Minimax estimation of Functional Principal Components from noisy discretized functional data

Ryad Belhakem *[1], Franck Picard †[2], Vincent Rivoirard ‡[3], and Angelina Roche §[4]

[1,3,4]CEREMADE, CNRS, Université Paris-Dauphine, Université PSL, 75016 PARIS, FRANCE
[2]Centre National de la Recherche Scientifique, Laboratoire de Biologie et Modélisation de la Cellule, 46, allée d'Italie 69007 Lyon

December 13, 2023

Abstract

Functional Principal Component Analysis is a reference method for dimension reduction of curve data. Its theoretical properties are now well understood in the simplified case where the sample curves are fully observed without noise. However, functional data are noisy and necessarily observed on a finite discretization grid. Common practice consists in smoothing the data and then to compute the functional estimates, but the impact of this denoising step on the procedure's statistical performance are rarely considered. Here we prove new convergence rates for functional principal component estimators. We introduce a double asymptotic framework: one corresponding to the sampling size and a second to the size of the grid. We prove that estimates based on projection onto histograms show optimal rates in a minimax sense. Theoretical results are illustrated on simulated data and the method is applied to the visualization of genomic data.

Keywords— Functional data analysis, Principal Components Analysis, minimax rates.

1 Introduction

Functional Data Analysis (FDA) is a statistical framework dedicated to curve data that are supposed to be the realizations of random functions [Ferraty and Vieu, 2006, Ramsay and Silverman,

*belhakem@ceremade.dauphine.fr
†franck.picard@ens-lyon.fr
‡rivoirard@ceremade.dauphine.fr
§roche@ceremade.dauphine.fr

2010, Ferraty and Romain, 2011]. Hence, in this framework, the infinite-dimensional nature of the process that generated the data is central to develop efficient estimation procedures. Functional Principal Components Analysis (fPCA) is a common method to reduce dimensionality of curve data and has been considered either as an exploration tool [Ramsay and Silverman, 2010] or as a pre-processing step for many statistical procedures [Kalogridis and Van Aelst, 2019, Goode et al., 2020, Perrin et al., 2021, Song et al., 2020, Seo, 2020, Jaimungal and Ng, 2007]. In most theoretical works (see e.g. Cardot and Johannes [2010], Cai and Yuan [2012], Mas and Ruymgaart [2015]) it is assumed that the functional data $Z_i(t)$ is observed for individual i at all points t in an interval (e.g. $[0,1]$). However, in practice, the Z_i's are observed on a finite grid $t_{i,0},\ldots,t_{i,p_i-1}$ and can be corrupted by noise. Few theoretical works in the literature of FDA study the effect of the sampling scheme on the performance of the estimation procedures.

Here, we consider the case where data are observed on a fixed and regular design (fixed grid), and we focus on the estimation of the elements of the functional principal components, from a non-asymptotic point of view. The case of a fixed regular grid actually corresponds to a large number of applications in FDA, for instance electricity consumption curves [Devijver et al., 2020], temperature or precipitation curves [Ramsay and Silverman, 2010], or spectrometric datasets [Pham et al., 2010]) to name a few. More precisely, we observe

$$\{Y_i(t_h), i=1,\ldots,n, h=0,\ldots,p-1\}, \tag{1}$$

generated from the following statistical model

$$Y_i(t_h) = Z_i(t_h) + \varepsilon_{i,h}, \qquad i=1,\ldots,n, \tag{2}$$

where $\{\varepsilon_{i,h}\}_{i=1,\ldots,n;h=0,\ldots,p-1}$ is an i.i.d. sequence of centered Gaussian errors with variance σ^2, the Z_i's are i.i.d. random elements of the space of continuous functions on the interval $[0,1]$ and $t_h = h/(p-1)$, $h=0,\ldots,p-1$. We also suppose that the $\varepsilon_{i,h}$'s are independent from the Z_i's.

Two statistical frameworks are often considered: longitudinal data analysis (LDA) or FDA and it seems important to clarify at this stage the links and differences between these two approaches. In LDA, data are often observed at random, hence the sampling points are random and depend on each individual. From a theoretical point of view, the number of sampling points can be supposed bounded, and results concerning convergence rates only involve the number of individuals n. There have been a lot of methodological and theoretical works in the case where the observations are observed on a random grid with a small number of observations per subject (sparse longitudinal data), we refer to Yao et al. [2005], Hall et al. [2006], Dai et al. [2018], Zhong et al. [2022] and references therein. In this paper, we consider the case where the number of sampling points is usually considered to be large and shared by all curves, which is frequently the case in FDA and motivates the characterization of a double asymptotic in n and p.

At first sight, inference for fPCA and classical PCA on the $n \times p$ matrix of observations is comparable. Following this idea, when $p \gg n$, functional PCA would be confronted to inconsistency problems as in standard multivariate PCA (see Johnstone and Lu 2009 and references therein). However, this is counter-intuitive in the functional framework: when p increases, more and more information is recorded on the underlying process, which should improve statistical performance. Indeed, in the continuous non-noisy case, where $Z_i(t)$ is observed at all points t, corresponding to $p=+\infty$ and $\sigma=0$, it is known from the works of Dauxois et al. [1982] that the estimation of the principal components is consistent. Moreover, an optimal n^{-1} parametric rate (up to a logarithmic factor) can be achieved for the risk associated to the $\mathbb{L}^2$-error [Bosq, 2000, Theorem 4.5, p. 106] or to the operator norm of the projector [Mas and Ruymgaart, 2015]. Considering our data as multivariate data would resume to ignore the underlying regularity of the processes Z_i and the fact that, when p is large, $Z_i(t_h)$ is close to $Z_i(t_{h+1})$. Then, specific attention should be paid.

From a theoretical perspective, the main challenge is to assess the rates of convergence of the estimators of the elements of the principal components basis, in a very specific framework. FDA combines two very different convergence settings: a first one associated with the sampling of n independent processes $Z_1, \ldots, Z_n$, and a non-parametric setting since data are functions, here observed at p points. Hall et al. [2006] investigate the estimation of the elements of the fPCA basis in the case where the number of discretization points by individual is bounded and the grid is random (i.e. $p_i \leq p$ with p fixed and $n \to +\infty$). They obtain non-parametric rates for a kernel smoothing estimator which are optimal under the assumption that the function to estimate is exactly two-times differentiable. However, the authors themselves point out that their results are no longer valid in the case of a fixed regular grid, where consistent estimation is not possible when p is fixed. In the context of the estimation of the mean function of a functional data sample, Cai and Yuan [2011] found that the optimal rates of convergence are completely different if we consider a fixed grid or a random grid. It appears that, in the case of a fixed grid, the minimax estimation rates of the function principal components basis remain unknown. Hence we propose to investigate the joint impact of noise and discretization (sampling scheme) on the estimation of the eigenelements of the covariance operator.

In addition, a common practice is to first smooth the data, usually by projecting it into a splines basis with a roughness penalty or via kernel smoothing (see e.g. Ramsay and Silverman 2010). However, the statistical implications of the smoothing step are rarely debated, whereas it raises some concern, mainly related to the level of regularity of the underlying process versus the choice of the smoothing basis, and the capacity of distinguishing noise from signal through this method. [Cai and Yuan, 2011, Section 2.2, p. 2336] point out that, in the case of mean estimation, there is no benefit from smoothing in terms of convergence rates when the observation grid is fixed. This implies that usual splines or kernel smoothing step can lead to suboptimal estimators if the smoothing parameter is not well chosen (see Theorem 2.2 of Cai and Yuan [2011]).

To fully understand the statistical complexity of functional principal component analysis, it is necessary to compute the minimax rates of estimation and to compare it with the parametric bounds obtained by Mas and Ruymgaart [2015]. Upper bounds in the case of noisy discretized data have also been proposed [Bunea and Xiao, 2015, Descary and Panaretos, 2019]. Bunea and Xiao [2015] established results under strong conditions of the eigenvalues of the operator. Descary and Panaretos [2019] studied a generalization to heterogeneous noise with possible time dependency at the price of two strong assumptions: analyticity of the eigenfunctions and finite rank of the covariance operator of the signal; the achieved rate is then $n^{-1} + p^{-2}$.

Here we study convergence rates for the estimation of the eigenelements of the covariance operator Γ under a mild regularity assumption on the process Z. Denoting by α this regularity, our assumption is equivalent to assuming that the kernel K is a bivariate α-Hölder continuous function. Under a moment assumption for process Z, we obtain rates of the form

$$n^{-1} + p^{-2\alpha}.$$

These rates, which are new, are, moreover, optimal in the minimax sense for the estimation of the first eigenfunction (we prove a lower bound). We illustrate these rates in practice on simulations.

These rates tell us a lot about the behavior of the estimated eigenfunctions under the double asymptotic in p and n. When p is large compared to $n^{1/(2\alpha)}$, we find the optimal parametric rate n^{-1} obtained by Dauxois et al. [1982], Mas and Ruymgaart [2015] when the curves Z_i are fully observed and without noise. Moreover, even though the problem is intrinsically non-parametric, and in the presence of noisy observations, the simple estimator obtained by projection on the p-bins histogram system reaches the optimal minimax rate. Therefore, we do not need regularization, which may be counter-intuitive. The knowledge of α is also not necessary. These results are confirmed by the simulation study we have conducted which also suggests that the same conclusion applies for the estimation of the eigenvalues for which we do not know so far whether the rate obtained is minimax optimal. Lastly, at the end of Section 3.4, we discuss our results and compare them with the most recent ones. Our results show that the underlying regularity of the data plays a special role in the theoretical developments, with important implications in practice. In our setting with equispaced deterministic observations, the sampling scheme cannot be too sparse to obtain parametric rates and even consistency. This is a main difference with the random sampling scheme considered by [Hall et al., 2006]. We complete our theoretical and empirical study by two original applications of functional principal component analysis, on single-cell expression data analysis to characterize the immune response to viral infection, and to genomic data for the characterization of replication origins along the human genome with respect to the spatial distribution of particular sequence motifs called G-quadruplexes [Zheng et al., 2020].

Notations: We denote $\|\cdot\|$ the $\mathbb{L}^2$-norm associated with the scalar product $\langle\cdot,\cdot\rangle$ and $\|\cdot\|_{\ell_2}$ the ℓ_2-norm for a vector. For any continuous operator T on $\mathbb{L}_2[0,1]$, we denote $|||T|||$ the operator norm of T associated to $\|\cdot\|$ and defined by $|||T||| = \sup_{f\in\mathbb{L}_2[0,1],\|f\|=1}\|Tf\|$. For P a probability measure, we denote E the associated expectation. We denote P_Z the distribution of the process Z and E_Z the associated expectation. The set of continuous functions on $[0,1]$ is denoted $\mathcal{C}^0$. We adopt the following notation: for two sequences $a = (a_{n,p})_{n,p\geq 1}$, $b = (b_{n,p})_{n,p\geq 1}$ of real fixed quantities or random variables, we denote $a \lesssim b$ if there exists a universal constant c such that $a_{n,p} \leq c b_{n,p}$ a.s. for all $n, p \geq 1$. We define $\text{sign}(u) = \mathbf{1}_{\{u\geq 0\}} - \mathbf{1}_{\{u<0\}}$ for any $u \in \mathbb{R}$.

2 Functional PCA for discretely observed random functions

We suppose in the following that $Z_1, \ldots, Z_n$ is a sample of independent centered continuous functional variables with same distribution as a process Z.

The covariance operator Γ defined by

$$\Gamma(f)(\cdot) = \mathbb{E}(\langle f, Z\rangle Z(\cdot)), \quad f \in \mathbb{L}^2,$$

is well defined provided $E(\|Z\|^2) < +\infty$, which is assumed in the following.

Defining the eigenfunctions $\eta^* = \{\eta_d^*, d \in \mathbb{N}^*\}$ and the associated eigenvalues, assumed to be distinct, $\mu^* = \{\mu_d^*, d \in \mathbb{N}^* \,;\, \mu_1^* > \mu_2^* > \ldots\}$ of the operator Γ, we obtain the Karhunen-Loève representation of Z [Bosq, 2000]:

$$Z = \sum_{d\in\mathbb{N}^*} \zeta_d^* \mu_d^{*1/2} \eta_d^*, \tag{3}$$

where $\zeta^* = \{\zeta_d^*, d \in \mathbb{N}^*\}$ is a sequence of non-correlated centered random variables of variance 1, usually called the principal components scores and the family $(\eta_j^*)_{j\geq 1}$ is an orthonormal basis of $\mathbb{L}^2$. Then, for any integer $\mathcal{D}$, the best $\mathcal{D}$-dimensional approximation of process Z is spanned by the first $\mathcal{D}$ eigenfunctions. Our aim is to provide estimators of these eigenelements based on noisy discretized data, and to assess their statistical performance.

In [Dauxois et al., 1982, Bosq, 2000, Ramsay and Silverman, 2010, Hall et al., 2006], the $Z_i(t)$'s are observed for all $t \in [0,1]$ without noise and the estimator of η_d^* is the eigenfunction $\widehat{\eta}_d$ associated to the d-th largest eigenvalue of the empirical covariance operator

$$\widehat{\Gamma}(f)(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \langle f, Z_i\rangle Z_i(\cdot), \quad f \in \mathbb{L}^2.$$

However, when process Z is observed on a grid, the empirical covariance operator $\widehat{\Gamma}$ can not be calculated and must be approximated.

Then, in the setting of Model (2), we first reconstruct the observed curves on the entire interval and we define, for $i = 1, \dots, n$,

$$\widetilde{Y}_i(t) = \sum_{\lambda \in \Lambda_D} \widetilde{y}_{i,\lambda} \phi_\lambda(t), \quad \widetilde{y}_{i,\lambda} = \frac{1}{p} \sum_{h=0}^{p-1} Y_i(t_h) \phi_\lambda(t_h), \quad t \in [0,1],$$

where $\{\phi_\lambda, \lambda \in \Lambda_D\}$ is an orthonormal system of $\mathbb{L}^2([0,1])$ of cardinality $D \geq 1$ and $\widetilde{y}_{i,\lambda}$ an approximation of $\langle Y_i, \phi_\lambda \rangle$. Similarly, we define $\widetilde{Z}_i(t)$, $\widetilde{z}_{i,\lambda}$, $\widetilde{E}_i(t)$, $\widetilde{\varepsilon}_{i,\lambda}$ by replacing $Y_i(t_h)$ in the previous expressions by $Z_i(t_h)$, and $\varepsilon_{i,h}$.

A natural estimator of the covariance operator is then

$$\widehat{\Gamma}_\phi(f)(\cdot) = \frac{1}{n} \sum_{i=1}^{n} \langle f, \widetilde{Y}_i \rangle \widetilde{Y}_i(\cdot), \quad f \in \mathbb{L}^2.$$

It is easily seen from the definition above that the operator $\widehat{\Gamma}_\phi$ is self-adjoint. It is also finite-rank hence compact. Then by the diagonalization theorem for self-adjoint compact operators [see Brezis, 2011, Theorem 6.11, p. 167], there exists a basis $(\widehat{\eta}_{\phi,d})_{d \geq 1}$ of $\mathbb{L}^2$ made of eigenfunctions of $\widehat{\Gamma}_\phi$. In the following, we study the $\mathbb{L}^2$-risk of the estimator $\widehat{\eta}_{\phi,d}$ for $d = 1, \dots, \mathcal{D}$.

3 Minimax rates of the eigenfunction estimator

3.1 Smoothness class for the functional curve Z

Minimax rates of convergence depend on the underlying smoothness of the process of interest. In the sequel, for any $\alpha \in (0,1]$ and $L > 0$ we consider the regularity class

$$\mathcal{R}_\alpha(L) = \Big\{P, \text{ probability measure on } \mathcal{C}^0 \text{ such that } \int_{\mathcal{C}^0} \{z(t) - z(s)\}^2 dP(z) \leq L|t-s|^{2\alpha}, \quad (s,t) \in [0,1]^2 \Big\}.$$

This regularity set is natural. Indeed, we can for instance remark that P_Z, the distribution of Z, satisfies

$$P_Z \in \mathcal{R}_\alpha(L) \Leftrightarrow E_Z[\{Z(t) - Z(s)\}^2] \leq L|t-s|^{2\alpha}, \quad (s,t) \in [0,1]^2.$$

This condition can be seen as a regularity assumption on the covariance kernel

$$K(s,t) = E\big\{Z(s)Z(t)\big\}, \quad (s,t) \in [0,1]^2.$$

Indeed, our regularity condition $E(\|Z\|^2) < +\infty$ combined with the condition $P_Z \in \mathcal{R}_\alpha(L)$ imply that kernel K is bounded

$$\|K\|_\infty = \sup_{(s,t) \in [0,1]^2} |K(s,t)| < \infty,$$

and is an α-Hölder continuous function. More precisely, for any $(s, s', t, t') \in [0,1]^4$,

$$P_Z \in \mathcal{R}_\alpha(L) \Rightarrow |K(s,t) - K(s',t')| \leq (\|K\|_\infty L)^{1/2} \left(|s-s'|^\alpha + |t-t'|^\alpha\right). \quad (4)$$

Conversely, if K is a bivariate α-Hölder continuous function, we know that there exists $L' > 0$ such that

$$|K(s,t) - K(s',t')| \leq L' \left(|s-s'|^2 + |t-t'|^2\right)^{\alpha/2}.$$

Then

$$E_Z[\{Z(t) - Z(s)\}^2] = K(s,s) - 2K(s,t) + K(t,t) \leq 2L'|s-t|^\alpha,$$

and $P_Z \in \mathcal{R}_\alpha(2L')$.

