frac{1}{n}y^\top R_{N,\theta}^{-1}y, \tag{15}$$

with $R_N(\theta) = R_{N,\theta} := (K_\theta(f_i^N, f_j^N))_{1\leq i,j\leq n}$ and $|R_{N,\theta}|$ the determinant of $R_{N,\theta}$. Then, the ML estimator is given by $\widehat{\theta}_{n,N} \in \mathrm{argmin}_{\theta\in\Theta}\,\mathcal{L}_{n,N}(\theta)$.

For the asymptotic results to hold for approximated data, we will require that the approximation order N satisfies Condition 5.1.

Condition 5.1 *There exists a sequence N_n such that as $n \to \infty$, we have*

$$\max_{i=1,\dots,n} \|f_i - f_i^{N_n}\| \xrightarrow[n\to\infty]{} 0.$$

In particular, it should be noted that this condition implies that the order of approximation N must depend on the number of observations n for convergence to hold in a general setting.
Additionally, for Condition 3.3 to be correctly passed down to the approximated data, we will require some local control over the rate of change of the covariance function. This control is detailed in Condition 5.2

Condition 5.2 *Keeping the same constant γ_0 as in Condition 3.3, there exists another constant $\rho_0 > 0$ such that for all $f, g \in \mathcal{H}$ satisfying $0 < \|f - g\| \leq \rho_0$, we have*

$$\sup_{\theta \in \Theta} \frac{|K_\theta(f) - K_\theta(g)|}{\|f - g\|} \lhd \frac{1}{1 + \min(\|f\|, \|g\|)^{\gamma_0}}.$$

To obtain the asymptotic normality of the ML estimator for exact data, i.e. Theorem 4.9, we also required some control over the partial derivatives of orders two and three of the function K_θ (see Condition 3.8). In order to pass down this condition to the approximated data we will require, in the same manner as Condition 5.2, a local control over the rate of change of the partial derivatives of the covariance function.

Condition 5.3 *Assume that Conditions 3.7 holds and keep the same constant $\gamma_1 > 0$. Then there exists another constant $\rho_1 > 0$ such that for all $i \in \{1, \ldots, p\}$, and for all $f, g \in \mathcal{H}$ such that $0 < \|f - g\| \leq \rho_1$, we have*

$$\sup_{\theta \in \Theta} \frac{1}{\|f - g\|} \left| \frac{\partial K_\theta}{\partial \theta_i}(f) - \frac{\partial K_\theta}{\partial \theta_i}(g) \right| \lhd \frac{1}{1 + \min(\|f\|, \|g\|)^{\gamma_1}}.$$

Condition 5.4 *Assume that Conditions 3.8 holds and keep the same constant $\gamma_2 > 0$. Then there exists another constant $\rho_2 > 0$ such that for all $q \in \{2, 3\}$, for all $i_1, \ldots, i_q \in \{1, \ldots, p\}$, and for all $f, g \in \mathcal{H}$ such that $0 < \|f - g\| \leq \rho_2$, we have*

$$\sup_{\theta \in \Theta} \frac{1}{\|f - g\|} \left| \frac{\partial^q K_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(f) - \frac{\partial^q K_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(g) \right| \lhd \frac{1}{1 + \min(\|f\|, \|g\|)^{\gamma_2}}.$$

The case of isotropic processes. As in Section 4, we can give some simplified versions of Conditions 5.2, 5.3 and 5.4 whenever the underlying process is assumed to be isotropic. The adaptations of these conditions are given in Appendix 7 (Conditions 7.4, 7.5, and 7.6 respectively).

5.2 Asymptotic consistency

We start by establishing the consistency of the ML estimator. To prove Theorem 5.2, we make use of the technical Lemma 5.1 which will allow us to control $\lambda_{\max}(R_{N_n, \theta} - R_\theta)$.

Lemma 5.1 *Assume that Conditions 3.1, 3.2, and 5.1 hold. If Condition 5.2 is verified, then*

$$\sup_{\theta \in \Theta} \lambda_{\max} \left(R_\theta - R_{N_n, \theta} \right) \xrightarrow[n \to \infty]{} 0.$$

If Condition 5.3 is verified, then we have

$$\sup_{\theta \in \Theta} \max_{i = 1, \ldots, p} \lambda_{\max} \left(\frac{\partial}{\partial \theta_i}(R_\theta - R_{N_n, \theta}) \right) \xrightarrow[n \to \infty]{} 0.$$

Next, we introduce Theorem 5.2. As mentioned before, this theorem guarantees that the numerical results obtained using only an approximation of the input functions will still converge correctly.

Theorem 5.2 *Let $\widehat{\theta}_{n,N} \in \Theta$ be the maximum likelihood estimator obtained by minimizing the negative log Gaussian likelihood of the approximated data in (15). Under Conditions 3.1 to 3.5, as well as Condition 3.7, Condition 5.1, Condition 5.2 and Condition 5.3, we have*

$$\widehat{\theta}_{n,N_n} \xrightarrow[n \to \infty]{\mathbb{P}} \theta_0.$$

Proof of Theorem 5.2.
Lemma 5.1 enables us to extend Conditions 3.4 and 3.5 to noisy inputs:

$$\sup_{\theta \in \Theta} \lambda_{\min}(R_{N_n, \theta}) \geq c, \qquad c > 0,$$

$$\liminf_{n\to\infty} \inf_{\|\theta-\theta_0\|\geq\alpha} \frac{1}{n} \sum_{i,j=1}^{n} [K_\theta(f_i^{N_n}, f_j^{N_n}) - K_{\theta_0}(f_i^{N_n}, f_j^{N_n})]^2 > 0, \qquad \forall \alpha > 0.$$

From Lemmas 4.4 and 5.1, we have that

$$\sup_{\theta\in\Theta} \lambda_{\max}(R_{N_n,\theta}) \quad \text{and} \quad \sup_{\theta\in\Theta} \max_{i=1,\dots,p} \lambda_{\max}\left(\frac{\partial R_{N_n}}{\partial\theta_i}(\theta)\right)$$

are bounded as $n \to \infty$. Then, we can follow the proof of Theorem 4.6 to conclude the proof. ∎

5.3 Asymptotic normality

For the asymptotic normality to hold, one requires some control over the partial derivatives of orders two and three of the modified negative log-likelihood. To pass down this property to the approximate data case, we will need to impose control over the rate of change of said derivatives; this will be done thanks to Conditions 5.3 and 5.4 through Lemma 5.3.

Lemma 5.3 *Assume that Conditions 3.1, 3.2, 5.1, 5.3, and 5.4 hold. Then for $q \in \{1,2,3\}$, $i_1, \dots, i_q \in \{1, \dots, p\}$, we have*

$$\sup_{\theta\in\Theta} \max_{i=1,\dots,p} \lambda_{\max}\left(\frac{\partial}{\partial\theta_{i_1}} \cdots \frac{\partial}{\partial\theta_{i_q}}(R_\theta - R_{N_n,\theta})\right) \xrightarrow[n\to\infty]{} 0.$$

We now provide the asymptotic normality of the estimator for the approximated data.