Classical Gaussian processes belong to $\mathcal{R}_\alpha(L)$ for α and L well chosen. For instance, if Z is a standard Brownian motion or a Brownian bridge then $P_Z \in \mathcal{R}_{1/2}(1)$. More generally, fractional Brownian motions with Hurst exponent α and Hurst index C_α belong to $\mathcal{R}_\alpha(C_\alpha)$. If Z is an Ornstein-Uhlenbeck process, its covariance function is $K(s,t) = \exp(-|t-s|/2)$, then it verifies

$$E_Z[\{Z(t) - Z(s)\}^2] = 2\big(1 - e^{-|t-s|/2}\big) \leq |t-s|, \quad (s,t) \in \mathbb{R}^2,$$

which implies $P_Z \in \mathcal{R}_{1/2}(1)$. We refer to Lifshits [1995] for the precise definitions and properties of these processes.

Remark 1 *We also remark that L depends on the eigenvalues sequence $(\mu_d^*)_{d\geq 1}$. Indeed, suppose e.g. that, for all $d \geq 1$, the eigenfunction η_d^* is α-holdérian (i.e. there exists $L_d > 0$ such that $|\eta_d^*(t) - \eta_d^*(s)| \leq L_d|t-s|^\alpha$, for all $t, s \in [0,1]$), then, if $\sum_{d\geq 1} \mu_d^* L_d^2 < +\infty$, from the Karhunen-Loève decomposition* (3)*,we can write, since the ζ_d^*'s are centered and uncorrelated,*

$$\begin{aligned} E_Z[\{Z(t) - Z(s)\}^2] &= E_Z\left[\Big\{\sum_{d\geq 1} \zeta_d^* \mu_d^{*1/2}(\eta_d^*(t) - \eta_d^*(s))\Big\}^2\right] \\ &= \sum_{d\geq 1} \mu_d^*(\eta_d^*(t) - \eta_d^*(s))^2 \\ &\leq \sum_{d\geq 1} \mu_d^* L_d^2 |t-s|^{2\alpha}. \end{aligned}$$

Then $Z \in \mathcal{R}_\alpha(L)$ with $L = \sum_{d\geq 1} L_d^2 \mu_d^*$.

3.2 Lower bound

The lower bound of the risk for estimating eigenfunctions can be viewed as a benchmark to achieve. We focus on the first eigenfunction, but a similar result, though more technical, could be obtained for the other eigenfunctions. However, since the estimation of higher order eigenfunctions is a more complex statistical problem, it seems intuitively reasonable to us that the lower bound on these eigenfunctions is (at worst) of the same order.

Theorem 1 *Let $\alpha \in (0,1]$ and $L > 0$. Assume that the rank of the covariance operator Γ is larger than 2. Then, for any $n \geq 1$ and $p \geq 1$, we have:*

$$\inf_{\widehat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} E(\|\widehat{\eta}_1 - \eta_1^*\|^2) \geq c\Big(p^{-2\alpha} + n^{-1}\Big),$$

where c is a positive constant depending on L, α and σ and the infimum is taken over all estimators i.e. all measurable functions of the observations $\{Y_i(t_h), h = 0, \ldots, p-1, i = 1, \ldots, n\}$.

Theorem 1 is obtained by combining Propositions 2 and 3 stated in Appendix B.

Proposition 2 provides the parametric rate n^{-1}, which is expected in our setting where we observe n curves. This rate has been proven to be optimal, up to logarthmic terms, in the case where the curves is supposed to be observed at all points (see e.g. Mas and Ruymgaart 2015). More precisely, Proposition 2 gives

$$\inf_{\widehat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\widehat{\eta}_1 - \eta_1^*\|^2] \geq c_1 n^{-1}, \quad \inf_{\widehat{\eta}_2} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\widehat{\eta}_2 - \eta_2^*\|^2] \geq c_1 n^{-1},$$

where c_1 depends on L, α and σ. The key point to establish the lower bound is the explicit form of the Kullback-Leibler divergence between two Gaussian distributions. The first two eigenvalues introduced in the building of models of the proof of Proposition 2 provide a constant spectral gap (i.e. $\mu_2^* - \mu_1^*$ is a constant independent of both n and p).

The lower bound of the minimax risk by $p^{-2\alpha}$ relies on the construction of two processes $Z_0, Z_1 \in \mathcal{R}_\alpha(L)$ with first eigenfunctions distant of $p^{-\alpha}$ from each other and such that $Z_0(t_h) = Z_1(t_h)$ almost surely for all $h = 0, \ldots, p-1$ (see Appendix B.2 and in particular Equation (8)).

Observe that if p, the number of observations per individual, is bounded, then rates cannot go to 0. In particular, consistency cannot be achieved by any estimate in our statistical model if p is a constant. Our results corroborate arguments of the discussion section 3.2 of Hall et al. [2006]. Parametric rates can be achieved only if p is large enough, namely larger than $n^{1/(2\alpha)}$.

We can remark that the constant L appearing in the regularity class is strongly linked with the eigenvalues sequence $(\mu_d^*)_{d\geq 1}$. This can be seen via Remark 1 but also via the proofs of Propositions 2 and 3 where the first eigenvalues of the processes we construct are upper-bounded, up to multiplicative constants, by L.

3.3 General upper bounds

We now derive upper bounds for estimates $\widehat{\eta}_{\phi,d}$. For this purpose, we set

$$b_1 = 8(\mu_1^* - \mu_2^*)^{-2}$$

and for $d = 2, \ldots, \mathcal{D}$,

$$b_d = 8/\min(\mu_d^* - \mu_{d+1}^*, \mu_{d-1}^* - \mu_d^*)^2.$$

Since we supposed that all the true eigenvalues μ_d^*'s are distinct, the quantities b_d's are well defined and finite.

The eigenfunction η_d^* being defined up to a sign change ($-\eta_d^*$ is also an eigenfunction associated to the eigenvalue μ_d^*), we cannot assess our procedure by using the classical risk $E(\|\widehat{\eta}_{\phi,d} - \eta_d^*\|^2)$. Following Bosq [2000], we evaluate the risk of

$$\eta_{\pm,d}^* = \text{sign}(\langle \widehat{\eta}_{\phi,d}, \eta_d^* \rangle) \times \eta_d^*.$$

We consider the following mild assumption on the 4th moment of the vector

$$\mathbf{Z} = \{Z(t_0), \ldots, Z(t_{p-1})\}^T.$$

Assumption 1 *We assume that there exists $C_1 > 0$ such that*

$$E\{(v^T\mathbf{Z})^4\} \leq C_1[E\{(v^T\mathbf{Z})^2\}]^2, \quad v \in \mathbb{R}^p. \tag{5}$$

Assumption 1 ensures a control of the fourth moment of $\tilde{z}_{1,\lambda}$. It is satisfied with $C_1 = 3$ if $\mathbf{Z}$ is Gaussian. Then we obtain the following result:

Theorem 2 *Let d be fixed. Under Assumption 1, we have*

$$E(\|\widehat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2) \leq 5b_d \left[|||\Pi_D \Gamma \Pi_D - \Gamma|||^2 + \frac{\max(C_1+3;6)}{n} \left\{ \sum_{\lambda \in \Lambda_D} \left(\sigma_\lambda^2 + s_\lambda^2 \right) \right\}^2 + A_p^{(K)}(\phi, D) + A_p^{(\sigma)}(\phi, D) + \frac{\sigma^4}{p^2} \right],$$

where Π_D *is the orthogonal projection onto* $S_D = span(\phi_\lambda, \lambda \in \Lambda_D)$*,*

$$\sigma_\lambda^2 = Var(\tilde{\varepsilon}_{1,\lambda}) = \frac{\sigma^2}{p^2}\sum_{h=0}^{p-1}\phi_\lambda^2(t_h), \quad s_\lambda^2 = Var(\tilde{z}_{1,\lambda}) = \frac{1}{p^2}\sum_{h,h'=0}^{p-1} K(t_h, t_{h'})\phi_\lambda(t_h)\phi_\lambda(t_{h'})$$

and

$$A_p^{(K)}(\phi, D) = \sum_{\lambda,\lambda' \in \Lambda_D}\left\{\frac{1}{p^2}\sum_{h,h'=0}^{p-1} K(t_h, t_{h'})\phi_\lambda(t_h)\phi_{\lambda'}(t_{h'}) - \int_0^1\int_0^1 K(s,t)\phi_\lambda(s)\phi_{\lambda'}(t)dsdt\right\}^2,$$

$$A_p^{(\sigma)}(\phi, D) = \frac{\sigma^4}{p^2}\sum_{\lambda,\lambda' \in \Lambda_D}\left\{\frac{1}{p}\sum_{h=0}^{p-1}\phi_\lambda(t_h)\phi_{\lambda'}(t_h) - \mathbf{1}_{\{\lambda=\lambda'\}}\right\}^2.$$

The first term of the upper bound is a bias term corresponding to the projection step, that decreases with D, the dimension of the approximation space. The second term is a variance term that increases with D but contrary to what happens generally in non-parametric statistics, it is bounded by n^{-1} up to a constant under mild assumptions on the orthonormal system (details in Section 3.4). Indeed, heuristically, when p grows, the term σ_λ^2 is of order σ^2/p (the variance due to the noise is tempered by the repetition of the observations) and the term s_λ^2 is of order $\iint K(s,t)\phi_\lambda(s)\phi_\lambda(t)dsdt = E(\langle Z, \phi_\lambda\rangle^2)$ so $\sum_{\lambda \in \Lambda_D} s_\lambda^2$ is bounded by a constant (independent of D) of order $E(\|Z\|^2) < +\infty$. By taking $D = \text{card}(\Lambda_D) \le p$, the second term is of order n^{-1}. The third and fourth terms are linked to the discretization and are usually negligible with respect to both the bias and variance terms (see Section 3.4). The term σ^4/p^2 is also negligible.

Remark 2 *We assume that* $\{\varepsilon_{i,h}\}_{i=1,\dots,n;h=0,\dots,p-1}$ *is a sequence of Gaussian errors. But the result of Theorem 2 still holds if the vectors* $\{\varepsilon_{i,h}\}_{h=0,\dots,p-1}$ *only satisfy Assumption 1 with* C_1 *a constant not depending on* i *(see the proof of Lemma 4).*

We can refine the previous result and obtain similar upper bounds in probability. To state them, we first recall the definition of sub-Gaussian variables. We refer to Section 2.5 of Vershynin [2018] for more details.

Definition 1 *We say that a random variable* W *is sub-Gaussian if*

$$\|W\|_{\psi_2} := \sup_{q\ge1}\left\{q^{-1/2}\left\{E(|W|^q)\right\}^{1/q}\right\} < \infty.$$

In this case, $\|W\|_{\psi_2}$ *is called the sub-Gaussian norm of* W.

Assumption 1 is extended to p-dimensional vectors as follows.

Assumption 2 *We assume that there exists* $C_2 > 0$ *such that*

$$\|v^T\mathbf{Z}\|_{\psi_2}^2 \le C_2 E(v^T\mathbf{Z})^2, \quad v \in \mathbb{R}^p.$$

Assumption 2 of Theorem 3 is stronger than Assumption 1 of Theorem 2 but it allows us to obtain an inequality in probability, which is stronger than in expectation; the price to pay is the logarithmic factor in the variance term as show in the following. Using quantities introduced in Theorem 2, we then obtain the following result.

Theorem 3 *Let d be fixed. Then, under Assumption 2, for all $\gamma > 0$, with probability larger than $1 - 2\exp(-1/64\min(\gamma^2, 16\gamma\sqrt{n}))$,*

$$\|\widehat{\eta}_{\phi,d} - \eta^*_{\pm,d}\|^2 \quad \leq \quad 5b_d \left[|||\Pi_D\Gamma\Pi_D - \Gamma|||^2 + \frac{(e^{1/2}+\gamma)^2\bar{C}^2(C_2+1)^2}{n}\left\{\sum_{\lambda\in\Lambda_D}\left(\sigma_\lambda^2 + s_\lambda^2\right)\right\}^2 \right.$$
$$\left. + A_p^{(K)}(\phi, D) + A_p^{(\sigma)}(\phi, D) + \frac{\sigma^4}{p^2}\right],$$

where $\bar{C}$ is an absolute constant.

Observe that if we take $\gamma = 8(\beta\log n)^{1/2}$, then for n large enough, the upper bound holds with probability larger than $1 - 2n^{-\beta}$. In this case, the order of the variance term is the same as for Theorem 2 up to the $\log n$-factor.

Theorem 3 is based on Assumption 2, namely a control of the sub-Gaussian norm of $v^T\mathbf{Z}$ for all vectors v. Such controls are standard to obtain concentration inequalities which are at the core of the proof of Theorem 3; see for instance Vershynin [2018], Koltchinskii and Lounici [2017] or Section 2.3 of Boucheron et al. [2013]. This assumption enables us to apply large deviation bounds for martingale differences established by Juditsky and Nemirovski [2008] to specific covariance matrices. See Proposition 5 in the Appendix for more details. Observe that Assumption 2 is satisfied if $\mathbf{Z}$ is Gaussian and in this case C_2 is an absolute constant (see Example 2.5.8 of Vershynin [2018]). In a more general setting, we have the following lemma proved in Appendix C.5.

Lemma 1 *We consider the Karhunen-Loève representation of Z given by* (3) *and assume that the scores ζ_d^*' are independent and that there exists $M < \infty$ such that*

$$\sup_{d\in\mathbb{N}^*} \|\zeta_d^*\|_{\psi_2} \leq M. \tag{6}$$

Then Assumption 2 is satisfied with $C_2 = \kappa M^2$, where κ is an absolute constant.

Under the assumptions of Lemma 1 (and others), Mas and Ruymgaart [2015] obtain rates of convergence for the eigenvectors and eigenprojectors in expectation in the case where the curves are observed at all points. In the context of functional Principal Components Regression, these assumptions are classical, we refer e.g. to Hall and Hosseini-Nasab [2009], Crambes and Mas [2013].

We have the analog of Remark 2.

Remark 3 *We assume that $\{\varepsilon_{i,h}\}_{i=1,\ldots,n;h=0,\ldots,p-1}$ is a sequence of Gaussian errors. But the result of Theorem 3 still holds if the vectors $\{\varepsilon_{i,h}\}_{h=0,\ldots,p-1}$ only satisfy Assumption 2 with C_2 a constant not depending on i (see the proofs of Lemmas 5 and 6).*

Remark 4 *The first step of the proof of our results consists in applying Bosq inequalities to bound $\|\widehat{\eta}_{\phi,d} - \eta^*_{\pm,d}\|$. Similar bounds also hold for $|\widehat{\mu}_d - \mu_d^*|$. See Section C of the Appendix for more details. Therefore, bounds of the previous theorems also hold for $|\widehat{\mu}_d - \mu_d^*|^2$. Obtaining lower bounds for the estimation of the eigenvalues remains an open interesting question.*

3.4 Upper bound for histograms

In this paragraph, we specify our results for the case of histograms. The histogram system is defined as follows (see Section 7.3 of Massart [2007] for instance).

Definition 2 *Let $\Lambda_D = \{0, \ldots, D-1\}$. For any $\lambda \in \Lambda_D$,*

$$\phi_\lambda(t) = D^{1/2} \times \mathbf{1}_{I_\lambda}(t), \quad t \in [0,1],$$

with $I_\lambda = (\lambda/D, (\lambda+1)/D]$. For any $(\lambda, \lambda') \in \Lambda^2$, $\langle\phi_\lambda, \phi_{\lambda'}\rangle = \mathbf{1}_{\{\lambda=\lambda'\}}$.

In the sequel, we consider the following assumption.
Assumption 3 *The integer D is such that D divides p.*

In this framework, all terms appearing in upper bounds of Theorems 2 and 3 can be easily controlled.
Proposition 1 *Under Assumption 3, if $P_Z \in \mathcal{R}_\alpha(L)$, we have*

$$|||\Pi_D \Gamma \Pi_D - \Gamma|||^2 \leq \frac{16L\|K\|_\infty}{(\alpha+1)^2} D^{-2\alpha},$$

$$A_p^{(K)}(\phi, D) \quad \leq \quad \frac{16\|K\|_\infty L}{(\alpha+1)^2} p^{-2\alpha}, \quad A_p^{(\sigma)}(\phi, D) = 0$$

and

$$\sum_{\lambda \in \Lambda_D} \left(\sigma_\lambda^2 + s_\lambda^2\right) \leq \|K\|_\infty + \frac{\sigma^2 D}{p}.$$

Combining Proposition 1 with Theorems 2 and 3, we finally deduce the following corollary.
Corollary 1 *Let d be fixed. Assume that $P_Z \in \mathcal{R}_\alpha(L)$ and $D = p$. Under Assumption 1,*

$$E\Big(\|\widehat{\eta}_{\phi,d} - \eta^*_{\pm,d}\|^2\Big) \quad \leq \quad b_d \left\{ \frac{B(L,K,\alpha)}{p^{2\alpha}} + \frac{5\sigma^4}{p^2} + \frac{V_1(K,\sigma,C_1)}{n} \right\}$$

and under Assumption 2, for any $\beta > 0$, for n large enough, with probability larger than $1 - 2n^{-\beta}$,

$$\|\widehat{\eta}_{\phi,d} - \eta^*_{\pm,d}\|^2 \quad \leq \quad b_d \left\{ \frac{B(L,K,\alpha)}{p^{2\alpha}} + \frac{5\sigma^4}{p^2} + \frac{V_2(K,\sigma,C_2,\beta)\log n}{n} \right\},$$

where $B(L,K,\alpha)$ depends on L, $\|K\|_\infty$ and α and $V_1(K,\sigma,C_1)$ (resp. $V_2(K,\sigma,C_2,\beta)$) depends on $\|K\|_\infty$, σ and C_1 (resp. $\|K\|_\infty$, σ, C_2 and β) (the constants $B(L,K,\alpha)$, $V_1(K,\sigma,C_1)$ and $V_2(K,\sigma,C_2,\beta)$ are deterministic).
Since $\alpha \leq 1$, the term $5\sigma^4/p^2$ is not larger than the first term $B(L,K,\alpha)/p^{2\alpha}$ (up to a constant), and the noise of the observations has no influence on the rates (as soon as the noise level is a constant). In particular, under Assumption 1,

$$\sup_{P_Z \in \mathcal{R}_\alpha(L)} E\Big(\|\widehat{\eta}_{\phi,d} - \eta^*_{\pm,d}\|^2\Big) \leq C\Big(p^{-2\alpha} + n^{-1}\Big),$$

for C a constant. This upper bound and the lower bound of Theorem 1 match, meaning that our estimation procedure is optimal in our setting. Observe that Assumption 1 is very mild. If we replace it with the stronger Assumption 2, we obtain a control in probability, coming from exponential bounds on probabilities of large deviations (for the Frobenius norm) for specific matrices. The price to pay is a logarithmic term in the variance term.

As expected, parameters p and n have a strong influence on rates. In our framework with two asymptotics very different in nature, we note that if p is large enough (depending on n and α), then our procedure achieves the parametric rate n^{-1} already obtained by Dauxois et al. [1982] and Mas and Ruymgaart [2015] when the curves $(Z_1, \ldots, Z_n)$ are fully observed and without noise. It means that discretization has no influence on theoretical performances. Conversely, if p is not very large with respect to n, discretization has a deep impact and rates depend strongly on the underlying smoothness of the curves observed in a noisy setting. The obtained rate $p^{-2\alpha} + n^{-1}$ describes very precisely the competition between the number of discretization points and the number of observations in functional principal component analysis. To the best of our knowledge, these rates are new.

Finally, let us emphasize the simplicity of our optimal estimation procedure. It is based on the most classical ideas: projection by using piecewise constant bases and empirical mean estimation. In particular, regularization is not necessary and the knowledge of α is not required. The use of such standard tools may be surprising in view of results obtained by Johnstone [2001] and Baik and Silverstein [2006]. But, as already mentioned, functional principal component analysis is a very specific setting. The rates we obtain have the same shape to those of Descary and Panaretos [2019] for which strong assumptions on the covariance operator (finite rank and analyticity of the eigenvalues η_d^*) are required and the noise may exhibit local correlations. Bunea and Xiao [2015] obtained a rate of convergence that is difficult to compare with ours, because their assumptions, concerning the rate of decay of the eigenvalues, differ significantly from our regularity assumption on the process. In the case of Brownian motion the rate of convergence of the $\mathbb{L}^2$- risk of reconstruction of the estimator of Bunea and Xiao [2015] is of order $\log^2(n)n^{-1} + p^{-1/3}$ which is suboptimal compared to the minimax rate of $n^{-1} + p^{-1}$ that we have proven. In a different statistical model where observational times are random, Hall et al. [2006] obtained an $\mathbb{L}^2$ convergence rate $n^{-2r/(2r+1)}$ for kernel estimators when η_d^* has a r-th order derivative even if the number of observations per curve is bounded by a constant. This clearly shows the impact of the nature of the observations. The randomness of observational times may allow to circumvent the sparse sampling scheme for individuals and consistency may be achievable, which is not the case for our statistical model with deterministic equispaced observational times.

4 Simulation results

We assess the statistical performance of functional principal components estimators with simulations. We consider two eigenfunctions such that $\eta_1^*(\cdot) = \sqrt{2}\sin(2\pi\cdot)$ and $\eta_2^*(\cdot) = \sqrt{2}\cos(2\pi\cdot)$, with eigenvalues $\mu_1^* = 1.1$ and $\mu_2^* = 0.1$. Simulated functional data are sampled on regularly spaced discretization points $t_h = h/(p-1)$ with $h = 0, \dots, p-1$, and we compute the covariance matrix Σ such that:

$$\Sigma_{h,h'} = \sum_{d=1}^{2} \mu_d^* \eta_d^*(t_h) \eta_d^*(t_{h'}) + \sigma^2 1_{h=h'}$$

from which we sample n random functions $Y_1, \dots, Y_n \sim \mathcal{N}(0, \Sigma)$, following Model (2). Then we consider different values for the number of observations $n \in \{256, 512, 1024, 2048, 4096\}$ to study the asymptotic performance of our estimators, and we will also consider different values of $p \in \{16, 32, 64, 128, 256\}$ to study the impact of discretization. The noise level σ is chosen to match a given signal to noise ratio defined by $\sigma^{-2}\sum_{d=1}^{2}\mu_d^*$ (the variance of the signal divided by the variance of the noise), that takes value in $\{0.25, 1, 4\}$. We consider two smoothing systems, histograms and the Haar wavelets, as detailed in the Appendix. We report average the performance on $nb_{test} = 100$ independent simulations. Even if our theoretical results do not include regularized estimators, we also consider a hard thresholded version of these estimators to improve reconstruction (as detailed in the Appendix).

4.1 Reconstruction Errors

To assess the empirical performance of our approach, we study the behavior of the mean reconstruction error

$$E\Big(\|\eta_{\pm,d}^* - \widehat{\eta}_{\phi,d}\|^2\Big)$$

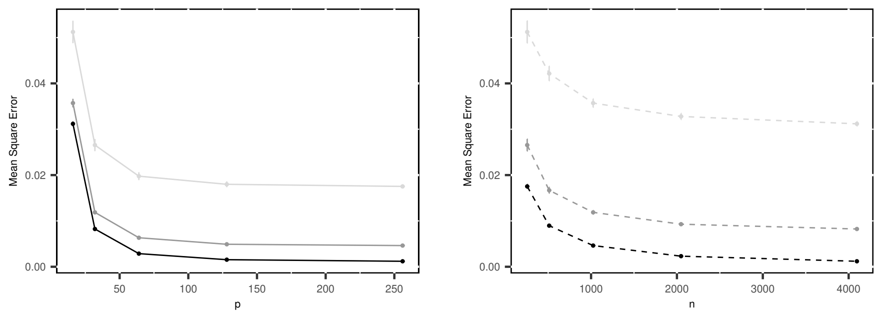

Figure 1: Mean square error for the first eigenfunction η_1^* according to the number of discretization points p (left), and the number of samples n (right). Left: the number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). Right: the number of discretization points is $p \in \{16, 32, 256\}$ (light gray, gray, black respectively). The signal to noise ratio is 0.25.

according to the number of observations n, the size of the discretization grid p, and the signal to noise ratio. More precisely, we introduce a second finer grid $s_h = h/p'$, with $h = 0, \dots, p'-1$, such that $p' \gg p$ ($p' = 8192$ in practice) and use the approximation

$$\mathbb{E}\Big(\|\eta^*_{\pm,d} - \widehat{\eta}_{\phi,d}\|^2\Big) \approx \frac{1}{nb_{test}} \sum_{j=1}^{nb_{test}} \frac{1}{p'} \sum_{h=1}^{p'} \Big\{\eta^*_{\pm,d}(s_h) - \widehat{\eta}^j_{\phi,d}(s_h)\Big\}^2.$$

To deduce the values of our estimator outside of the original grid, we use the piecewise constant property of the Haar and the histogram systems. In the following we also compute the estimation error on eigenvalues and assess $E\{(\mu^*_d - \widehat{\mu}_{\phi,d})^2\}$.

4.2 Results

The empirical error rates on eigenfunctions match the theoretical ones (Fig. 1), with orders $(\mu^*_1 - \mu^*_2)^{-2}(n^{-1} + p^{-2\alpha})$ ($\alpha = 1$ in our case) for the first eigenfunction estimator and $(\mu^*_2 - \mu^*_3)^{-2}(n^{-1} + p^{-2\alpha})$ for the second (Fig. 5 and 6, $\mu^*_3 = 0$ in our setting). Computed errors exhibit a double asymptotic behavior in n and p. The rates in n are slower than those in p, and exactly match n^{-1} and p^{-2}. Also, the difference in terms of mean square errors between the first and the second eigenfunctions is due to the gap, since $\mu^*_1 - \mu^*_2 = 10(\mu^*_2 - \mu^*_3)$, which means that the estimation of the second eigenfunction is 10 times harder in terms of speed of convergence. The estimation error on eigenvalues also matches the theoretical upper bound (Fig 2).

We also show that regularization does not necessarily improve the rate of convergence of eigen-elements (Fig. 7 and 8). We showed that projection-based functional principal component should attain minimax rates without any regularization. Consequently, at best, regularization should induce a better variance but comparable rates. Since one could operate only on the observed part of the function, at best, one could improve results on the grid without going beyond n^{-1}, which is already attained by the non-smoothed estimator.

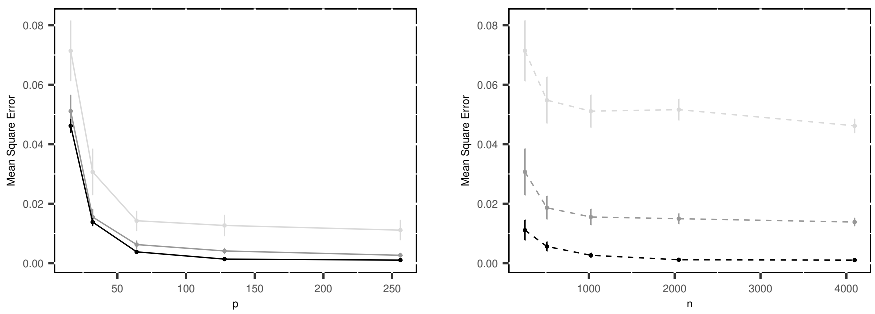

Figure 2: Mean square error for the first eigenvalue μ_1^* according to the number of discretization points p (left), and the number of samples n (right). Left: the number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). Right: the number of discretization points is $p \in \{16, 32, 256\}$ (light gray, gray, black respectively). The signal to noise ratio is 0.25.

5 Applications

Single-cell genomics has been made possible thanks to technological breakthroughs that have allowed the measurement of gene expression Macosko et al. [2015] at the single-cell resolution. These advances have revolutionized our view of the complexity of living tissues, in their normal or pathological states, and have produced complex high dimensional data. Immunology, and in particular studies of lymphocytes differentiation have focused much attention. Indeed, upon an acute infection, pathogen-specific CD8 T cells are activated, proliferate drastically and differentiate short-term (ou short-lived) effector cells displaying the ability to eliminate infected cells. the response also generates a small population of cells being part of the long-term immune response, the so-called memory cells conferring protection to the host. In Kurd et al. [2020], the authors collected single antigen-specific CD8 T cells in the spleen of mice after LCMV (lymphocytic choriomeningitis virus) acute infection and measured the expression of genes at time points 0, 3, 4, 5, 6, 7, 10, 14, 21, 32 and 90 days post-infection (dpi). Hence, these data offer the unique opportunity to study the evolution of gene expression throughout an immune response. Many methodological questions can be addressed with these data, and we consider here a short analysis to illustre functional PCA on original data. Average expressions were considered over cells such that the data correspond to the expression of genes over time. Expression data were normalized thanks to the `SCTransform` procedure of the `Seurat` package [Satija et al., 2015] that corrects counts for overdispersion, and provides corrected counts for which the Gaussian approximation is reasonable. Then genes were selected by using the `FindVariableFeatures` function [Satija et al., 2015], resulting in 4851 genes with averaged temporal expression that constitute the input of our model ($Y_i(t)$, $n = 4851$, $p = 11$). When performing functional PCA on those data using the histogram basis, the rule of thumbs suggests to select 3 functional principal components, and k-means clustering exhibits 3 clusters with very distinct average expression profiles (Fig. 3). Interestingly, cluster 1 consists of only two genes Ccl5 and Malat1 that are known to be involved in immune cells activation and recruitement [Araujo et al., 2018] and CD8+ differentiation [Kanbar et al., 2022], respectively. Their kinetics of expression appears then as a major feature of the gene-expression response to viral infection: they are down-regulated between days 0-8, then up-regulated between 9-90 days, whereas the 31 genes of cluster 2 are up

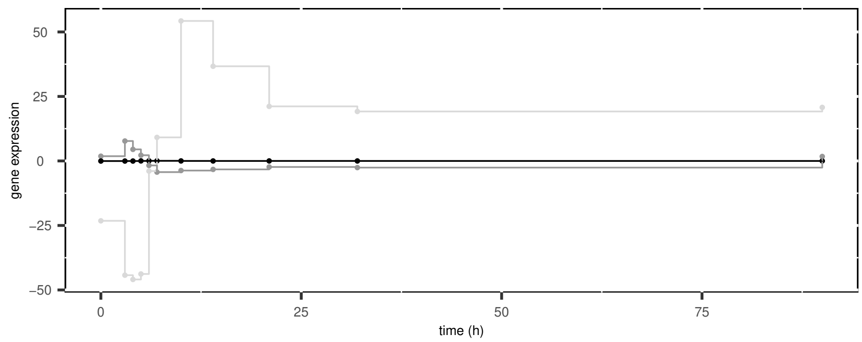

Figure 3: Average gene expression over time for lymphocytes CD8+ cells following a viral infection, as inferred by functional PCA followed by k-means clustering. Curves correspond to the averages of gene expression for the 3 clusters (1-light grey, 2-dark grey, 3-black).

and then down-regulated (Gzmb, Ncl, Tuba1b, Hsp90aa1, Npm1, Hspe1, Ptma, Ran, Hsp90ab1, Actg1, Mif, Tubb5, Eif5a, Ldha, Plac8 and ribosome proteins), the other genes showing a stable average expression stable over time (cluster 3). This first analysis allows us to illustrate the interest of functional PCA on gene expression data, and the underlying gene regulatory networks that structure these gene expression dynamics will deserve a deeper study.