Theorem 5.4 *Let $M_{n,N}$ be the $p \times p$ asymptotic covariance matrix defined by*

$$(M_{n,N})_{i,j} = \frac{1}{2n} \operatorname{tr}\left(R_{N,\theta_0}^{-1} \frac{\partial R_N}{\partial\theta_i}(\theta_0) R_{N,\theta_0}^{-1} \frac{\partial R_N}{\partial\theta_j}(\theta_0)\right).$$

Under Conditions 3.1 to 3.8, and Conditions 5.1 to 5.4, we have

$$\sqrt{n} M_{n,N_n}^{1/2}(\widehat{\theta}_{n,N_n} - \theta_0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, I_p). \tag{16}$$

Furthermore,

$$0 < \liminf_{n\to\infty} \lambda_{\min}(M_{n,N_n}) \leq \limsup_{n\to\infty} \lambda_{\max}(M_{n,N_n}) < \infty. \tag{17}$$

Proof. From Lemmas 4.4 and 5.1, and Condition 3.4, we show (17) similarly as in the proof of Theorem 4.9. The arguments of the remaining of the proof are similar to the ones considered in the proof of Theorem 4.9. ∎

As observed in (16), the renormalization matrix M_{n,N_n} depends on the approximation order N_n. This dependency on N_n poses some challenges in computing confidence ensembles for the estimated parameters θ_0. One way to solve this issue is to prove that the covariance matrix for the approximated data, that is M_{n,N_n}, is asymptotically close to the asymptotic covariance matrix M_n which is the aim of Lemma 5.5.

Lemma 5.5 *Consider the same conditions as in Theorem 5.4. We have the following convergence of the covariance matrix for the approximated data, denoted by M_{n,N_n}, toward the theoretical covariance matrix M_n: as $n \to \infty$,*

$$M_{n,N_n} - M_n \xrightarrow[n\to\infty]{} 0.$$

Using Lemma 5.5, we can obtain an extension of Theorem 5.4 renormalized using the theoretical covariance matrix M_n instead.

Corollary 5.6 *Under the hypothesis of Theorem 5.4, we have*

$$\sqrt{n} M_n^{1/2}(\widehat{\theta}_{n,N_n} - \theta_0) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, I_p). \tag{18}$$

Proof. According to Theorem 5.4, when $n \to \infty$, we have the convergence

$$\sqrt{n} M_{n,N_n}^{1/2} (\widehat{\theta}_{n,N_n} - \theta_0) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, I_p). \tag{19}$$

Moreover, for any $n \in \mathbb{N}^*$, it holds

$$\sqrt{n} M_n^{1/2} (\widehat{\theta}_{n,N_n} - \theta_0) = \sqrt{n} M_{n,N_n}^{1/2} (\widehat{\theta}_{n,N_n} - \theta_0) + \sqrt{n} (M_n^{1/2} - M_{n,N_n}^{1/2})(\widehat{\theta}_{n,N_n} - \theta_0).$$

However, according to Lemma 5.5, we have

$$M_n^{1/2} - M_{n,N_n}^{1/2} \xrightarrow[n \to \infty]{} 0.$$

Then applying Slutsky theorem, we can conclude the proof. ∎

6 Analytical examples and numerical illustrations

6.1 Analytical examples

Here, we give some examples of the application of the theorems mentioned in Section 5. We consider the Hilbert space $\mathcal{H} = L^2([0,1])$ endowed with its usual inner product as well as the subspace

$$E_\omega := \{f \in \mathcal{H}; \forall u, v \in [0,1], |f(u) - f(v)| \leq \omega(|u - v|)\},$$

of functions of $\mathcal{H}$ which admits $\omega : [0,1] \to [0,1]$ as a global modulus of continuity. Recall that $\omega(t) \to 0$ as $t \to 0$. Examples of such spaces include spaces of continuous functions, spaces of Lispchitz continuous functions, or spaces of Hölder continuous functions.

Recall that we denote by $\| \cdot \|$ the norm on $\mathcal{H}$ induced by the usual inner product, that is for any $f \in \mathcal{H}$

$$\|f\|^2 = \int_0^1 f^2(x) dx.$$

We also denote by $\| \cdot \|_\infty$ the uniform norm, that is for $f : [0,1] \to \mathbb{R}$, $\|f\|_\infty = \sup_{x \in [0,1]} |f(x)|$. Finally, we assume that the input functions f_n, $n \in \mathbb{N}$, are all elements of the linear subspace E_ω.

Example 6.1 (Sampling) *Let f_i^N be the piecewise constant approximation of f_i given by*

$$f_i^N(x) = f_i \left(\frac{\overline{(N-1)x}}{N-1} \right) \qquad x \in [0,1],$$

with $\overline{(N-1)x}$ the integer part of $(N-1)x$. Then,

$$\|f_i - f_i^{N_n}\| \leq \|f_i - f_i^{N_n}\|_\infty \leq \omega \left(\frac{1}{N_n - 1} \right),$$

which verifies Condition 5.1 for any $N_n \to \infty$.

Example 6.2 (Random sampling with time-independent samples) *Consider random sampling where samples are fixed for all $i \in \mathbb{N}$. Let $U_1, \ldots, U_N \underset{iid}{\sim} \mathcal{U}([0,1])$ and denote the order process by $0 = U_{(0)} < U_{(1)} < \cdots < U_{(N)}$. The random sampling approximation is given by*

$$f_i^N(x) = f_i \left(U_{(I_N(x))} \right),$$

with $I_N(x) = \max \{j \in \{0, \ldots, N\}; U_{(j)} \leq x\}$.
As $N \to \infty$, we have $\max_{j=1,\ldots,N} (U_{(j)} - U_{(j-1)}) \to 0$ almost surely. Hence, for any $N_n \to \infty$, Condition 5.1 is verified almost surely, by the same arguments as in Example 6.1.

Example 6.3 (Random sampling with time-varying samples) *Consider now that $U_1^i, \ldots, U_N^i \underset{iid}{\sim}$ $\mathcal{U}([0,1])$. Denote the order process by $0 = U_{(0)}^i < U_{(1)}^i < \cdots < U_{(N)}^i$ for all $i \in \mathbb{N}$. The random sampling approximation is given by*

$$f_i^N(x) = f_i\left(U_{(I_N^i(x))}^i\right),$$

with $I_N^i(x) = \max\left\{j \in \{0, \ldots, N\}; U_{(j)}^i \leq x\right\}$. For all n and for all $i = 1, \ldots, n$, it holds

$$\|f_i - f_i^{N_n}\| < \sup_{\substack{f \in E_\omega \\ x \in [0,1]}} \left|f(x) - f\left(U_{(I_{N_n}^i(x))}^i\right)\right| \leq \max_{j=1,\ldots,N_n} \omega(U_{(j)}^i - U_{(j-1)}^i).$$