Genomics offers another original application of functional principal component analysis, to reduce the dimensionality of data that are structured in one dimension along the genome. As an illustration, we consider the fine mapping of replication origins in the human genome, that constitutes the starting points of chromosomes duplication. Replication origins are under a very strong spatio-temporal control, and are part of the integrity maintenance of genomes. The investigation of their spatial organization has become central to better understand genomes architecture and regulation, which remains challenging due to a complex interplay between genetic and epigenetic regulations. Part of the genetic component of their regulation involve particular sequence motifs, called G-quadruplexes, that have the property to form complex four-stranded structures whose role in replication remains unsolved. A crucial aspect to better understand their implication is to determine if these sequence motifs have a preferential positioning upstream replication origins. To investigate this matter, we considered the $\sim$130,000 replication origins of the human genome [Picard et al., 2014], and we defined by $Y_i(t)$ the process that equals 1 if there is a G-quadruplex at position t in replication origin i, taking motifs coordinates from Zheng et al. [2020]. Hence this application goes beyond the Gaussian setting of our model, and shows that extension to count data is also effective with histogram-based functional-PCA. By convention, $t = 0$ corresponds to the peak of replication, and we consider positions 500 bases upstream this peak (in negative coordinates). The continuous aspect of the model is not mandatory since positions along the genome are discrete. However, the functional setting allows us to consider the spatial dependencies between the occurrences of these motifs, which is very informative. Given the discrete nature of the data, we smoothed the data using the histogram system, with bin size of 25 bases (corresponding to the average size of G-quadruplexes). Then we performed functional principal component analysis, and we used functional principal components to perform a downstream clustering. We projected every observed curve on principal components to obtain a new representation of the functional data based on general terms $\langle \widetilde{Y}_i, \widehat{\eta}_{\phi,d} \rangle$, and performed

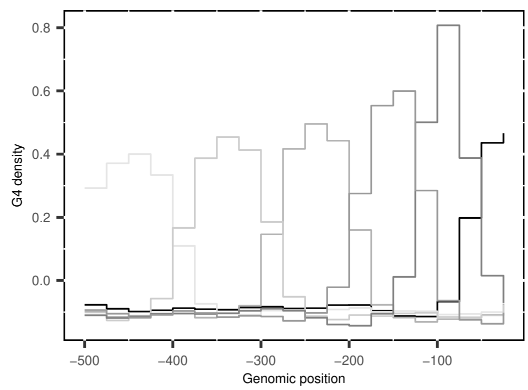

Figure 4: Density of G-quadruplexes accumulation in human replication origins clusters, determined by functional principal component analysis combined with k-means clustering. Each color correspond to a particular cluster.

a k-means clustering to regroup replication origins that share the same spatial distribution of these G-quadruplex motifs. We considered 6 principal components along with 6 clusters, and we considered the spatial distribution of G-quadruplexes within clusters as a result (Figure 4). Functional principal component analysis appears to catch the spatial structure that makes the clusters, as different clusters of replication origins are characterized by specific patterns of G-quadruplexes accumulation upstream the replication peak. Interestingly, the observed periodicity can be related to a biophysical property of chromatin fibers. Indeed, the DNA molecule is in the form of chromatin fibers in the nucleus, wrapped around the so-called nucleosomes with a periodicity of 144 base pairs. The formation of stable G-quadruplexes has been shown to take place in nucleosome-free regions [Prorok et al., 2019], hence, the periodicity of their accumulation upstream replication origins indicates that their positioning is directly linked to the epigenetic context of replication initiation. These new biological results are currently under further investigation, and developing a framework for functional-PCA dedicated to count data could be an interesting research direction [Backenroth et al., 2020].

6 Conclusion and perspectives

In this paper, we have established optimal rates of convergence for estimating the eigenfunctions of the covariance operator of a corrupted process observed on a fixed and regular grid. It is shown that the minimax rate is of order $p^{-2\alpha} + n^{-1}$, revealing the behavior of rates with respect to the parameters n, p and $\alpha \in (0, 1]$ that design the number of repetitions, the size of the regular grid and the smoothness of the process respectively. In this framework with a double asymptotic in n and p, the phase transition occurs when p is of order $n^{1/(2\alpha)}$:

- For p larger than $n^{1/(2\alpha)}$, the obtained rate is n^{-1} and the problem has intrinsically the same difficulty if the n curves would be available entirely.
- For p smaller $n^{1/(2\alpha)}$, the rate is $p^{-2\alpha}$ regardless the number of observed curves.

In particular, $p \to +\infty$ is required to achieve consistency. These rates are new and are fundamentally different from classical nonparametric rates of the form $(np)^{-2\alpha/(2\alpha+1)}$ in a setting with $n \times p$ observations, revealing main differences with the framework where the sampling scheme is random [Hall et al., 2006] and typical of longitudinal data analysis. In our setting, optimality is achieved by diagonalizing the empirical operator after simple projections by using histograms with p bins. Surprisingly, it means that smoothing and the knowledge of α are not required. These results have important theoretical and practical implications in many aspects regarding the sampling scheme, the used methodology or the grid size if it is left to the practitioner (the choice of the grid itself is a research field on its own and out of the scope of the paper, we refer to Müller-Gronbach [1996], Seleznjev [2000] for articles on optimal grid selection). These conclusions bear some similarities with those drawn by Cai and Yuan [2011] for mean function estimation.

Following this work, several perspectives can be envisaged. The first one concerns the range of α and it should be interesting to extend our theoretical results to any smoothness $\alpha \in \mathbb{R}_+^*$, even if it means changing the definition of the class $\mathcal{R}_\alpha(L)$. This extended framework with higher smoothness raises several open questions: Are rates different? If yes, do we need to introduce smoothing? Can we still consider histograms or do we have to project on smoother bases such as wavelets, splines or Fourier bases?

Our minimax rates depend on the sequence of eigenvalues $(\mu_d^*)_d$ through the terms $(b_d)_d$'s that are viewed as (unknown) constants in our paper. It would be very interesting to investigate how such parameters influence rates when they become very small. Concerning the problem of estimating eigenvalues, Remark 4 provides some upper bounds but optimality of these upper bounds remain a very challenging problem, out of the scope of this paper.

Besides, as pointed out by Hörmann and Jammoul [2022], assuming that the noise is i.i.d. may be unrealistic in some applications. Hörmann and Jammoul [2022] assume for instance that the vector of errors $(\varepsilon_{i,h})_{h=0,\ldots,p}$ is a stationary process, while Descary and Panaretos [2019] model the errors as realizations of a stochastic process with short term dependency. Both articles obtain convergence rates for the eigenfunctions in the case where the covariance operator Γ is of finite-rank. Based on these works, it could be of interest, in a future work, to extend the minimax approach developed in this paper to models allowing non i.i.d. errors.

References

J. M. Araujo, A. C. Gomez, A. Aguilar, R. Salgado, J. M. Balko, L. Bravo, F. Doimi, D. Bretel, Z. Morante, C. Flores, H. L. Gomez, and J. A. Pinto. Effect of CCL5 expression in the recruitment of immune cells in triple negative breast cancer. Sci Rep, 8(1):4899, Mar 2018.

D. Backenroth, R. T. Shinohara, J. A. Schrack, and J. Goldsmith. Nonnegative decomposition of functional count data. Biometrics, 76(4):1273–1284, 2020.

J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. J. Multivariate Anal., 97(6):1382–1408, 2006.

D. Bosq. Linear Processes in Function Spaces: Theory and Applications. Springer New York, 2000.

S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

H. Brezis. Functional analysis, Sobolev spaces and partial differential equations. Universitext. Springer, New York, 2011. ISBN 978-0-387-70913-0.

F. Bunea and L. Xiao. On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA. Bernoulli, 21(2):1200–1230, 2015.

T. T. Cai and M. Yuan. Optimal estimation of the mean function based on discretely sampled functional data: phase transition. Ann. Statist., 39(5):2330–2355, 2011.

T. T. Cai and M. Yuan. Minimax and adaptive prediction for functional linear regression. J. Amer. Statist. Assoc., 107(499):1201–1216, 2012.

H. Cardot and J. Johannes. Thresholding projection estimators in functional linear models. J. Multivariate Anal., 101(2):395–408, 2010.

C. Crambes and A. Mas. Asymptotics of prediction in functional linear regression with functional outputs. Bernoulli, 19(5B):2627–2651, 2013. ISSN 1350-7265. doi: 10.3150/12-BEJ469. URL https://dx-doi-org.frodon.univ-paris5.fr/10.3150/12-BEJ469.

X. Dai, H.-G. Müller, and W. Tao. Derivative principal component analysis for representing the time dynamics of longitudinal and functional data. Statist. Sinica, 28(3):1583–1609, 2018. ISSN 1017-0405.

J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. Journal of Multivariate Analysis, 12(1):136–154, 1982.

M.-H. Descary and V. M. Panaretos. Functional data analysis by matrix completion. Ann. Statist., 47(1):1–38, 2019. ISSN 0090-5364.

E. Devijver, Y. Goude, and J.-M. Poggi. Clustering electricity consumers using high-dimensional regression mixture models. Applied Stochastic Models in Business and Industry, 36(1):159–177, 2020.

F. Ferraty and Y. Romain, editors. The Oxford handbook of functional data analysis. Oxford University Press, Oxford, 2011.

F. Ferraty and P. Vieu. Nonparametric functional data analysis. Springer Series in Statistics. Springer, New York, 2006. Theory and practice.

K. Goode, D. Ries, and J. Zollweg. Explaining neural network predictions for functional data using principal component analysis and feature importance, 2020.

P. Hall and M. Hosseini-Nasab. Theory for High-Order bounds in functional principal components analysis. Mathematical Proceedings of the Cambridge Philosophical Society, 146(01): 225–256, 2009. doi: 10.1017/S0305004108001850.

P. Hall, H.-G. Müller, and J.-L. Wang. Properties of principal component methods for functional and longitudinal data analysis. The Annals of Statistics, 34(3):1493 – 1517, 2006.

S. Hörmann and F. Jammoul. Preprocessing noisy functional data: A multivariate perspective. Electronic Journal of Statistics, 16(2):6232 – 6266, 2022. doi: 10.1214/22-EJS2083. URL https://doi.org/10.1214/22-EJS2083.

S. Jaimungal and E. Ng. Consistent functional pca for financial time-series, 05 2007.

I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. Ann. Statist., 29(2):295–327, 2001.

I. M. Johnstone and A. Y. Lu. On consistency and sparsity for principal components analysis in high dimensions. Journal of the American Statistical Association, 104(486):682–693, 2009.

A. Juditsky and A. S. Nemirovski. Large deviations of vector-valued martingales in 2-smooth normed spaces, 2008.

I. Kalogridis and S. Van Aelst. Robust functional regression based on principal components. Journal of Multivariate Analysis, 173:393–415, 2019.

J. N. Kanbar, S. Ma, E. S. Kim, N. S. Kurd, M. S. Tsai, T. Tysl, C. E. Widjaja, A. E. Limary, B. Yee, Z. He, Y. Hao, X. D. Fu, G. W. Yeo, W. J. Huang, and J. T. Chang. The long noncoding RNA Malat1 regulates CD8+ T cell differentiation by mediating epigenetic repression. J Exp Med, 219(6), Jun 2022.

V. Koltchinskii and K. Lounici. Concentration inequalities and moment bounds for sample covariance operators. Bernoulli, 23(1):110–133, 2017.

N. S. Kurd, Z. He, T. L. Louis, J. J. Milner, K. D. Omilusik, W. Jin, M. S. Tsai, C. E. Widjaja, J. N. Kanbar, J. G. Olvera, T. Tysl, L. K. Quezada, B. S. Boland, W. J. Huang, C. Murre, A. W. Goldrath, G. W. Yeo, and J. T. Chang. T lymphocytes revealed by single-cell RNA sequencing. Sci Immunol, 5(47), May 2020.

M. A. Lifshits. Gaussian random functions, volume 322 of Mathematics and its Applications. Kluwer Academic Publishers, Dordrecht, 1995.

E. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. Bialas, N. Kamitaki, E. Martersteck, J. Trombetta, D. Weitz, J. Sanes, A. Shalek, A. Regev, and S. McCarroll. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell, 161(5):1202–1214, May 2015. ISSN 00928674. doi: 10.1016/j.cell.2015.05.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0092867415005498.

A. Mas and F. Ruymgaart. High-dimensional principal projections. Complex Analysis and Operator Theory, 9:35–63, 2015.

P. Massart. Concentration inequalities and model selection, volume 1896 of Lecture Notes in Mathematics. Springer, Berlin, 2007. ISBN 978-3-540-48497-4; 3-540-48497-3. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

T. Müller-Gronbach. Optimal designs for approximating the path of a stochastic process. J. Statist. Plann. Inference, 49(3):371–385, 1996.

T. Perrin, O. Roustant, J. Rohmer, O. Alata, J. Naulin, D. Idier, R. Pedreros, D. Moncoulon, and P. Tinard. Functional principal component analysis for global sensitivity analysis of model with spatial output. Reliability Engineering & System Safety, 211, 2021.

H. Pham, S. Mottelet, O. Schoefs, A. Pauss, V. Rocher, C. Paffoni, F. Meunier, S. Rechdaoui, and S. Azimi. Estimation simultanée et en ligne de nitrates et nitrites par identification spectrale UV en traitement des eaux usées. L'eau, l'industrie, les nuisances, 335:61–69, 2010.

F. Picard, J. C. Cadoret, B. Audit, A. Arneodo, A. Alberti, C. Battail, L. Duret, and M. N. Prioleau. The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. PLoS Genet, 10(5):e1004282, May 2014.

P. Prorok, M. Artufel, A. Aze, P. Coulombe, I. Peiffer, L. Lacroix, A. Guédin, J. L. Mergny, J. Damaschke, A. Schepers, C. Cayrou, M. P. Teulade-Fichou, B. Ballester, and M. Méchali. Involvement of G-quadruplex regions in mammalian replication origin activity. Nat Commun, 10(1):3274, 07 2019.

J. Ramsay and B. W. Silverman. Functional Data Analysis. Springer New York, 2010.

C. E. Rasmussen and C. K. I. Williams. Gaussian processes for machine learning. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2006. ISBN 978-0-262-18253-9.

R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, and A. Regev. Spatial reconstruction of single-cell gene expression data. Nature Biotechnology, 33:495–502, 2015. doi: 10.1038/nbt.3192. URL https://doi.org/10.1038/nbt.3192.

O. Seleznjev. Spline approximation of random processes and design problems. J. Statist. Plann. Inference, 84(1-2):249–262, 2000.

W.-K. Seo. Functional principal component analysis of cointegrated functional time series. 2011.12781, arXiv.org, 2020.

W. Song, H.-S. Oh, Y. Lim, and Y. K. Cheung. Multi-feature clustering of step data using multivariate functional principal component analysis, 2020.

A. B. Tsybakov. Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York, 2009.

R. Vershynin. High-dimensional probability: An introduction with applications in data science, volume 47. Cambridge university press, 2018.

F. Yao, H.-G. Müller, and J.-L. Wang. Functional data analysis for sparse longitudinal data. J. Amer. Statist. Assoc., 100(470):577–590, 2005. ISSN 0162-1459. doi: 10.1198/016214504000001745. URL https://doi.org/10.1198/016214504000001745.

K. W. Zheng, J. Y. Zhang, Y. D. He, J. Y. Gong, C. J. Wen, J. N. Chen, Y. H. Hao, Y. Zhao, and Z. Tan. Detection of genomic G-quadruplexes in living cells using a small artificial protein. Nucleic Acids Res, 48(20):11706–11720, 11 2020.

R. Zhong, S. Liu, H. Li, and J. Zhang. Robust functional principal component analysis for non-gaussian longitudinal data. Journal of Multivariate Analysis, 189:104864, 2022. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2021.104864. URL https://www.sciencedirect.com/science/article/pii/S0047259X21001421.

Acknowledgements

The authors want to thank both referees for their helpful comments which have improved the presentation of this work. We also would like to thank Christophe Arpin and Margaux Preux for their help on the scRNAseq data analysis. The research was supported by a grant from Région Ile-de-France, and by grants from the Agence Nationale de la Recherche ANR-18-CE45-0023 SingleStatOmics and ANR-18-CE40-0014 Smiles.

A Simulation study

We consider two smoothing systems, the histogram system, and the Haar wavelet system. In the case of histograms, we denote by D the number of bins (such that D divides p in practice), then $\Lambda_D = \{0, \ldots, D-1\}$ and

$$\phi_\lambda(t) = D^{1/2} \times \mathbf{1}_{(\lambda/D,(\lambda+1)/D]}(t), \quad t \in [0,1], \lambda \in \Lambda_D.$$

Then in the case of the Haar system, we consider $(\varphi_{0,0}, \psi_{j,k}, j = 0, \ldots, J, k = 0, \ldots, 2^j - 1)$, with $J + 1 = \log_2(p)$, $\varphi_{0,0}$ the scaling function of a multi-resolution analysis (father wavelet) and $\psi_{j,k}$ the associated mother wavelets, such that

$$\begin{aligned}
&\varphi_{0,0}(t) = \mathbf{1}_{[0,1]}(t),\\
&\psi_{j,k}(x) = 2^{j/2}\mathbf{1}_{\left[\frac{2k-2}{2^{j+1}};\frac{2k-1}{2^{j+1}}\right]}(t) - 2^{j/2}\mathbf{1}_{\left(\frac{2k-1}{2^{j+1}};\frac{2k}{2^{j+1}}\right]}(t), \quad t \in [0,1].
\end{aligned}$$

We introduce a cross-validation procedure to regularize the eigenfunctions estimators. For each fold $r \in \{1, \ldots, n_{\text{folds}}\}$ we split the observations $Y = (Y_1, \ldots, Y_n)$ into two training and test sets $Y^{\text{train}_r}, Y^{\text{test}_r}$ such that $\mid \text{train}_r \cup \text{test}_r \mid = n$. Then we introduce ζ, a thresholding parameter, and we set

$$\widehat{Y}_{i,\zeta}(t) = \widetilde{y}_{i,0,0}\varphi_{0,0}(t) + \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} \widetilde{y}_{i,j,k}\mathbf{1}_{|\widetilde{y}_{i,j,k}|>\zeta}\psi_{j,k}(t), \quad i = 1, \ldots, n, \quad t \in [0,1],$$

with

$$\begin{aligned}
\widetilde{y}_{i,0,0} &= \frac{1}{p}\sum_{h=0}^{p-1} Y_i(t_h)\varphi_{0,0}(t_h),\\
\widetilde{y}_{i,j,k} &= \frac{1}{p}\sum_{h=0}^{p-1} Y_i(t_h)\psi_{j,k}(t_h).
\end{aligned}$$

Then we compute the $\widehat{\eta}^r_{d,\zeta}$'s on Y^{train_r} for each fold such that

$$(\widehat{\eta}^r_{d,\zeta})_d \in \underset{\langle f_d, f_{d'}\rangle = 1_{d=d'}}{\arg\min} \sum_{i \in \text{train}_r} \|\widehat{Y}_{i,\zeta} - \sum_{d=1}^{2}\langle \widehat{Y}_{i,\zeta}, f_d\rangle f_d\|^2.$$

We select $\widehat{\zeta}$, the minimizer of the cross validated errors :

$$\frac{1}{n_{\text{folds}}}\sum_{r=1}^{n_{\text{folds}}}\sum_{i \in \text{test}_r} \|Y_i - \sum_{d=1}^{2}\langle Y_i, \widehat{\eta}^r_{d,\zeta}\rangle \widehat{\eta}^r_{d,\zeta}\|^2.$$

Once $\widehat{\zeta}$ is chosen we compute the final estimator $\widehat{\eta}_{d,\widehat{\zeta}}$ as :

$$(\widehat{\eta}_{d,\widehat{\zeta}})_d \in \underset{\langle f_d, f_{d'}\rangle = 1_{d=d'}}{\arg\min} \sum_{i=1}^{n} \|\widehat{Y}_{i,\widehat{\zeta}} - \sum_{d=1}^{2}\langle \widehat{Y}_{i,\widehat{\zeta}}, f_d\rangle f_d\|^2.$$

We use the same score function to select the number of bins for histograms.

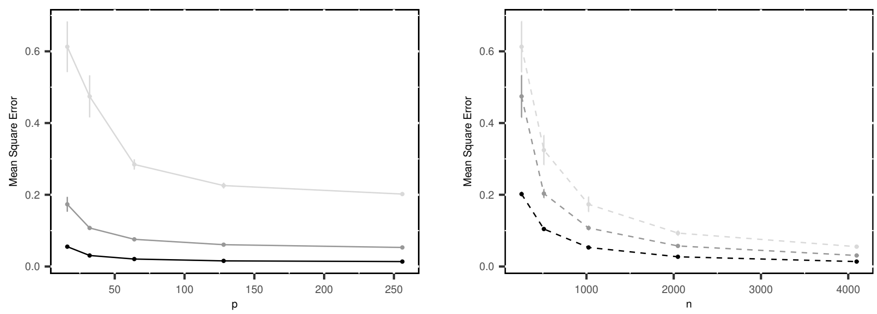

Figure 5: Mean square error for the second eigenfunction η_2^* according to the number of discretization points p (left), and the number of samples n (right). Left: the number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). Right: the number of discretization points is $p \in \{16, 32, 256\}$ (light gray, gray, black respectively). The signal to noise ratio is 0.25.

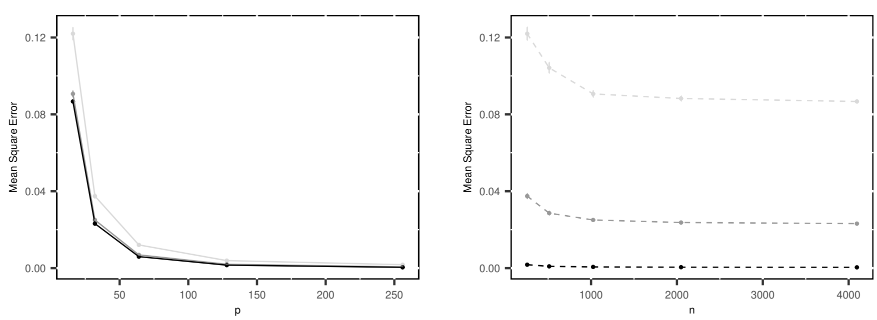

Figure 6: Mean square error for the second eigenvalue μ_2^* according to the number of discretization points p (left), and the number of samples n (right). Left: the number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). Right: the number of discretization points is $p \in \{16, 32, 256\}$ (light gray, gray, black respectively). The signal to noise ratio is 0.25.

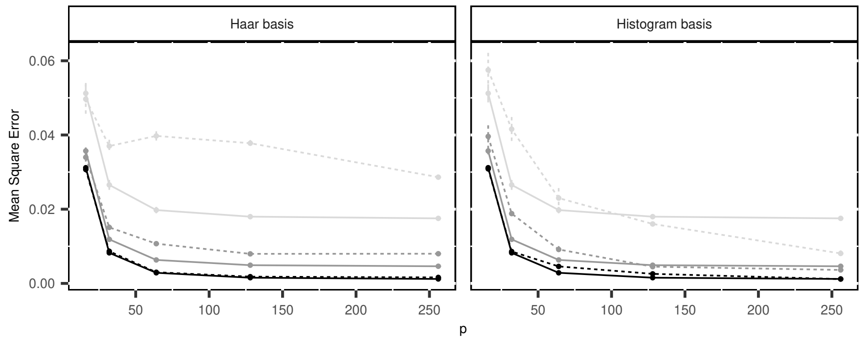

Figure 7: Mean square error for the first eigenfunction η_1^* according to the number of discretization points p and the smoothing system (Haar, left panel, histograms, right panel). The number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). Dashed line: regularized estimator based on cross validation, plain line: non regularized estimator. The signal to noise ratio is 0.25. Regularization is performed by cross-validation to choose the number of wavelet coefficients or bins for histograms, as detailed in Appendix A.

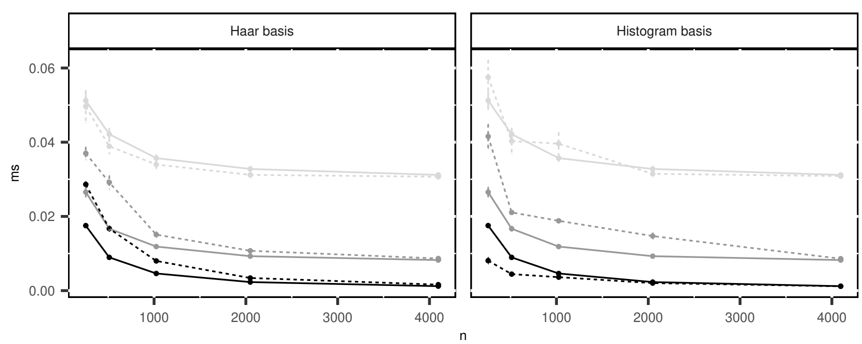

Figure 8: Mean square error for the first eigenfunction η_1^* according to the number samples n and the smoothing system (Haar, left panel, histograms, right panel). The number of discretization points is $p \in \{16, 32, 256\}$ (light gray, gray, black respectively). Dashed line: regularized estimator based on cross validation, plain line: non regularized estimator. The signal to noise ratio is 0.25. Regularization is performed by cross-validation to choose the number of wavelet coefficients or bins for histograms, as detailed in Appendix A. .

B Proof of Theorem 1

To establish Theorem 1, we prove following Propositions 2 and 3.

Proposition 2 *Assume that the rank of the operator Γ is larger than 2 and $p \geq 4$. Then,*

$$\inf_{\widehat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\widehat{\eta}_1 - \eta_1^*\|^2] \geq c_1 n^{-1}, \quad \inf_{\widehat{\eta}_2} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\widehat{\eta}_2 - \eta_2^*\|^2] \geq c_1 n^{-1},$$

where $c_1 > 0$ is a constant depending on L, α and σ.

Proposition 3 *There exists a universal constant $c_2 > 0$ such that*

$$\inf_{\widehat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\widehat{\eta}_1 - \eta_1^*\|^2] \geq c_2 p^{-2\alpha}.$$

The result of Theorem 1 is deduced from Propositions 2 and 3, by taking

$$c = \frac{1}{2} \min(c_1; c_2) > 0.$$

Studying the case $p \in \{1, 2, 3\}$ separately, which is easy, provides the result of Theorem 1 for any $n \geq 1$ and any $p \geq 1$.

B.1 Proof of Proposition 2

Proof: Denoting by

$$\mathcal{R}_\alpha^{\mathcal{N}}(L) = \{P_Z \in \mathcal{R}_\alpha(L) \text{ and } Z \text{ is a Gaussian process}\},$$

we remark that $\mathcal{R}_\alpha^{\mathcal{N}}(L) \subset \mathcal{R}_\alpha(L)$, then, for $d = 1, 2$,

$$\inf_{\widehat{\eta}_d} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\widehat{\eta}_d - \eta_d^*\|^2] \geq \inf_{\widehat{\eta}_d} \sup_{P_Z \in \mathcal{R}_\alpha^{\mathcal{N}}(L)} \mathbb{E}[\|\widehat{\eta}_d - \eta_d^*\|^2].$$

Hence, we can restrict the proof of our lower bound to the case of Gaussian processes.

For $t \in [0, 1]$, let

$$a(t) = \sqrt{2}\cos(2\pi t), \quad b(t) = \sqrt{2}\sin(2\pi t).$$

We observe that

$$\|a\| = \|b\| = 1, \quad \langle a, b \rangle = 0.$$

We then define, for $t \in [0, 1]$,

$$\eta_{A,0}(t) = a(t), \quad \eta_{B,0}(t) = b(t),$$

$$\eta_{A,1}(t) = c_n \left(a(t) + \frac{b(t)}{\sqrt{n}} \right), \quad \eta_{B,1}(t) = c_n \left(b(t) - \frac{a(t)}{\sqrt{n}} \right)$$

and c_n such that

$$\|\eta_{A,1}\| = \|\eta_{B,1}\| = 1$$

and we obtain

$$c_n^2 = \left(1 + \frac{1}{n}\right)^{-1} < 1.$$

We then have for $j = 0, 1$,

$$\|\eta_{A,j}\| = \|\eta_{B,j}\| = 1, \quad \langle \eta_{A,j}, \eta_{B,j} \rangle = 0.$$

Functions $\eta_{A,j}$ and $\eta_{A,j}$ are also C^∞. Now, we introduce for $j = 0, 1$,

$$Z^j(t) = \sqrt{\mu_A}\xi_A\eta_{A,j}(t) + \sqrt{\mu_B}\xi_B\eta_{B,j}(t), \quad t \in [0,1],$$

where μ_A and μ_B are two positive constants such that $L/(64\pi^2) \geq \mu_A > \mu_B$ and $\xi_A \sim \xi_B \sim \mathcal{N}(0,1)$ with ξ_A and ξ_B independent.

Remark 5 *Observe that for any $t \in [0,1]$,*

$$Z^0(t) \sim \mathcal{N}\big(0, \mu_A a^2(t) + \mu_B b^2(t)\big),$$

$$Z^1(t) \sim \mathcal{N}\left(0, \mu_A c_n^2\Big(a^2(t) + \frac{b^2(t)}{n} + 2\frac{a(t)b(t)}{\sqrt{n}}\Big) + \mu_B c_n^2\Big(b^2(t) + \frac{a^2(t)}{n} - 2\frac{a(t)b(t)}{\sqrt{n}}\Big)\right).$$

We consider Model (2) with $\sigma = 0$ such that $Z_1^0, \ldots, Z_n^0$ (resp. $Z_1^1, \ldots, Z_n^1$) are i.i.d copies of Z^0 (resp. Z^1). It is straightforward to observe that the lower bound established for $\sigma = 0$ provides a lower bound for any $\sigma \geq 0$. We then observe n i.i.d. copies of

$$\begin{cases} Z^0(t_h) = \sqrt{\mu_A}\xi_A\eta_{A,0}(t_h) + \sqrt{\mu_B}\xi_B\eta_{B,0}(t_h) \\ Z^1(t_h) = \sqrt{\mu_A}\xi_A\eta_{A,1}(t_h) + \sqrt{\mu_B}\xi_B\eta_{B,1}(t_h) \end{cases}$$

for $h = 0, \ldots, p-1$. Let, for $j = 0, 1$, P_j^Z the distribution of Z^j. We have for any $(t,u) \in [0,1]^2$,

$$\begin{aligned}\int_{\mathcal{C}^0} (z(t) - z(u))^2 dP_j^Z(z) &= E[(Z^j(t) - Z^j(u))^2] \\ &= \mu_A\left(\eta_{A,j}(t) - \eta_{A,j}(u)\right)^2 + \mu_B\left(\eta_{B,j}(t) - \eta_{B,j}(u)\right)^2 \\ &\leq C\mu_A|t-u|^{2\alpha},\end{aligned}$$

for $C = 64\pi^2$, since

$$|a(t) - a(u)|^2 \leq 8\pi^2|t-u|^2, \quad |b(t) - b(u)|^2 \leq 8\pi^2|t-u|^2$$

implies

$$\int_{\mathcal{C}^0} (z(t) - z(u))^2 dP_0^Z(z) \leq 16\pi^2\mu_A|t-u|^{2\alpha}$$
$$\int_{\mathcal{C}^0} (z(t) - z(u))^2 dP_1^Z(z) \leq 64\pi^2\mu_A|t-u|^{2\alpha}$$

(we have used that $\mu_A > \mu_B$ and $n \geq 1$ and $|t-u|^{1-2\alpha} \leq 1$, since $\alpha \leq 1$ and the mean value theorem). We easily deduce that $P_j^Z \in \mathcal{R}_\alpha^{\mathcal{N}}(L)$ since $64\pi^2\mu_A \leq L$. This allows to deduce that

$$\inf_{\widehat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} E[\|\widehat{\eta}_1 - \eta_1^*\|^2] \geq \inf_{\widehat{\eta}_1} \sup_{j=0,1} E[\|\widehat{\eta}_1 - \eta_{A,j}\|^2].$$

We obtain similarly

$$\inf_{\widehat{\eta}_2} \sup_{P_Z \in \mathcal{R}_\alpha(L)} E[\|\widehat{\eta}_2 - \eta_2^*\|^2] \geq \inf_{\widehat{\eta}_2} \sup_{j=0,1} E[\|\widehat{\eta}_2 - \eta_{B,j}\|^2].$$

We now prove a lower bound for $E[\|\widehat{\eta}_1 - \eta_{A,j}\|^2]$. Let $\widehat{\eta}_1$ an estimator and $\hat{\psi}$ the minimum distance test defined by

$$\hat{\psi} = \arg\min_{j=0,1}\|\widehat{\eta}_1 - \eta_{A,j}\|^2,$$

we have for $j = 0, 1$,

$$\|\widehat{\eta}_1 - \eta_{A,j}\| \geq \frac{1}{2}\|\eta_{A,\hat{\psi}} - \eta_{A,j}\|.$$

Now, we have, with $c_n^2 = n/(n+1)$,

$$\begin{aligned}\|\eta_{A,\hat{\psi}} - \eta_{A,j}\|^2 &= \mathbf{1}_{\{\hat{\psi}\neq j\}}\|\eta_{A,0} - \eta_{A,1}\|^2 = \mathbf{1}_{\{\hat{\psi}\neq j\}}\left\|a - c_n\Big(a + \frac{b}{\sqrt{n}}\Big)\right\|^2 \\ &\geq \mathbf{1}_{\{\hat{\psi}\neq j\}}\Big((1-c_n)^2 + c_n^2/n\Big) \geq \mathbf{1}_{\{\hat{\psi}\neq j\}}\frac{c_n^2}{n} = \mathbf{1}_{\{\hat{\psi}\neq j\}}\frac{1}{n+1},\end{aligned}$$

for any n. Then,

$$\inf_{\widehat{\eta}_1} \sup_{P_Z\in\mathcal{R}_\alpha(L)} \mathbb{E}[\|\widehat{\eta}_1 - \eta_1^*\|^2] \geq \frac{1}{4(n+1)} \times \inf_{\hat{\psi}} \max_{j=0,1} \mathbb{P}(\widehat{\psi} \neq j). \tag{7}$$

We now prove that the quantity $\inf_{\hat{\psi}} \max_{j=0,1} \mathbb{P}(\widehat{\psi} \neq j)$ can be bounded from below by an absolute positive constant. For this purpose, we control the Kullback divergence between the two models. We have the following lemma proved in Section B.1.1.