Let $\delta > 0$ and denote by $F^{(N)}$ the empirical cumulative distribution function of $U_1, \ldots, U_N$. Then, since $F^{(N)}$ is almost surely piecewise constant in between the observations $U_1^1, \ldots, U_N^1$, we deduce that

$$\left\{\exists j \in \{1, \ldots, N_n\}, U_{(j)}^1 - U_{(j-1)}^1 > \delta\right\} \subset \left\{\sup_{x \in [0,1]} \left|F^{(N)}(x) - x\right| \geq \frac{\delta}{2}\right\},$$

and therefore

$$\mathbb{P}\left(\max_{j=1,\ldots,N_n}\left(U_{(j)}^1 - U_{(j-1)}^1\right) > \delta\right) \leq \mathbb{P}\left(\sup_{x \in [0,1]} \left|F^{(N)}(x) - x\right| \geq \frac{\delta}{2}\right).$$

Moreover, letting $\omega^{-1}(t) = \inf\{s \geq 0; \omega(s) \geq t\}$,

$$\mathbb{P}\left(\max_{i=1,\ldots,n} \|f_i - f_i^{N_n}\| > \delta\right) \leq n\mathbb{P}(\|f_1 - f_1^{N_n}\| > \delta)$$

$$\leq n\mathbb{P}\left(\omega\left(\max_{j=1,\ldots,N_n}\left(U_{(j)}^1 - U_{(j-1)}^1\right)\right) \geq \delta\right)$$

$$\leq n\mathbb{P}\left(\max_{j=1,\ldots,N_n}\left(U_{(j)} - U_{(j-1)}\right) \geq \omega^{-1}(\delta)\right)$$

$$\leq n\mathbb{P}\left(\|F^{(N_n)} - Id\|_\infty \geq \frac{\omega^{-1}(\delta)}{2}\right).$$

Then, by using the Dvoretzky-Kiefer-Wolfowitz inequality [22], we have

$$\mathbb{P}\left(\max_{i=1,\ldots,n} \|f_i - f_i^{N_n}\| > \delta\right) \leq 2ne^{-N_n(\omega^{-1}(\delta))^2/2}.$$

Hence, if $\frac{N_n}{\log n} \to \infty$ as $n \to \infty$, then we have

$$\max_{i=1,\ldots,n} \|f_i - f_i^{N_n}\| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Then, there exists a sequence $\xi_n \xrightarrow[n\to\infty]{} 0$ such that

$$\mathbb{P}\left(\max_{i=1,\ldots,n} \|f_i - f_i^{N_n}\| > \xi_n\right) \xrightarrow[n\to\infty]{} 0.$$

Denote

$$\overline{f}_i^{N_n} = \begin{cases} f_i^{N_n}, & \text{if } \max_{i=1,\ldots,n} \|f_i - f_i^{N_n}\| \leq \xi_n, \\ f_i, & \text{otherwise.} \end{cases}$$

Let $\widehat{\theta}_{n,N_n}$ and $\overline{\widehat{\theta}}_{n,N_n}$ be the estimated parameters from $f_i^{N_n}$ and $\overline{f}_i^{N_n}$, respectively. Then, almost surely

$$\max_{i=1,\ldots,n} \|f_i - \overline{f}_i^{N_n}\| \leq \xi_n,$$

which implies, for almost all samplings, both the consistency and normality of $\overline{\widehat{\theta}}_{n,N_n}$ through Theorems 5.2 and 5.4 as long as the rest of the hypotheses are satisfied. Since $\mathbb{P}(\overline{\widehat{\theta}}_{n,N_n} = \widehat{\theta}_{n,N_n}) \to 1$ as $n \to \infty$, both guarantees also hold for $\widehat{\theta}_{n,N_n}$ for almost all samplings.

Example 6.4 (Bernstein approximation) *For any N, and any $k \in \{1, \ldots, N\}$, denote by $B_{N,k}$: $x \mapsto \binom{N}{k} x^k (1-x)^{N-k}$ the Bernstein polynomials of degree N with binomial coefficients $\binom{N}{k}$. The Bernstein approximation f_i^N of f_i is then given by $f_i^N(x) = \sum_{k=0}^N f\left(\frac{k}{N}\right) B_{N,k}(x)$ for all $x \in [0,1]$. In this case ([30], Theorem 5.3.2) states*

$$\|f_i - f_i^{N_n}\| \leq \frac{9}{4} \omega \left(\frac{1}{\sqrt{N_n}}\right),$$

which satisfies Condition 5.1 for any $N_n \xrightarrow{n \to \infty} \infty$.

Example 6.5 (Cubic spline approximation) *Let $\xi_1, \ldots, \xi_N$ be the knots of the spline approximation such that $0 = \xi_1 < \cdots < \xi_N = 1$ and $\Delta \xi_j = \xi_{j+1} - \xi_j$ for all $j \in \{1, \ldots, N-1\}$. Assume that $\max_{j=1,\ldots,N-1} \Delta \xi_i \to 0$ as $N \to \infty$. For all $i \in \mathbb{N}$, consider $f_i \in \mathcal{C}^4([0,1])$ with $\mathcal{C}^4$ the set of four times continuously differentiable functions. Define $f_i^{(4)}$ as the 4-th derivative of f_i, and f_i^N as the cubic spline approximation of f_i. Then ([17], Theorem 1),*

$$\|f_i - f_i^{N_n}\| \leq \|f_i - f_i^{N_n}\|_\infty \triangleleft \left(\max_{j=1,\ldots,N_n-1} \Delta \xi_j\right)^4 \|f^{(4)}\|_\infty,$$

which verifies Condition 5.1 for any $N_n \to \infty$.

Note that there exist many different error bounds for spline approximations under a variety of hypotheses (for a more in-depth analysis, see [28]) which would lead to different but similar cases as Example 6.5.

6.2 Numerical illustrations

We now numerically verify Theorems 5.2 and 5.4 for the examples in Section 6.1. Due to computational challenges involved in both data generation and covariance parameter estimation, we have limited the asymptotic analysis to $n \leq 160$ and $N_n \leq 160$. This choice is based on the fact that conducting these tasks for $n > 160$ would require more than dozens of hours. Furthermore, as observed in Figures 1 and 2, it is worth noting that convergence results become visible when $n = 160$. Our experiments have been executed on an 11th Gen Intel(R) Core(TM) i5-1145G7 @ 2.60GHz 16 Gb RAM. The Python source codes are available on the GitHub repository: `https://github.com/anfelopera/fGPs`. The different plots have been generated using the Python toolbox Seaborn [34].

We consider inputs given by the functional form:

$$f_t(x) = 2t + \cos(U_0(t)x)e^{-U_1(t)x}, \tag{20}$$

with $x \in [0,1]$, and $(U_0(t))_{t \in \mathbb{N}}, (U_1(t))_{t \in \mathbb{N}}$ two i.i.d. sequences of i.i.d. random variables sampled from a uniform distribution $U[0,1]$. While U_0 introduces oscillations in f, U_1 controls its decay rate. The parameter $t \in \mathbb{N}$ is introduced to satisfy Condition 3.1 related to the increasing-domain asymptotics.

We sample n random inputs $f_1, \ldots, f_n$ from (20). These inputs are then used to compute the corresponding outcomes $y_i = Y(f_i)$, where $Y(f_i)$ are obtained as a random realization of a GP with the SE kernel in (2) and covariance parameters $\theta_0 = (0.5, 0.1)$. These values have been arbitrarily chosen for illustrative purposes. The resulting collection of tuples $(f_i, y_i)_{i=1,\ldots,n}$ form the training dataset. For the asymptotic analysis, we consider $n \geq 20$.