Lemma 2 *Denoting P_j^{obs} the distribution of the random vector $\mathbf{Z}^{j,obs} := (Z^j(t_0), \dots, Z^j(t_{p-1}))$, $KL(P_1^{obs}, P_0^{obs})$, the Kullback divergence between P_1^{obs} and P_0^{obs} satisfies, if $p \geq 4$,*

$$KL(P_1^{obs}, P_0^{obs}) = \frac{1}{2(n+1)}\Big(\frac{\mu_A}{\mu_B} + \frac{\mu_B}{\mu_A} - 2\Big).$$

The result of the lemma entails that $KL((P_1^{obs})^{\otimes n}, (P_0^{obs})^{\otimes n})$, the Kullback divergence between $(P_1^{obs})^{\otimes n}$ and $(P_0^{obs})^{\otimes n}$ satisfies

$$KL((P_1^{obs})^{\otimes n}, (P_0^{obs})^{\otimes n}) = n \times KL(P_1^{obs}, P_0^{obs}) = \frac{n}{2(n+1)}\Big(\frac{\mu_A}{\mu_B} + \frac{\mu_B}{\mu_A} - 2\Big).$$

and then is bounded by a constant κ only depending on μ_A and μ_B. Therefore, Theorem 2.2 of Tsybakov [2009] shows that

$$\inf_{\hat{\psi}} \max_{j=0,1} \mathbb{P}(\widehat{\psi} \neq j) \geq \max\left\{\frac{1}{4}\exp(-\kappa), \frac{1 - \sqrt{\kappa/2}}{2}\right\} > 0.$$

and Inequality (7) provides the desired lower bound. In the same way, we obtain a lower bound for $\inf_{\widehat{\eta}_2} \sup_{P_Z\in\mathcal{R}_\alpha(L)} E[\|\widehat{\eta}_2 - \eta_2^*\|^2]$.

B.1.1 Proof of Lemma 2

We first remark that

$$\mathbf{Z}^{j,obs} \sim \mathcal{N}(0, G_j),$$

where $G_j = ([G_j]_{k,\ell})_{0\leq k,\ell\leq p-1}$ and

$$[G_j]_{k,\ell} = \mathbb{E}[Z_j(t_k)Z_j(t_\ell)] = \mu_A\eta_{A,j}(t_k)\eta_{A,j}(t_\ell) + \mu_B\eta_{B,j}(t_k)\eta_{B,j}(t_\ell).$$

Let us explicit G_0 and G_1. We have

$$\begin{aligned}[G_0]_{k,\ell} &= \mu_A\eta_{A,0}(t_k)\eta_{A,0}(t_\ell) + \mu_B\eta_{B,0}(t_k)\eta_{B,0}(t_\ell) \\ &= \mu_A a(t_k)a(t_\ell) + \mu_B b(t_k)b(t_\ell) \\ &= 2\mu_A\cos(2\pi t_k)\cos(2\pi t_\ell) + 2\mu_B\sin(2\pi t_k)\sin(2\pi t_\ell)\end{aligned}$$

and

$$
\begin{aligned}
[G_1]_{k,\ell} &= \mu_A \eta_{A,1}(t_k)\eta_{A,1}(t_\ell) + \mu_B \eta_{B,1}(t_k)\eta_{B,1}(t_\ell) \\
&= \mu_A c_n^2 \Big(a(t_k) + \frac{b(t_k)}{\sqrt{n}} \Big) \Big(a(t_\ell) + \frac{b(t_\ell)}{\sqrt{n}} \Big) + \mu_B c_n^2 \Big(b(t_k) - \frac{a(t_k)}{\sqrt{n}} \Big) \Big(b(t_\ell) - \frac{a(t_\ell)}{\sqrt{n}} \Big) \\
&= 2\mu_A c_n^2 \Big(\cos(2\pi t_k) + \frac{\sin(2\pi t_k)}{\sqrt{n}} \Big) \Big(\cos(2\pi t_\ell) + \frac{\sin(2\pi t_\ell)}{\sqrt{n}} \Big) \\
&\qquad + 2\mu_B c_n^2 \Big(\sin(2\pi t_k) - \frac{\cos(2\pi t_k)}{\sqrt{n}} \Big) \Big(\sin(2\pi t_\ell) - \frac{\cos(2\pi t_\ell)}{\sqrt{n}} \Big).
\end{aligned}
$$

We now determine eigenelements of G_0 and G_1. For this purpose, we use the following lemma proved in Section B.1.2.

Lemma 3 *Let $p \geq 3$. With $t_k = k/(p-1)$, for $k = 0, \ldots, p-1$, we kave*

$$
\sum_{k=0}^{p-1} \cos(2\pi t_k) = 1, \quad \sum_{k=0}^{p-1} \sin(2\pi t_k) = 0, \quad \sum_{k=0}^{p-1} \sin(4\pi t_k) = 0.
$$

Furthermore, for $p \geq 4$,

$$
\sum_{k=0}^{p-1} \cos^2(2\pi t_k) = \frac{p+1}{2}, \quad \sum_{k=0}^{p-1} \sin^2(2\pi t_k) = \frac{p-1}{2}.
$$

We set $e = (e_k)_{0 \leq k \leq p-1}$ and $f = (f_k)_{0 \leq k \leq p-1}$ with

$$
e_k = \sqrt{\frac{2}{p+1}} \cos(2\pi t_k) = \frac{a(t_k)}{\sqrt{p+1}}, \quad f_k = \sqrt{\frac{2}{p-1}} \sin(2\pi t_k) = \frac{b(t_k)}{\sqrt{p-1}}.
$$

Lemma 3 shows that

$$
\|e\|_{\ell_2} = \|f\|_{\ell_2} = 1, \quad \langle e, f \rangle_{\ell_2} = 0.
$$

We then complete (e, f) so that we have an orthonormal basis of $\mathbb{R}^p$, denoted $\mathcal{B}$. We observe that

$$
[G_0]_{k,\ell} = \mu_A (p+1) e_k e_l + \mu_B (p-1) f_k f_\ell,
$$

which entails

$$
G_0 e = \mu_A (p+1) e, \quad G_0 f = \mu_B (p-1) f.
$$

Similarly,

$$
\begin{aligned}
[G_1]_{k,\ell} = \mu_A c_n^2 &\left(\sqrt{p+1} e_k + \sqrt{\frac{p-1}{n}} f_k \right) \left(\sqrt{p+1} e_\ell + \sqrt{\frac{p-1}{n}} f_\ell \right) \\
&+ \mu_B c_n^2 \left(\sqrt{p-1} f_k - \sqrt{\frac{p+1}{n}} e_k \right) \left(\sqrt{p-1} f_\ell - \sqrt{\frac{p+1}{n}} e_\ell \right),
\end{aligned}
$$

which entails

$$
G_1 e = \left(\mu_A c_n^2 (p+1) + \mu_B c_n^2 \frac{p+1}{n} \right) e + \left(\mu_A c_n^2 \sqrt{\frac{p^2-1}{n}} - \mu_B c_n^2 \sqrt{\frac{p^2-1}{n}} \right) f
$$

and

$$
G_1 f = \left(\mu_A c_n^2 \sqrt{\frac{p^2-1}{n}} - \mu_B c_n^2 \sqrt{\frac{p^2-1}{n}} \right) e + \left(\mu_A c_n^2 \frac{p-1}{n} + \mu_B c_n^2 (p-1) \right) f.
$$

We have shown that $\mathbf{Z}^{0,obs}$ and $\mathbf{Z}^{1,obs}$ are supported by the hyperplane spanned by e and f and the variance-covariance matrices for $\mathbf{Z}^{0,obs}$ and $\mathbf{Z}^{1,obs}$ expressed on (e, f) are respectively

$$G_{red,0} = \begin{pmatrix} \mu_A(p+1) & 0 \\ 0 & \mu_B(p-1) \end{pmatrix}$$

and

$$G_{red,1} = c_n^2 \begin{pmatrix} (p+1)\big(\mu_A + \frac{\mu_B}{n}\big) & \sqrt{\frac{p^2-1}{n}}(\mu_A - \mu_B) \\ \sqrt{\frac{p^2-1}{n}}(\mu_A - \mu_B) & (p-1)\big(\frac{\mu_A}{n} + \mu_B\big) \end{pmatrix}.$$

We recall

$$KL(P_1^{obs}, P_0^{obs}) = \frac{1}{2}\left(\log\left(\frac{\det(G_{red,0})}{\det(G_{red,1})}\right) - 2 + \text{Trace}(G_{red,0}^{-1}G_{red,1})\right).$$

see for instance Equation (A23) of Rasmussen and Williams [2006]. We have:

$$\det(G_{red,0}) = \mu_A\mu_B(p^2-1)$$

and

$$\begin{aligned}\det(G_{red,1}) &= c_n^4(p^2-1)\left(\big(\mu_A + \frac{\mu_B}{n}\big)\big(\frac{\mu_A}{n} + \mu_B\big) - \frac{(\mu_A-\mu_B)^2}{n}\right) \\ &= \mu_A\mu_B(p^2-1)c_n^4\Big(1+\frac{1}{n}\Big)^2 \\ &= \mu_A\mu_B(p^2-1).\end{aligned}$$

Finally,

$$G_{red,0}^{-1} = \begin{pmatrix} \mu_A^{-1}(p+1)^{-1} & 0 \\ 0 & \mu_B^{-1}(p-1)^{-1} \end{pmatrix}$$

and

$$G_{red,0}^{-1}G_{red,1} = c_n^2 \begin{pmatrix} \big(1+\frac{\mu_B}{\mu_A n}\big) & \sqrt{\frac{p-1}{n(p+1)}}\big(1-\frac{\mu_B}{\mu_A}\big) \\ -\sqrt{\frac{p+1}{n(p-1)}}\big(1-\frac{\mu_A}{\mu_B}\big) & \big(1+\frac{\mu_A}{\mu_B n}\big) \end{pmatrix},$$

which yields, with $c_n^2 = n/(n+1)$,

$$KL(P_1^{obs}, P_0^{obs}) = \frac{1}{2}\left(-2 + 2c_n^2 + \frac{c_n^2}{n}\Big(\frac{\mu_A}{\mu_B} + \frac{\mu_B}{\mu_A}\Big)\right) = \frac{1}{2(n+1)}\Big(\frac{\mu_A}{\mu_B} + \frac{\mu_B}{\mu_A} - 2\Big).$$

Lemma 2 is proved.

B.1.2 Proof of Lemma 3

Let $x \in (0, 2\pi)$. We have:

$$\sum_{k=0}^{p-1} e^{ixk} = \frac{1-e^{ixp}}{1-e^{ix}} = \frac{e^{ixp/2}\big(e^{-ixp/2} - e^{ixp/2}\big)}{e^{ix/2}\big(e^{-ix/2} - e^{ix/2}\big)} = e^{ix(p-1)/2}\frac{\sin(xp/2)}{\sin(x/2)}.$$

Let $p \geq 3$. We take $x = 2\pi/(p-1)$ which lies in $(0, 2\pi)$. Considering the real and imaginary parts, we obtain:

$$\sum_{k=0}^{p-1}\cos(2\pi t_k) = \cos(\pi)\frac{\sin(\pi p/(p-1))}{\sin(\pi/(p-1))} = 1$$

and

$$\sum_{k=0}^{p-1} \sin(2\pi t_k) = \sin(\pi) \frac{\sin(\pi p/(p-1))}{\sin(\pi/(p-1))} = 0.$$

Similarly, if $p \geq 4$, with $x = 4\pi/(p-1)$ which lies in $(0, 2\pi)$, we have

$$\sum_{k=0}^{p-1} \sin(2\pi t_k) = 0.$$

This result remains true for $p = 3$. Now, we have, for $p \geq 4$,

$$\sum_{k=0}^{p-1} \cos^2(2\pi t_k) = \sum_{k=0}^{p-1} \frac{\cos(4\pi t_k) + 1}{2} = \frac{\frac{\sin(2\pi p/(p-1))}{\sin(2\pi/(p-1))} + p}{2} = \frac{p+1}{2}$$

and

$$\sum_{k=0}^{p-1} \sin^2(2\pi t_k) = \sum_{k=0}^{p-1} \Big(1 - \cos^2(2\pi t_k)\Big) = \frac{p-1}{2}.$$

Lemma 3 is proved.

B.2 Proof of Proposition 3

The proof is based on Assouad's Lemma and follows the general scheme described in Tsybakov [2009, Sections 2.6 and 2.7]. Let

$$\phi(t) = e^{-\frac{1}{1-t^2}} 1_{(-1,1)}(t).$$

We then define

$$\varphi(t) = \begin{cases} \phi(4t-1) & \text{if } t \in [0, 1/2), \\ -\phi(4t+1) & \text{if } t \in (-1/2, 0], \\ 0 & \text{if } t \notin (-1/2, 1). \end{cases}$$

Both functions ϕ and φ are C^∞ on $\mathbb{R}$ with bounded support, then are α-Hölder continuous, for all $\alpha > 0$. The function φ has its support included in $(-1/2, 1/2)$ and verifies $\int_{-1/2}^{1/2} \varphi(t)dt = 0$. We note L_α such that, for all $t, u \in \mathbb{R}$,

$$|\varphi(t) - \varphi(u)| \leq L_\alpha |t-u|^\alpha.$$

Let us now define test eigenfunctions. For $\boldsymbol{\omega} = (w_0, \ldots, w_{p-1}) \in \{0,1\}^p$, we set

$$\eta^*_{1,\boldsymbol{\omega}}(t) = C_{\boldsymbol{\omega}} \left(\gamma + \sum_{k=0}^{p-1} \omega_k \left(p^{-\alpha} \varphi\left(p(t-t_k) - 1/2\right)\right) \right),$$

with $C_{\boldsymbol{\omega}}$ and $\gamma > 0$ two positive constants to be specified later. To be an eigenfunction, $\eta^*_{1,\boldsymbol{\omega}}$ has to be of norm 1, which writes

$$\begin{aligned} \|\eta^*_{1,\boldsymbol{\omega}}\|^2 &= C^2_{\boldsymbol{\omega}} \int_0^1 \left(\gamma + \sum_{k=0}^{p-1} \omega_k \left(p^{-\alpha} \varphi(p(t-t_k) - 1/2)\right) \right)^2 dt \\ &= C^2_{\boldsymbol{\omega}} \left(\gamma^2 + 2\gamma \sum_{k=0}^{p-1} \omega_k \left(p^{-\alpha} \int_0^1 \varphi(p(t-t_k) - 1/2) dt \right) \right. \\ &\quad \left. + \int_0^1 \left(\sum_{k=0}^{p-1} \omega_k \left(p^{-\alpha} \varphi(p(t-t_k) - 1/2)\right) \right)^2 dt \right). \end{aligned}$$

Using successively that the support of φ is in $(-1/2, 1/2)$ and that $\int_{-1/2}^{1/2} \varphi(t)dt = 0$, we have

$$\int_0^1 \varphi(p(t-t_k)-1/2)dt = \int_{t_k}^{t_{k+1}} \varphi(p(t-t_k)-1/2)dt = p^{-1}\int_{-1/2}^{1/2} \varphi(t)dt = 0,$$

and

$$\int_0^1 \left(\sum_{k=0}^{p-1} \omega_k \varphi(p(t-t_k)-1/2)\right)^2 dt = \sum_{k=0}^{p-1} \omega_k \int_0^1 \varphi^2(p(t-t_k)-1/2)dt = p^{-1}\sum_{k=0}^{p-1} \omega_k \|\varphi\|^2.$$

This implies that

$$\|\eta_{1,\boldsymbol{\omega}}^*\|^2 \quad = \quad C_{\boldsymbol{\omega}}^2 \left(\gamma^2 + p^{-2\alpha-1}\|\varphi\|^2 \sum_{k=0}^{p-1} \omega_k\right).$$