For a given value of n, we repeat the aforementioned strategy for generating 10^3 distinct replicates of the training dataset. Then, for each replicate, we train a new GP with a SE kernel. We consider approximations based on sampling, random sampling with time-independent samples or with time-varying samples, and Bernstein polynomials. For the latter three cases, the L_2-norm is approximated
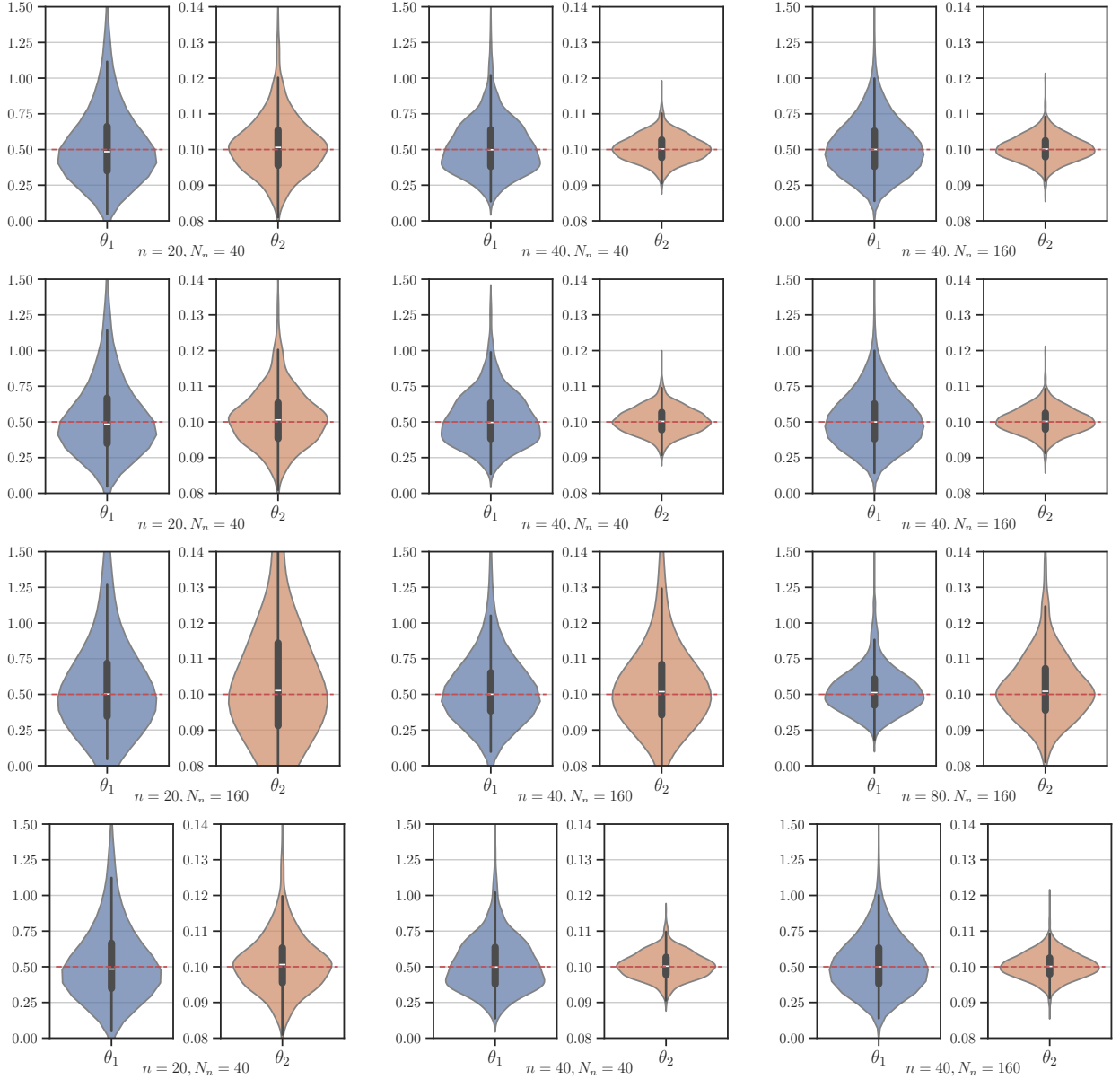
Figure 1: Violin plots of the estimated covariance parameters $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2)$ for Examples 6.1-6.4 (from top to bottom: sampling, random sampling with time-independent samples, random sampling with time-varying samples, Bernstein), and for different values of $n \geq 20$ and $N_n \geq 40$ (results shown by rows). The panel illustrates the estimation results (boxplot and KDE plot) of the variance $\widehat{\theta}_1$ and length-scale $\widehat{\theta}_2$ computed over 10^3 replicates. The ground truth θ_0 is represented by a horizontal red dashed line.

at $d = 500$ equispaced values of $x \in [0, 1]$. For random sampling with time-varying samples, due to the condition $\frac{N_n}{\log n} \to \infty$ discussed in Example 6.3, we fix $N_n = 160$ independently of the value of n. Finally, the covariance parameters $\theta = (\theta_1, \theta_2)$ are estimated via ML in a multi-start context, with five initial values randomly sampled from $[0.1, 10]^2$. To do so, the gradient-based modified Powell algorithm available in the Python toolbox SciPy [33] has been used. This entire procedure results in 10^3 estimations of θ_1 and θ_2, which will be used for assessing Theorems 5.2 and 5.4. For the latter, we focus on Corollary 5.6.

Figure 1 shows the violin plots illustrating both the boxplots and the kernel density estimate (KDE) plots for $\widehat{\theta}_1$ and $\widehat{\theta}_2$. We observe that, for all four cases, as n and N_n increase, the distributions of the estimators become closer to the Gaussian ones with median values closer to the ground truth
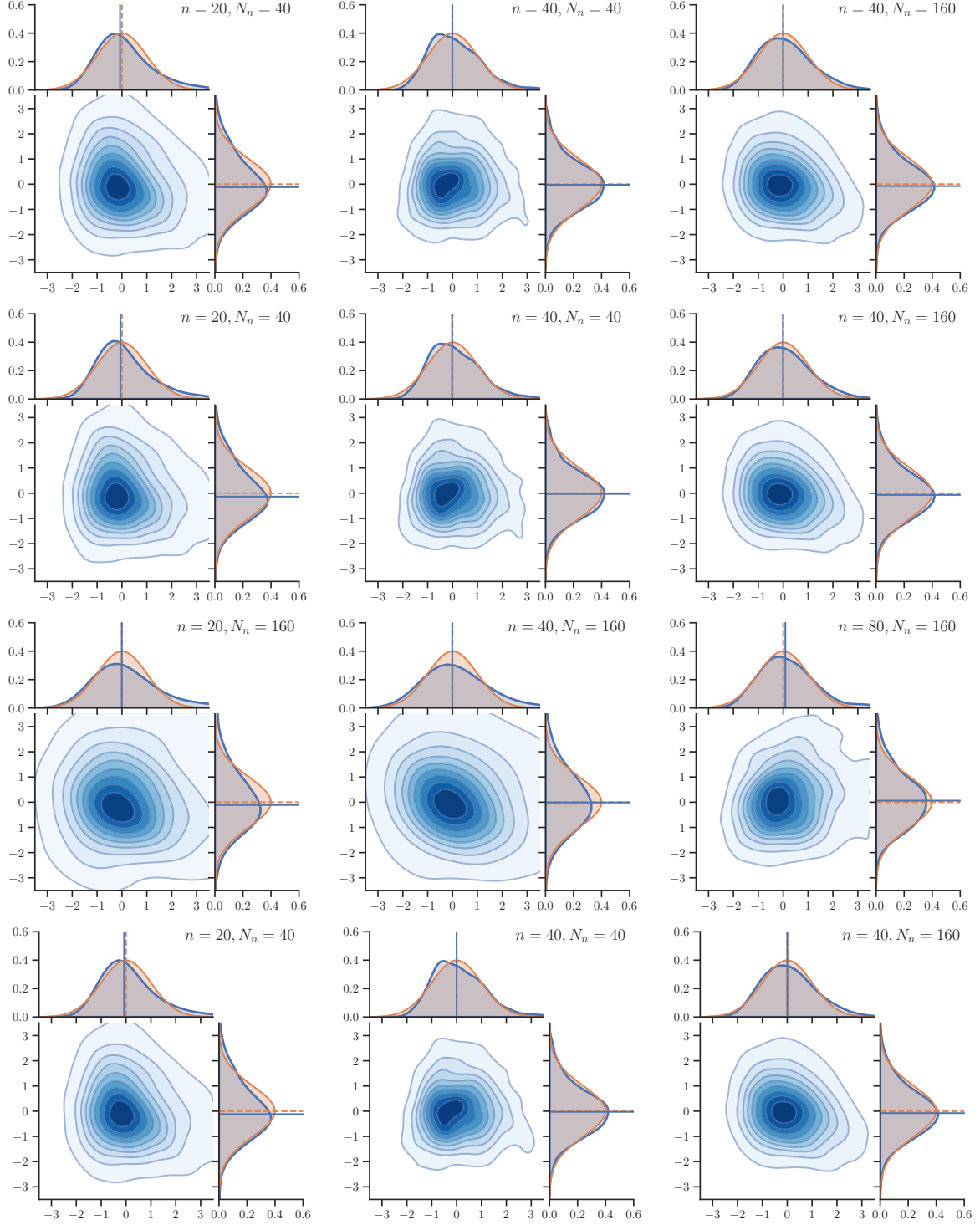
Figure 2: 2D KDE plots of $\beta = \sqrt{n} M_n^{1/2}(\widehat{\theta}_{n,N_n} - \theta_0) \in \mathbb{R}^2$ for Examples 6.1-6.4 (from top to bottom: sampling, random sampling with time-independent or time-varying samples, Bernstein), and for different values of $n \geq 20$ and $N_n \geq 40$ (results shown by rows). As shown in Corollary 5.6, $\beta \sim \mathcal{N}(0, I_2)$ as $n, N_n \to \infty$. For the case of random sampling with time-varying samples (third row), due to the condition $\frac{N_n}{\log n} \to \infty$, we let $N_n = 160$. The KDE (blue) and asymptotic (orange) marginal distributions of β are shown in the $x - y$ axes, with dashed lines indicating the median values.

$\theta_0 = (0.5, 0.1)$. It is noteworthy that, due to the complexity of the random sampling with time-varying samples, a larger value of n has been needed for more accurate median values with contracted distributions around the median. Overall, the results support the verification of Theorem 5.2 related to the asymptotic consistency of the ML estimator.

Figure 2 presents 2D KDE plots for $\beta = \sqrt{n} M_n^{1/2}(\widehat{\theta}_{n,N_n} - \theta_0) \in \mathbb{R}^2$. As shown in Corollary 5.6, the random vector β is expected to follow an asymptotic distribution $\mathcal{N}(0, I_2)$ as $n, N_n \to \infty$. The blue areas represent the 2D confidence ensembles at varying levels of confidence for β. Additionally, we display the KDE plots of the marginal distributions of β. We must clarify that these marginals do not correspond directly to the estimated parameters $\widehat{\theta}_1$ and $\widehat{\theta}_2$, but rather to centered linear combinations of them. Consistently with the results in Figure 1, we observe that, as n and N_n increase, the marginals become closer to the standard Gaussian distribution, thus supporting Theorem 5.4 and Corollary 5.6 related to the asymptotic normality of the ML estimator. Although the marginals indeed match the asymptotic distributions, this is not clear from the 2D KDE plots. We believe that this discrepancy is likely due to numerical issues stemming from the estimation method used by the 2D KDE function available in the Python toolbox Seaborn. We anticipate that considering tens of thousands of random replicates could yield better results, closer to the asymptotic distribution $\mathcal{N}(0, I_2)$.

7 Conclusions

To the best of our knowledge, we have introduced the first general convergence guarantees for the maximum likelihood (ML) estimator of a Gaussian process (GP) with functional inputs. Within the framework of increasing-domain asymptotics, we have shown both the consistency and normality of the estimator. Moreover, we extend these findings to encompass scenarios accounting for approximation errors in the inputs. These extensions offer valuable insights into certifying the robustness of practical implementations that rely on approximations obtained through conventional sampling methods or projections onto a fixed functional basis. Through several analytical and numerical examples, we illustrate the validity of the asymptotic guarantees.

A natural extension of this work includes convergence results within the fixed-domain asymptotic setting. This will require specific derivation on a case-by-case basis, tailored to particular covariance functions.

Appendix A: Conditions for isotropic covariance functions

This appendix provides the adaptions of Conditions 3.3, 3.7, 3.8, 5.2, 5.3 and 5.4 to the case of isotropic covariance functions.

Condition 7.1 *Suppose that the covariance structure given by* (1) *is both stationary and isotropic, that is we can write the kernels K_θ as*

$$\forall \theta \in \Theta, \forall f, g \in \mathcal{H}, \ K_\theta(f, g) = F_\theta(\|f - g\|),$$

with $F_\theta : \mathbb{R}^+ \to \mathbb{R}$. Furthermore, suppose that Conditions 3.1 and 3.2 hold, and that there exists a fixed $\gamma_0 > d_0$, such that for all $t \geq 0$, it holds

$$\sup_{\theta \in \Theta} |F_\theta(t)| \ \lhd \ \frac{1}{1 + t^{\gamma_0}}.$$

We can also rewrite Conditions 3.7 and 3.8 in the isotropic case as follows.