We then fix the quantity

$$C_{\boldsymbol{\omega}} = \left(\gamma^2 + p^{-2\alpha-1}\|\varphi\|^2 \sum_{k=0}^{p-1} \omega_k\right)^{-1/2},$$

so that $\|\eta_{1,\boldsymbol{\omega}}^*\| = 1$ and observe that $C_{\boldsymbol{\omega}}$ verifies

$$\left(\gamma^2 + \|\varphi\|^2\right)^{-1/2} \le \left(\gamma^2 + p^{-2\alpha}\|\varphi\|^2\right)^{-1/2} \le C_{\boldsymbol{\omega}} \le \gamma^{-1}.$$

We now define the associated distribution of our observations: for ξ a centered random variable with variance 1 and $\mu_{1,\boldsymbol{\omega}}^* = \frac{L}{2L_\alpha^2 C_{\boldsymbol{\omega}}^2}$, we set

$$Z_{\boldsymbol{\omega}}(t) = \sqrt{\mu_{1,\boldsymbol{\omega}}^*}\,\xi \eta_{1,\boldsymbol{\omega}}^*(t). \tag{8}$$

Let $P_{\boldsymbol{\omega}}^Z$ be the distribution of $Z_{\boldsymbol{\omega}}$. We have that $P_{\boldsymbol{\omega}}^Z \in \mathcal{R}_\alpha(L)$ since

$$\begin{aligned}
\int_{C([0,1])} (z(t)-z(s))^2 dP_{\boldsymbol{\omega}}^Z(z) &= \mathbb{E}[(Z_{\boldsymbol{\omega}}(t) - Z_{\boldsymbol{\omega}}(s))^2] = \mu_{1,\boldsymbol{\omega}}^*(\eta_{1,\boldsymbol{\omega}}(t) - \eta_{1,\boldsymbol{\omega}}(s))^2\mathbb{E}[\xi^2] \\
&= \mu_{1,\boldsymbol{\omega}}^*(\eta_{1,\boldsymbol{\omega}}(t) - \eta_{1,\boldsymbol{\omega}}(s))^2 \\
&= \mu_{1,\boldsymbol{\omega}}^* C_{\boldsymbol{\omega}}^2 \left(\sum_{k=0}^{p-1} \omega_k p^{-\alpha}(\varphi(p(t-t_k)-1/2) - \varphi(p(s-t_k)-1/2))\right)^2.
\end{aligned}$$

Then, using the properties of φ, we have two cases:

- If $s, t \in [t_\ell, t_{\ell+1}[$ for some $\ell \in \{0, \dots, p-1\}$,

$$\begin{aligned}
&\left(\sum_{k=0}^{p} \omega_k p^{-\alpha}(\varphi(p(t-t_\ell)-1/2) - \varphi(p(s-t_\ell)-1/2))\right)^2 \\
&\quad = \omega_\ell^2 p^{-2\alpha}(\varphi(p(t-t_\ell)-1/2) - \varphi(p(s-t_\ell)-1/2))^2 \\
&\quad \le p^{-2\alpha} L_\alpha^2 |p(t-t_\ell) - p(s-t_\ell)|^{2\alpha} = L_\alpha^2 |t-s|^{2\alpha}.
\end{aligned}$$

- If $s \in [t_\ell, t_{\ell+1}[$ and $t \in [t_{\ell'}, t_{\ell'+1}[$ with $\ell \neq \ell'$,

$$\left(\sum_{k=0}^{p} \omega_k p^{-\alpha} (\varphi(p(t-t_k)-1/2) - \varphi(p(s-t_k)-1/2))\right)^2$$
$$= \omega_\ell^2 p^{-2\alpha} |\varphi(p(t-t_\ell)-1/2) - \varphi(p(s-t_\ell)-1/2)|^2.$$
$$+\omega_{\ell'}^2 p^{-2\alpha} |\varphi(p(t-t_{\ell'})-1/2) - \varphi(p(s-t_{\ell'})-1/2)|^2$$
$$\leq 2L_\alpha^2 |t-s|^{2\alpha}.$$

Finally

$$\int_{C([0,1])} (z(t)-z(s))^2 dP_{\boldsymbol{\omega}}(z) \leq 2\mu_{1,\boldsymbol{\omega}}^* C_{\boldsymbol{\omega}}^2 L_\alpha^2 |t-s|^{2\alpha} = L|t-s|^{2\alpha}.$$

This allows to deduce that

$$\inf_{\widehat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\widehat{\eta}_1 - \eta_1^*\|^2] \geq \inf_{\widehat{\eta}_1} \sup_{\boldsymbol{\omega} \in \{0,1\}^p} \mathbb{E}[\|\widehat{\eta}_1 - \eta_{1,\boldsymbol{\omega}}^*\|^2],$$

and the aim of what follows is to prove a lower bound for $\mathbb{E}[\|\widehat{\eta}_1 - \eta_{1,\boldsymbol{\omega}}^*\|^2]$.

Let $\widehat{\eta}_1$ an estimator and

$$\widehat{\boldsymbol{\omega}} \in \arg\min_{\boldsymbol{\omega} \in \{0,1\}^p} \|\widehat{\eta}_1 - \eta_{1,\boldsymbol{\omega}}^*\|^2,$$

we have

$$\|\widehat{\eta}_1 - \eta_{1,\widehat{\boldsymbol{\omega}}}^*\| \geq \frac{1}{2} \|\eta_{1,\widehat{\boldsymbol{\omega}}}^* - \eta_{1,\boldsymbol{\omega}}^*\|.$$

Now, still from the support properties of φ,

$$\|\eta_{1,\widehat{\boldsymbol{\omega}}}^* - \eta_{1,\boldsymbol{\omega}}^*\|^2$$
$$= \sum_{k=0}^{p-1} \int_{t_k}^{t_{k+1}} \left(C_{\widehat{\boldsymbol{\omega}}}(\gamma + \widehat{\omega}_k p^{-\alpha} \varphi(p(t-t_k)-1/2)) - C_{\boldsymbol{\omega}}(\gamma + \omega_k p^{-\alpha} \varphi(p(t-t_k)-1/2))\right)^2 dt$$
$$= p^{-1} \sum_{k=0}^{p-1} \int_{-1/2}^{1/2} \left(C_{\widehat{\boldsymbol{\omega}}}(\gamma + \widehat{\omega}_k p^{-\alpha} \varphi(u)) - C_{\boldsymbol{\omega}}(\gamma + \omega_k p^{-\alpha} \varphi(u))\right)^2 du$$
$$= (C_{\widehat{\boldsymbol{\omega}}} - C_{\boldsymbol{\omega}})^2 \gamma^2 + \|\varphi\|^2 p^{-2\alpha-1} \sum_{k=0}^{p-1} (C_{\widehat{\boldsymbol{\omega}}} \widehat{\omega}_k - C_{\boldsymbol{\omega}} \omega_k)^2 \geq \|\varphi\|^2 p^{-2\alpha-1} \sum_{k=0}^{p-1} (C_{\widehat{\boldsymbol{\omega}}} \widehat{\omega}_k - C_{\boldsymbol{\omega}} \omega_k)^2$$
$$\geq \|\varphi\|^2 p^{-2\alpha-1} \min\{C_{\widehat{\boldsymbol{\omega}}}^2, C_{\boldsymbol{\omega}}^2\} \sum_{k=0}^{p-1} \mathbf{1}_{\{\widehat{\omega}_k \neq \omega_k\}}$$
$$\geq (\gamma^2 + \|\varphi\|^2)^{-1} \|\varphi\|^2 p^{-2\alpha-1} \rho(\widehat{\boldsymbol{\omega}}, \boldsymbol{\omega}),$$

where $\rho(\boldsymbol{\omega}, \boldsymbol{\omega}') = \sum_{k=0}^{p-1} \mathbf{1}_{\omega_k \neq \omega_k'}$ is the Hamming distance on $\{0,1\}^p$.

Combining all the inequalities above, we have the existence of a constant $\tilde{c} = \|\varphi\|^2/(4(\gamma^2 + \|\varphi\|^2))$ such that

$$\inf_{\widehat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\widehat{\eta}_1 - \eta_{1,\boldsymbol{\omega}}^*\|^2] \quad \geq \quad \tilde{c} p^{-2\alpha-1} \inf_{\widehat{\boldsymbol{\omega}}} \max_{\boldsymbol{\omega} \in \{0,1\}^p} \mathbb{E}[\rho(\widehat{\boldsymbol{\omega}}, \boldsymbol{\omega})].$$

By Assouad's lemma (see e.g. Tsybakov, 2009, Theorem 2.12), there exists a constant $c > 0$ such that

$$\inf_{\widehat{\boldsymbol{\omega}}} \max_{\boldsymbol{\omega} \in \{0,1\}^p} \mathbb{E}[\rho(\widehat{\boldsymbol{\omega}}, \boldsymbol{\omega})] \geq cp, \tag{9}$$

provided we are able to prove that for some constant $K_{\max} \geq 0$,

$$KL((P_{\boldsymbol{\omega}}^{obs})^{\otimes n}, (P_0^{obs})^{\otimes n}) \leq K_{\max}, \text{ for all } \boldsymbol{\omega} \in \{0,1\}^p,$$

where $P_{\boldsymbol{\omega}}^{obs}$ is the law of the random vector

$$\mathbf{Y}_{\boldsymbol{\omega}}^{obs} := (Y_{\boldsymbol{\omega}}(t_0), \ldots, Y_{\boldsymbol{\omega}}(t_{p-1}))$$

such that

$$Y_{\boldsymbol{\omega}}(t_j) = Z_{\boldsymbol{\omega}}(t_j) + \varepsilon_j$$

with $\varepsilon_0, \ldots, \varepsilon_{p-1} \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$ and $KL(P, Q)$ is the Kullback-Leibler divergence between two measures P and Q. In (9), the constant c only depends on $K_{\max}$. We observe that, for all $\boldsymbol{\omega} \in \{0,1\}^p$, for all $j = 0, \ldots, p-1$,

$$Y_{\boldsymbol{\omega}}(t_j) = Z_{\boldsymbol{\omega}}(t_j) + \varepsilon_j = \sqrt{\mu_{1,\boldsymbol{\omega}}^*}\xi\eta_{1,\boldsymbol{\omega}}^*(t_j) + \varepsilon_j.$$

Now

$$\eta_{1,\boldsymbol{\omega}}^*(t_j) = C_{\boldsymbol{\omega}}\left(\gamma + \sum_{k=0}^{p-1}\omega_k(p^{-\alpha}\varphi(p(t_j - t_k) - 1/2))\right) = C_{\boldsymbol{\omega}}\gamma,$$

since $\varphi((p(t_j - t_k) - 1/2) = \varphi(-1/2) = 0$ if $j = k$ and $\varphi((p(t_j - t_k) - 1/2) = 0$ if $j \neq k$ by the support properties of φ and the fact that

$$p(t_j - t_k) - 1/2 = \frac{p}{p-1}(j-k) - 1/2 \geq \frac{p}{p-1} - 1/2 \geq 1/2$$

if $j > k$ and $p(t_j - t_k) - 1/2 \leq 1/2$ if $j < k$. Hence

$$Y_{\boldsymbol{\omega}}(t_j) = \sqrt{\mu_{1,\boldsymbol{\omega}}^*}\xi C_{\boldsymbol{\omega}}\gamma + \varepsilon_j = \frac{\gamma\sqrt{L}}{L_\alpha\sqrt{2}}\xi + \varepsilon_j$$

and the distribution of $\mathbf{Y}_{\boldsymbol{\omega}}^{obs}$ does not depend on $\boldsymbol{\omega}$. Therefore,

$$KL((P_{\boldsymbol{\omega}}^{obs})^{\otimes n}, (P_0^{obs})^{\otimes n}) = nKL(P_{\boldsymbol{\omega}}^{obs}, P_0^{obs}) = 0.$$

C Proof of Theorems 2 and 3

C.1 Preliminary result

The proof of Theorems 2 and 3 is based on Bosq inequalities stated in the following theorem.

Theorem 4 (Bosq 2000) *Let Γ and $\widehat{\Gamma}$ be two linear compact operators on $\mathbb{L}^2([0,1])$. We denote by*

$$\Gamma = \sum_{d=1}^{\infty}\mu_d^*\eta_d^* \otimes \eta_d^* \quad \textit{and} \quad \widehat{\Gamma} = \sum_{d=1}^{\infty}\widehat{\mu}_d\widehat{\eta}_d \otimes \widehat{\eta}_d$$

their spectral decomposition with the eigenvalues $(\mu_d^*)_{d\geq 1}$ and $(\widehat{\mu}_d)_{d\geq 1}$ sorted in decreasing order. Then

$$|\widehat{\mu}_d - \mu_d^*| \leq \||\widehat{\Gamma} - \Gamma\||. \tag{10}$$

Suppose moreover that, for $d \geq 1$, the eigenspace associated to the eigenfunction η_d^* is one-dimensional and denote, to avoid sign confusion, $\eta_{\pm,d}^* = sign(\langle\widehat{\eta}_{\phi,d}, \eta_d^*\rangle) \times \eta_d^*$. Then, we have

$$\|\widehat{\eta}_d - \eta_{\pm,d}^*\| \leq b_d^{1/2}\||\widehat{\Gamma} - \Gamma\||, \tag{11}$$

where

$$b_1 = 8(\mu_1^* - \mu_2^*)^{-2}$$

and for any $d \in \{2, \dots, \mathcal{D}\}$

$$b_d = 8/\min(\mu_d^* - \mu_{d+1}^*, \mu_{d-1}^* - \mu_d^*)^2.$$

The proof of Theorem 4 comes directly from Bosq [2000, Lemma 4.2, p. 103] for the upper bound (10) on the eigenvalues and Bosq [2000, Lemma 4.3, p.104] for the upper bound (11) on the eigenfunctions.

We remark that Γ, $\widehat{\Gamma}$ and $\widehat{\Gamma}_\phi$ are integral operators with kernel respectively K

$$\widehat{K}(s,t) = \frac{1}{n}\sum_{i=1}^{n} Z_i(t) Z_i(s), \quad (s,t) \in [0,1]^2.$$

and

$$\widehat{K}_\phi(s,t) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{Y}_i(t) \widetilde{Y}_i(s), \quad (s,t) \in [0,1]^2.$$

We use the previous result to establish the following proposition.