Condition 7.2 *Assume Conditions 3.1 and 3.2 holds. For all $t \geq 0$, $F_\theta(t)$ is continuously differentiable with respect to θ and, for some $\gamma_1 > d_0$, we have*

$$\sup_{\theta \in \Theta} \max_{i=1,\ldots,p} \left| \frac{\partial F_\theta}{\partial \theta_i}(t) \right| \ \lhd \ \frac{1}{1 + t^{\gamma_1}}.$$

Condition 7.3 *Assume Conditions 3.1 and 3.2 holds. For all $t \geq 0$, $\theta \mapsto F_\theta(t)$ is three times continuously differentiable and we have, for $q \in \{2, 3\}$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, and for some $\gamma_2 > d_0$,*

$$\sup_{\theta \in \Theta} \left| \frac{\partial^q F_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(t) \right| \ \lhd \ \frac{1}{1 + t^{\gamma_2}}.$$

The isotropic equivalents to Conditions 5.2, 5.3, and 5.4 are given below.

Condition 7.4 *Keeping the same constant γ_0 as in Condition 3.3, there exists another constant $\rho_0 > 0$ such that for all $t_1, t_2 \geq 0$ such that $0 < |t_1 - t_2| \leq \rho_0$, we have*

$$\sup_{\theta \in \Theta} \left| \frac{F_\theta(t_1) - F_\theta(t_2)}{t_1 - t_2} \right| \ \lhd \ \frac{1}{1 + \min(t_1, t_2)^{\gamma_0}}.$$

Condition 7.5 *Assume that Conditions 3.7 holds with the same constant $\gamma_1 > 0$. Then there exists another constant $\rho_1 > 0$ such that for all $i \in \{1, \ldots, p\}$, and for all $t_1, t_2 \geq 0$ such that $0 < |t_1 - t_2| \leq \rho_1$, we have*

$$\sup_{\theta \in \Theta} \frac{1}{|t_1 - t_2|} \left| \frac{\partial F_\theta}{\partial \theta_i}(t_1) - \frac{\partial F_\theta}{\partial \theta_i}(t_2) \right| \ \lhd \ \frac{1}{1 + \min(t_1, t_2)^{\gamma_1}}.$$

Condition 7.6 *Assume that Condition 3.8 holds with the same constant $\gamma_2 > 0$. Then there exists another constant $\rho_2 > 0$ such that for all $q \in \{2, 3\}$, for all $i_1, \ldots, i_q \in \{1, \ldots, p\}$, and for all $t_1, t_2 \geq 0$ such that $0 < |t_1 - t_2| \leq \rho_2$, we have*

$$\sup_{\theta \in \Theta} \frac{1}{|t_1 - t_2|} \left| \frac{\partial^q F_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(t_1) - \frac{\partial^q F_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(t_2) \right| \ \lhd \ \frac{1}{1 + \min(t_1, t_2)^{\gamma_2}}.$$

Appendix B: Proof of Lemmas in Section 4

Proof of Lemma 4.1. Let g, $(x_i)_{i \in \mathbb{N}^*}$ and n be like in the lemma. Then, through a partitioning argument, we get

$$\sum_{i=1}^{n} g(x_i) = \sum_{k=1}^{\infty} \sum_{i \in \{i \ : \ x_i \in [k-1,k)\}} g(x_i) \leq \sum_{k=1}^{\infty} \nu_k g(k-1).$$

The last inequality holds since g is non-increasing. ∎

Proof of Lemma 4.2. Let $n \in \mathbb{N}^*$. From Condition 3.3, we have

$$S_j = \sup_{\theta \in \Theta} \sum_{i=1}^{n} |K_\theta(f_j, f_i)| \vartriangleleft \sum_{i=1}^{n} \frac{1}{1 + \|f_j - f_i\|^{\gamma_0}},$$

for all $j \in \{1, \ldots, n\}$. From Condition 3.2 and Parseval's identity, we get that, for all $k, \ell \in \{1, \ldots, n\}$,

$$\|(f_{k,d_0} - f_{\ell,d_0}) - (f_k - f_\ell)\| \leq 2 \left(\sum_{m=d_0+1}^{\infty} s_m^2 \right)^{\frac{1}{2}} =: \delta. \tag{21}$$

However, we have the following inequality for all $x \geq 2\delta$

$$\frac{C}{1 + (x - \delta)^{\gamma_0}} \leq \frac{1}{1 + x^{\gamma_0}}, \tag{22}$$

with $C = \min \left(\frac{1 + \delta^{\gamma_0}}{1 + (2\delta)^{\gamma_0}}, \left(\frac{1}{2}\right)^{\gamma_0 - 1} \right) < 1$. Thus using this inequality as well as (21) and Lemma 3.1, we get that

$$S_j \ \vartriangleleft \ \sum_{i=1}^{n} \frac{1}{1 + \|f_{j,d_0} - f_{i,d_0}\|^{\gamma_0}}.$$

Applying Lemma 4.1, we get

$$\sum_{i=1}^{n} \frac{1}{1 + \|f_{j,d_0} - f_{i,d_0}\|^{\gamma_0}} \leq \sum_{k=1}^{\infty} \frac{\nu_k}{1 + (k-1)^{\gamma_0}}$$

$$\vartriangleleft \frac{1}{\Delta^{d_0}} \sum_{k=1}^{\infty} \frac{k^{d_0 - 1}}{1 + (k-1)^{\gamma_0}}.$$

with $\nu_k = \#\{i : \|f_{i,d_0} - f_{j,d_0}\| \in [k-1,k)\} \vartriangleleft \frac{k^{d_0 - 1}}{\Delta^{d_0}}$ according to Condition 3.1. Since $\gamma_0 > d_0$,

$$S_j \ \vartriangleleft \frac{1}{\Delta^{d_0}} \ \sum_{k=1}^{\infty} \frac{k^{d_0 - 1}}{1 + (k-1)^{\gamma_0}} < \infty.$$

∎

Proof of Lemma 4.3. Let $n \in \mathbb{N}^*$, then from Condition 3.7, we have

$$\widetilde{S}_j = \sup_{\theta \in \Theta} \sum_{i=1}^{n} \left| \frac{\partial K_\theta}{\partial \theta_k}(f_j, f_i) \right| \vartriangleleft \sum_{i=1}^{n} \frac{1}{1 + \|f_j - f_i\|^{\gamma_1}},$$

for all $j \in \{1, \ldots, n\}$. We can conclude by using similar arguments as in the proof of Lemma 4.2. ∎

Proof of Lemma 4.4. Using Lemma 4.2 as well as the Gerschgorin's circle theorem, we can show that

$$\sup_{\theta \in \Theta} \lambda_{\max}(R_\theta) \leq \sup_{\theta \in \Theta} \max_{j=1,\ldots,n} \sum_{i=1}^{n} |K_\theta(f_j, f_i)|$$

23

is bounded as $n \to \infty$. For the second part, it suffices to replace Lemma 4.2 by Lemma 4.3 and argue in the same manner as before to get that

$$\sup_{\theta \in \Theta} \max_{i=1,\dots,p} \lambda_{\max} \left(\frac{\partial R}{\partial \theta_i}(\theta) \right)$$

is bounded as $n \to \infty$. ■

Proof of Lemma 4.5. The function $\theta \mapsto \mathcal{L}_n(\theta)$ is C^1, note that

$$M := \sup_{\theta \in \Theta} \left| \frac{\partial \left(\mathcal{L}_n - \mathbb{E}\left[\mathcal{L}_n(\cdot) \right] \right)}{\partial \theta_i}(\theta) \right| = O_P(1).$$

Let $\epsilon > 0$ and consider $\widetilde{\theta}_1, \dots, \widetilde{\theta}_k \in \Theta$ such that all $(B(\widetilde{\theta}_i, \epsilon))_i$ constitute an open cover. Let $\theta \in \Theta$ and denote by $\widetilde{\theta}_{i_0}$ the closest $\widetilde{\theta}_i$ from θ. Then, by the mean value inequality, we have

$$\left| \mathcal{L}_n(\theta) - \mathbb{E}\left[\mathcal{L}_n(\theta) \right] - \mathcal{L}_n(\widetilde{\theta}_{i_0}) + \mathbb{E}\left[\mathcal{L}_n(\widetilde{\theta}_{i_0}) \right] \right| \leq 2M \left| \theta - \widetilde{\theta}_{i_0} \right|.$$

Thus

$$\left| \mathcal{L}_n(\theta) - \mathbb{E}\left[\mathcal{L}_n(\theta) \right] \right| \leq 2M \left| \theta - \widetilde{\theta}_{i_0} \right| + \left| \mathcal{L}_n(\widetilde{\theta}_j) - \mathbb{E}\left[\mathcal{L}_n(\widetilde{\theta}_{i_0}) \right] \right|.$$

However $|\mathcal{L}_n(\widetilde{\theta}_{i_0}) - \mathbb{E}[\mathcal{L}_n(\widetilde{\theta}_{i_0})]| = o_P(1)$ as well as $|\theta - \widetilde{\theta}_{i_0}| \leq \epsilon$ with ϵ arbitrarily chosen. This concludes the proof. ■

Proof of Lemma 4.7. Let $n \in \mathbb{N}^*$, $q \in \{2,3\}$ and $i_1, \dots, i_q \in \{1, \dots, p\}$, then from Condition 3.8, we have

$$\sup_{\theta \in \Theta} \sum_{i=1}^{n} \left| \frac{\partial^q K_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(f_j, f_i) \right| \vartriangleleft \sum_{i=1}^{n} \frac{1}{1 + \| f_j - f_i \|^{\gamma_2}},$$

for all $j \in \{1, \dots, n\}$. We can conclude by using similar arguments as in the proof of Lemma 4.2. ■

Proof of Lemma 4.8. As in the proof of Lemma 4.4, we obtain this result through the appropriate use of Gershgorin's circle theorem as well as Lemma 4.7. ■

Appendix C: Proof of Lemmas in Section 5

Proof of Lemma 5.1. According to Gerschgorin's circle theorem, we have for any $\theta \in \Theta$

$$\lambda_{\max}(R_\theta - R_{N_n,\theta}) \leq \max_{i=1,\ldots,n} \sum_{j=1}^{n} |(R_\theta)_{i,j} - (R_{N_n,\theta})_{i,j}|$$

$$= \max_{i=1,\ldots,n} \sum_{j=1}^{n} \left| K_\theta(f_i, f_j) - K_\theta(f_i^{N_n}, f_j^{N_n}) \right|.$$

Let $b_n = \max_{i=1}^{n} \|f_i - f_i^{N_n}\|$. Then

$$\|(f_i - f_j) - (f_i^{N_n} - f_j^{N_n})\| \leq \|f_i - f_i^{N_n}\| + \|f_j - f_j^{N_n}\| \leq 2b_n.$$

Since $\lim_{n\to\infty} b_n = 0$ from Condition 5.1, from Condition 5.2, as well as the inequality in (22), for n large enough

$$\sup_{\theta \in \Theta} \lambda_{\max}(R_\theta - R_{N_n,\theta}) \triangleleft b_n \max_{i=1,\ldots,n} \sum_{j=1}^{n} \frac{1}{1 + \min(\|f_i - f_j\|, \|f_i^{N_n} - f_j^{N_n}\|)^{\gamma_0}}$$

$$\triangleleft b_n \max_{i=1,\ldots,n} \sum_{j=1}^{n} \frac{1}{1 + \|f_i - f_j\|^{\gamma_0}}.$$

We recall that $\|f_{j,d_0} - f_{i,d_0}\| \geq \Delta$ and $\frac{1}{1+|f_i-f_j\|^{\gamma_0}} \triangleleft \frac{1}{1+\|f_{j,d_0}-f_{i,d_0}\|^{\gamma_0}}$. Therefore, applying Lemma 4.1, we get

$$\sup_{\theta \in \Theta} \lambda_{\max}(R_\theta - R_{N_n,\theta}) \triangleleft b_n \max_{i=1,\ldots,n} \sum_{j=1}^{n} \frac{1}{1 + \|f_i - f_j\|^{\gamma_0}}$$

$$\triangleleft \frac{b_n}{\Delta^{d_0}} \sum_{k=1}^{\infty} \frac{k^{d_0-1}}{1 + (k-1)^{\gamma_0}}.$$

Since $\lim_{n\to\infty} b_n = 0$, then the first part of the lemma is proven. The second part can be established using similar arguments. ∎

Proof of Lemma 5.5. Let $i, j \in \{1, \ldots p\}$ and $n \in \mathbb{N}^*$. Remark that all the following matrices are symmetric.

$$A := R_{\theta_0}^{-1}, \qquad B := \frac{\partial R}{\partial \theta_i}(\theta_0), \qquad C := \frac{\partial R}{\partial \theta_j}(\theta_0),$$

$$A_n := R_{N_n,\theta_0}^{-1}, \qquad B_n := \frac{\partial R_{N_n}}{\partial \theta_i}(\theta_0), \qquad C_n := \frac{\partial R_{N_n}}{\partial \theta_j}(\theta_0).$$

Using the Cauchy-Schwarz inequality on the space of square matrices endowed with the norm $\|\cdot\|_2$, where we recall that $\|R\|_2^2 = \frac{1}{n}\sum_{i,j=1}^{n} R_{i,j}^2$, we obtain

$$|(M_{n,N_n} - M_n)_{i,j}| = \frac{1}{2n} \left| \mathrm{tr}\left(R_{N_n,\theta_0}^{-1} \frac{\partial R_{N_n}}{\partial \theta_i}(\theta_0) R_{N_n,\theta_0}^{-1} \frac{\partial R_{N_n}}{\partial \theta_j}(\theta_0) \right) \right.$$

$$\left. - \mathrm{tr}\left(R_{\theta_0}^{-1} \frac{\partial R}{\partial \theta_i}(\theta_0) R_{\theta_0}^{-1} \frac{\partial R}{\partial \theta_j}((\theta_0)) \right) \right|$$

$$= \frac{1}{2n} \left| \mathrm{tr}(A_n B_n A_n C_n) - \mathrm{tr}(ABAC) \right|$$

$$= \frac{1}{2n} \left| \mathrm{tr}\left((A_n - A)B_n A_n C_n + A(B_n - B)A_n C_n \right. \right.$$

$$\left. \left. + AB(A_n - A)C_n + ABA(C_n - C) \right) \right|$$

$$\leq \|A_n - A\|_2 \|B_n A_n C_n\|_2 + \|B_n - B\|_2 \|A_n C_n A\|_2$$

$$+ \|A_n - A\|_2 \|C_n AB\|_2 + \|C_n - C\|_2 \|ABA\|_2.$$

However, we have
$$\|A_n - A\|_2 = \|R_{N_n,\theta}^{-1} - R_\theta^{-1}\|_2 \xrightarrow[n\to\infty]{} 0,$$
as well as
$$\|B_n - B\|_2 \xrightarrow[n\to\infty]{} 0 \quad \text{and} \quad \|C_n - C\|_2 \xrightarrow[n\to\infty]{} 0.$$
Moreover the quantities $\|B_n A_n C_n\|_2$, $\|A_n C_n A\|_2$, $\|C_n A B\|_2$ and $\|ABA\|_2$ are bounded in n. Thus
$$M_{n,N_n} - M_n \xrightarrow[n\to\infty]{} 0.$$

∎

References

[1] François Bachoc. Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of Multivariate Analysis*, 125:1–35, 2014.

[2] François Bachoc, Louis Béthune, Alberto Gonzalez-Sanz, and Jean-Michel Loubes. Gaussian processes on distributions based on regularized optimal transport. In *International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 4986–5010, 25–27 Apr 2023.

[3] François Bachoc, Fabrice Gamboa, Jean-Michel Loubes, and Nil Venet. A Gaussian process regression model for distribution inputs. *IEEE Transactions on Information Theory*, 64(10):6620–6637, 2018.

[4] François Bachoc. *Asymptotic analysis of maximum likelihood estimation of covariance parameters for Gaussian processes: An introduction with proofs*, pages 283–303. Springer International Publishing, Cham, 2021.

[5] François Bachoc, Baptiste Broto, Fabrice Gamboa, and Jean-Michel Loubes. Gaussian field on the symmetric group: Prediction and learning. *Electronic Journal of Statistics*, 14(1):503–546, 2020.

[6] François Bachoc, Alexandra Suvorikova, David Ginsbourger, Jean-Michel Loubes, and Vladimir Spokoiny. Gaussian processes with multidimensional distribution inputs via optimal transport and Hilbertian embedding. *Electronic Journal of Statistics*, 14(2):2742–2772, 2020.

[7] François Bachoc. Cross validation and maximum likelihood estimations of hyper-parameters of gaussian processes with model misspecification. *Computational Statistics & Data Analysis*, 66:55–69, 2013.

[8] Bachoc, François, Contal, Emile, Maatouk, Hassan, and Rullière, Didier. Gaussian processes for computer experiments. *ESAIM: Procs*, 60:163–179, 2017.

[9] José Betancourt, François Bachoc, Thierry Klein, Déborah Idier, Rodrigo Pedreros, and Jérémy Rohmer. Gaussian process metamodeling of functional-input code for coastal flood hazard assessment. *Reliability Engineering & System Safety*, 198:106870, 2020.

[10] Mickaël Binois and Nathan Wycoff. A survey on high-dimensional gaussian process modeling with application to bayesian optimization. *ACM Transactions on Evolutionary Learning and Optimization*, 2(2), aug 2022.

[11] Viacheslav Borovitskiy, Iskander Azangulov, Alexander Terenin, Peter Mostowsky, Marc Deisenroth, and Nicolas Durrande. Matérn gaussian processes on graphs. In *International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 2593–2601, 13–15 Apr 2021.

[12] Gustau Camps-Valls, Jochem Verrelst, Jordi Munoz-Mari, Valero Laparra, Fernando Mateo-Jimenez, and Jose Gomez-Dans. A survey on gaussian processes for earth-observation data analysis: A comprehensive investigation. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):58–78, 2016.

[13] A. Forrester, A. Sóbester, and A. Keane. *Engineering Design via Surrogate Modelling: A Practical Guide*. Wiley, 2008.

[14] Reinhard Furrer and Michael Hediger. Asymptotic analysis of ml-covariance parameter estimators based on covariance approximations. *Electronic Journal of Statistics*, 17(2):3050–3102, 2023.

[15] Eduardo C. Garrido-Merchán and Daniel Hernández-Lobato. Dealing with categorical and integer-valued variables in bayesian optimization with gaussian processes. *Neurocomputing*, 380:20–35, 2020.

[16] Marc G. Genton. Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research*, 2:299–312, 2001.

[17] C.A Hall. On error bounds for spline interpolation. *Journal of Approximation Theory*, 1(2):209–218, 1968.

[18] Hemachandran K, S. Tayal, P.M. George, P. Singla, and U. Kose. *Bayesian Reasoning and Gaussian Processes for Machine Learning Applications*. CRC Press, 2022.

[19] C. G. Kaufman and B. A. Shaby. The role of the range parameter for estimation and prediction in geostatistics. *Biometrika*, 100:473–484, 2013.

[20] Chi-Kwongn Li and Fuzhen Zhang. Eigenvalue continuity and gersgorin's theorem. *Electronic Journal of Linear Algebra*, 35:619–625, 2019.

[21] Andrés F. López-Lopera, D. Idier, J. Rohmer, and F. Bachoc. Multioutput Gaussian processes with functional data: A study on coastal flood hazard assessment. *Reliability Engineering & System Safety*, 218, 2022.

[22] P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability*, 18(3):1269–1283, 1990.

[23] T. Muehlenstaedt, J. Fruth, and O. Roustant. Computer experiments with functional inputs and scalar outputs by a norm-based approach. *Statistics and Computing*, 27(2):1083–1097, 2017.

[24] Simon Nanty, Céline Helbert, Amandine Marrel, Nadia Pérot, and Clémentine Prieur. Sampling, metamodeling, and sensitivity analysis of numerical simulators with functional stochastic inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):636–659, 2016.

[25] Christopher J. Paciorek and Mark J. Schervish. Nonstationary covariance functions for Gaussian process regression. In *Advances in Neural Information Processing Systems*, volume 16, pages 273–280. 2004.

[26] James Ramsay and Bernard W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer, 2005.

[27] Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, Cambridge, MA, 2005.

[28] J.B. Rosen. Minimum error bounds for multidimensional spline approximation. *Journal of Computer and System Sciences*, 5(4):430–452, 1971.

[29] Olivier Roustant, Espéran Padonou, Yves Deville, Aloïs Clément, Guillaume Perrin, Jean Giorla, and Henry Wynn. Group kernels for gaussian process metamodels with categorical inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 8(2):775–806, 2020.

[30] M. Schatzman. *Numerical Analysis: A Mathematical Introduction*. Clarendon Press, 2002.

[31] Michael L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer, 1999.

[32] Zoltán Szabó, Bharath K. Sriperumbudur, Barnabás Póczos, and Arthur Gretton. Learning theory for distribution regression. *Journal of Machine Learning Research*, 17(152):1–40, 2016.

[33] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero,

Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[34] Michael L. Waskom. Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.

[35] H. Zhang. Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99:250–261, 2004.

[36] Yin-Cong Zhi, Yin Cheng Ng, and Xiaowen Dong. Gaussian processes on graphs via spectral kernel learning. *IEEE Transactions on Signal and Information Processing over Networks*, 9:304–314, 2023.