Proposition 4 *Setting* $K_\phi = E[\widehat{K}_\phi]$*, we have*

$$\|\widehat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2 \leq 5 b_d \left[|||\widehat{\Gamma}_\phi - \Gamma_\phi|||^2 + |||\Pi_D \Gamma \Pi_D - \Gamma|||^2 + \frac{\sigma^4}{p^2} + A_p^{(K)}(\phi, D) + A_p^{(\sigma)}(\phi, D) \right]. \tag{12}$$

Proof (of Proposition 4) *In the sequel, we denote* $\Gamma_\phi = E[\widehat{\Gamma}_\phi]$ *and remark that* Γ_ϕ *is an integral operator with kernel* K_ϕ*.*

$$\begin{aligned} K_\phi(s,t) &= \sum_{\lambda,\lambda' \in \Lambda_D} \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) \phi_\lambda(s) \phi_{\lambda'}(t) \\ &\quad + \frac{\sigma^2}{p^2} \sum_{\lambda,\lambda' \in \Lambda_D} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) \phi_\lambda(s) \phi_{\lambda'}(t) \\ &= \Pi_{S_D^2} K(s,t) + \frac{\sigma^2}{p} \sum_{\lambda \in \Lambda_D} \phi_\lambda(s) \phi_\lambda(t) + R^{(K)}(s,t) + R^{(\sigma)}(s,t), \end{aligned} \tag{13}$$

where $\Pi_{S_D^2}$ *is the orthogonal projection onto* $S_D^2 = span\{(s,t) \mapsto \phi_\lambda(s)\phi_{\lambda'}(t), \lambda, \lambda' \in \Lambda_D\}$*,*

$$\begin{aligned} R^{(K)}(s,t) &= \sum_{\lambda,\lambda' \in \Lambda_D} \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) \phi_\lambda(s) \phi_{\lambda'}(t) - \Pi_{S_D^2} K(s,t) \\ &= \sum_{\lambda,\lambda' \in \Lambda_D} \left(\frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) - \int_0^1 \int_0^1 K(s,t) \phi_\lambda(s) \phi_{\lambda'}(t) ds dt \right) \phi_\lambda(s) \phi_{\lambda'}(t) \end{aligned}$$

and

$$\begin{aligned} R^{(\sigma)}(s,t) &= \frac{\sigma^2}{p^2} \sum_{\lambda,\lambda' \in \Lambda_D} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) \phi_\lambda(s) \phi_{\lambda'}(t) - \frac{\sigma^2}{p} \sum_{\lambda \in \Lambda_D} \phi_\lambda(s) \phi_\lambda(t) \\ &= \frac{\sigma^2}{p} \sum_{\lambda,\lambda' \in \Lambda_D} \left(\frac{1}{p} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) - \mathbf{1}_{\{\lambda = \lambda'\}} \right) \phi_\lambda(s) \phi_{\lambda'}(t). \end{aligned}$$

Then, from the decomposition of the kernel K_ϕ given in Equation (13), we have for any fonction f and any $t \in [0,1]$,

$$\begin{aligned}\Gamma_\phi(f)(t) &= \int_0^1 K_\phi(s,t)f(s)ds \\ &= \int_0^1 \Pi_{S_D^2} K(s,t)f(s)ds + \frac{\sigma^2}{p}\sum_{\lambda\in\Lambda_D}\int_0^1 \phi_\lambda(s)f(s)ds\,\phi_\lambda(t) + T^{(K)}(f)(t) + T^{(\sigma)}(f)(t) \\ &= \int_0^1 \Pi_{S_D^2} K(s,t)f(s)ds + \frac{\sigma^2}{p}\Pi_D(f)(t) + T^{(K)}(f)(t) + T^{(\sigma)}(f)(t),\end{aligned}$$

where Π_D is the orthogonal projection onto $S_D = \mathrm{span}\{\phi_\lambda, \lambda\in\Lambda_D\}$ and $T^{(K)}$ (resp. $T^{(\sigma)}$) is the integral operator associated to the kernel $R^{(K)}$ (resp. $R^{(\sigma)}$):

$$T^{(K)}(t) := \int_0^1 R^{(K)}(s,t)f(s)ds, \quad T^{(\sigma)}(f)(t) := \int_0^1 R^{(\sigma)}(s,t)f(s)ds.$$

Now,

$$\begin{aligned}\int_0^1 \Pi_{S_D^2} K(s,t)f(s)ds &= \sum_{\lambda,\lambda'\in\Lambda_D}\int_0^1\int_0^1\int_0^1 K(u,v)\phi_\lambda(u)\phi_{\lambda'}(v)dudv\,\phi_\lambda(s)\phi_{\lambda'}(t)f(s)ds \\ &= \sum_{\lambda,\lambda'\in\Lambda_D}\langle\phi_\lambda, f\rangle\int_0^1\int_0^1 K(u,v)\phi_\lambda(u)\phi_{\lambda'}(v)dudv\,\phi_{\lambda'}(t) \\ &= \sum_{\lambda,\lambda'\in\Lambda_D}\langle\phi_\lambda, f\rangle\langle\Gamma(\phi_\lambda),\phi_{\lambda'}\rangle\,\phi_{\lambda'}(t) \\ &= \sum_{\lambda'\in\Lambda_D}\langle\Gamma(\sum_{\lambda\in\Lambda_D}\langle\phi_\lambda, f\rangle\phi_\lambda),\phi_{\lambda'}\rangle\,\phi_{\lambda'}(t) \\ &= \Pi_D(\Gamma(\Pi_D(f)))(t). \end{aligned} \tag{14}$$

Hence, we obtain:

$$\Gamma_\phi = \Pi_D\Gamma\Pi_D + \frac{\sigma^2}{p}\Pi_D + T^{(K)} + T^{(\sigma)}.$$

Now, since the eigenvalues $(\mu_d^*)_{d\geq 1}$ are all distincts, the eigenspace associated to the eigenvalue μ_d^* is one-dimensional and we can apply Theorem 4 to the operators Γ and $\widehat{\Gamma}_\phi$, which yields

$$\begin{aligned}\|\widehat{\eta}_{\phi,d} - \eta^*_{\pm,d}\| &\leq b_d^{1/2}|||\widehat{\Gamma}_\phi - \Gamma||| \\ &\leq b_d^{1/2}\Big(|||\widehat{\Gamma}_\phi - \Gamma_\phi||| + |||\Pi_D\Gamma\Pi_D - \Gamma||| \\ &\qquad + \frac{\sigma^2}{p} + |||T^{(K)}||| + |||T^{(\sigma)}|||\Big). \end{aligned} \tag{15}$$

In the previous inequality, we have used that $|||\Pi_D||| = 1$. We now control each term of the previous inequality. For this purpose, introducing $\|\cdot\|_{HS}$, the Hilbert-Schmidt norm of an operator defined by $\|T\|_{HS}^2 = \sum_{\lambda\in\Lambda}\|Te_\lambda\|^2$ where $(e_\lambda)_{\lambda\in\Lambda}$ is an orthonormal basis of $\mathbb{L}^2$ (recall that the Hilbert-Schmidt norm is independent of the choice of the basis), we have, for all operator $T : \mathbb{L}^2 \mapsto \mathbb{L}^2$, $|||T||| \leq \|T\|_{HS}$ since

$$|||T|||^2 = \sup_{f\in\mathbb{L}^2, f\neq 0}\frac{\|Tf\|^2}{\|f\|^2}$$

and, by Cauchy-Schwarz's Inequality,

$$\begin{aligned}\|Tf\|^2 &= \sum_{\lambda\in\Lambda}\langle Tf, e_\lambda\rangle^2 = \sum_{\lambda\in\Lambda}\left(\sum_{\lambda'\in\Lambda}\langle f, e_{\lambda'}\rangle\langle Te_{\lambda'}, e_\lambda\rangle\right)^2 \\ &\leq \sum_{\lambda\in\Lambda}\left(\sum_{\lambda'\in\Lambda}\langle f, e_{\lambda'}\rangle^2\sum_{\lambda'\in\Lambda}\langle Te_{\lambda'}, e_\lambda\rangle^2\right) = \|f\|^2\sum_{\lambda'\in\Lambda}\|Te_{\lambda'}\|^2 = \|f\|^2\|T\|^2_{HS}.\end{aligned}$$

Moreover, we also remark that if T is a kernel operator associated to a kernel R,

$$\begin{aligned}\|T\|^2_{HS} &= \sum_{\lambda\in\Lambda}\|Te_\lambda\|^2 = \sum_{\lambda\in\Lambda}\left\|\int_0^1 R(s,\cdot)e_\lambda(s)ds\right\|^2 \\ &= \sum_{\lambda\in\Lambda}\int_0^1\left(\int_0^1 R(s,t)e_\lambda(s)ds\right)^2 dt = \int_0^1\int_0^1 R^2(s,t)dsdt = \|R\|^2.\end{aligned}$$

In addition, if the kernel $R \in S^2_D$, i.e. if there exists a matrix $G = (G_{\lambda,\lambda'})_{\lambda,\lambda'\in\Lambda_D}$ such that

$$R(s,t) = \sum_{\lambda,\lambda'\in\Lambda_D} G_{\lambda,\lambda'}\phi_\lambda(s)\phi_{\lambda'}(t),$$

we have $\|R\|_{L^2} = \|G\|_F$, where, for a matrix G,

$$\|G\|_F = \sqrt{Tr(G^TG)} = \Big(\sum_{\lambda,\lambda'\in\Lambda_D} G^2_{\lambda,\lambda'}\Big)^{1/2}$$

is the Frobenius norm of the matrix G. The fourth and fifth terms of Equation (15) *are then bounded by the squared Frobenius norm of the associated matrices and we obtain*

$$\|\widehat{\eta}_{\phi,d} - \eta^*_{\pm,d}\|^2 \leq 5b_d\left[|||\widehat{\Gamma}_\phi - \Gamma_\phi|||^2 + |||\Pi_D\Gamma\Pi_D - \Gamma|||^2 + \frac{\sigma^4}{p^2} + A_p^{(K)}(\phi, D) + A_p^{(\sigma)}(\phi, D)\right].$$

Proposition 4 is proved.

To end the proof of Theorems 2 and 3, it remains to deal with the stochastic term $|||\widehat{\Gamma}_\phi - \Gamma_\phi|||^2$, still bounded by using the Frobenius norm:

$$|||\widehat{\Gamma}_\phi - \Gamma_\phi|||^2 \leq \|\widehat{G}_\phi - G_\phi\|^2_F,$$

where

$$\widehat{G}_\phi := \left(\frac{1}{n}\sum_{i=1}^n \widetilde{Y}_{i,\lambda}\widetilde{Y}_{i,\lambda'}\right)_{\lambda,\lambda'\in\Lambda_D}$$

and $G_\phi = E[\widehat{G}_\phi]$. The upper bound of $E[\|\widehat{G}_\phi - G_\phi\|^2_F]$ gives Theorem 2, whereas Theorem 3 is deduced from the control in probability of $\|\widehat{G}_\phi - G_\phi\|_F$ provided by Proposition 5 below.

C.2 End of the proof of Theorem 2

Lemma 4 *Under Assumption 1, we have:*

$$E[\|\widehat{G}_\phi - G_\phi\|^2_F] \leq \frac{\max(C_1+3;6)}{n}\left(\sum_{\lambda\in\Lambda_D}\left[\sigma^2_\lambda + s^2_\lambda\right]\right)^2.$$

Proof (of Lemma 4) *We have*

$$\begin{aligned}
E[\|\widehat{G}_\phi - G_\phi\|_F^2] &= \sum_{\lambda,\lambda'\in\Lambda_D} E\left[\left(\frac{1}{n}\sum_{i=1}^n\left[\widetilde{Y}_{i,\lambda}\widetilde{Y}_{i,\lambda'} - E[\widetilde{Y}_{i,\lambda}\widetilde{Y}_{i,\lambda'}]\right]\right)^2\right] \\
&= \sum_{\lambda,\lambda'\in\Lambda_D} \mathit{Var}\left(\frac{1}{n}\sum_{i=1}^n \widetilde{Y}_{i,\lambda}\widetilde{Y}_{i,\lambda'}\right) \\
&\le \frac{1}{n}\sum_{\lambda,\lambda'\in\Lambda_D} E[\widetilde{Y}_{1,\lambda}^2\widetilde{Y}_{1,\lambda'}^2] \\
&\le \frac{1}{n}\left(\sum_{\lambda\in\Lambda_D}(E[\widetilde{Y}_{1,\lambda}^4])^{1/2}\right)^2.
\end{aligned}$$

Now, since $\tilde{\varepsilon}_{1\lambda} \sim \mathcal{N}(0,\sigma_\lambda^2)$*, and* $\tilde{z}_{1\lambda} = \frac{1}{p}\sum_{h=0}^{p-1} Z_1(t_h)\phi_\lambda(t_h)$*, we have*

$$\begin{aligned}
E[\widetilde{Y}_{1,\lambda}^4] &= E[(\tilde{z}_{1,\lambda} + \tilde{\varepsilon}_{1,\lambda})^4] \\
&= E[\tilde{z}_{1,\lambda}^4] + 6E[\tilde{z}_{1,\lambda}^2]E[\tilde{\varepsilon}_{1,\lambda}^2] + E[\tilde{\varepsilon}_{1,\lambda}^4] \\
&\le C_1(E[\tilde{z}_{1,\lambda}^2])^2 + 6E[\tilde{z}_{1,\lambda}^2]E[\tilde{\varepsilon}_{1,\lambda}^2] + 3\sigma_\lambda^4 \\
&\le (C_1+3)s_\lambda^4 + 6\sigma_\lambda^4
\end{aligned}$$

and

$$\begin{aligned}
E[\|\widehat{G}_\phi - G_\phi\|_F^2] &\le \frac{1}{n}\left(\sum_{\lambda\in\Lambda_D}\left((C_1+3)s_\lambda^4 + 6\sigma_\lambda^4\right)^{1/2}\right)^2 \\
&\le \frac{\max(C_1+3;6)}{n}\left(\sum_{\lambda\in\Lambda_D}\left[\sigma_\lambda^2 + s_\lambda^2\right]\right)^2.
\end{aligned}$$

This ends the proof of Lemma 4.

Combining the upper bound of the previous lemma with (12) provides the stated result in Theorem 2 .

C.3 End of the proof of Theorem 3

To complete the proof of Theorem 3, we need some technical lemmas. Before stating them, we recall that for all $i = 1,\ldots,n$, we have set

$$\widetilde{Y}_{i,\lambda} = \frac{1}{p}\sum_{h=0}^{p-1} Y_i(t_h)\phi_\lambda(t_h), \quad \tilde{z}_{i,\lambda} = \frac{1}{p}\sum_{h=0}^{p-1} Z_i(t_h)\phi_\lambda(t_h), \quad \tilde{\varepsilon}_{i,\lambda} = \frac{1}{p}\sum_{h=0}^{p-1}\varepsilon_{i,h}\phi_\lambda(t_h)$$

and

$$s_\lambda^2 = \mathrm{Var}(\tilde{z}_{i,\lambda}), \quad \sigma_\lambda^2 = \mathrm{Var}(\tilde{\varepsilon}_{i,\lambda}).$$

In the sequel, we consider $\widetilde{Y}_i = (\widetilde{Y}_{i,\lambda})_{\lambda\in\Lambda_D}$, $\tilde{z}_i = (\tilde{z}_{i,\lambda})_{\lambda\in\Lambda_D}$ and $\tilde{\varepsilon}_i = (\tilde{\varepsilon}_{i,\lambda})_{\lambda\in\Lambda_D}$.

Lemma 5 *Under Assumption 2 , for any* $u \in \mathbb{R}^{|\Lambda_D|}$*,*

$$\|u^T\tilde{z}_1\|_{\psi_2}^2 \le C_2 E[(u^T\tilde{z}_1)^2]. \tag{16}$$

If we consider $\tilde{\varepsilon}_1$ instead of $\tilde{z}_1$, Inequality (16) *holds with an absolute constant instead of C_2. Furthermore,*

$$Tr\Big(E\big[\tilde{z}_1\tilde{z}_1^T\big]\Big) = \sum_{\lambda\in\Lambda_D} s_\lambda^2, \quad Tr\Big(E\big[\tilde{\varepsilon}_1\tilde{\varepsilon}_1^T\big]\Big) = \sum_{\lambda\in\Lambda_D} \sigma_\lambda^2. \tag{17}$$

Proof (of Lemma 5) *Since $\mathbf{Z_1} := \{Z_1(t_0), \dots, Z_1(t_{p-1})\}^T$ is a zero-mean sub-Gaussian vector, the vector $\tilde{z}_1$ is also a zero-mean sub-Gaussian vector. We have, for any $u \in \mathbb{R}^{|\Lambda_D|}$,*

$$\begin{aligned}
\|u^T\tilde{z}_1\|_{\psi_2}^2 &= \left\| \sum_{\lambda\in\Lambda_D} u_\lambda \tilde{z}_{1,\lambda} \right\|_{\psi_2}^2 \\
&= \left\| \sum_{\lambda\in\Lambda_D} u_\lambda \times \frac{1}{p}\sum_{h=0}^{p-1} Z_1(t_h)\phi_\lambda(t_h) \right\|_{\psi_2}^2 \\
&= \left\| v^T\mathbf{Z_1} \right\|_{\psi_2}^2,
\end{aligned}$$

with $v = (v_h)_{h=0,\dots,p-1}$ and $v_h := \frac{1}{p}\sum_{\lambda\in\Lambda_D} u_\lambda\phi_\lambda(t_h)$. Therefore,

$$\begin{aligned}
\|u^T\tilde{z}_1\|_{\psi_2}^2 &\le C_2 E[(v^T\mathbf{Z_1})^2] \\
&\le C_2 E\Big[\sum_{h,h'=0}^{p-1} v_h Z_1(t_h)Z_1(t_{h'})v_{h'}\Big] \\
&\le C_2 \sum_{\lambda,\lambda'\in\Lambda_D} u_\lambda u_{\lambda'} \frac{1}{p^2} E\Big[\sum_{h,h'=0}^{p-1} \phi_\lambda(t_h)\phi_{\lambda'}(t_{h'})Z_1(t_h)Z_1(t_{h'})\Big] \\
&\le C_2 E[(u^T\tilde{z}_1)^2].
\end{aligned}$$

Now, if we consider $\tilde{\varepsilon}_1$ instead of $\tilde{z}_1$, setting $\varepsilon_{\mathbf{1}} := (\varepsilon_{1,0}, \dots, \varepsilon_{1,p-1})^T$, and using Section 5.2.3 and Lemma 5.24 of Vershynin (2012),

$$\left\| v^T\varepsilon_{\mathbf{1}} \right\|_{\psi_2}^2 \le C\sigma^2\|v\|_{\ell_2}^2 = CE[(u^T\tilde{\varepsilon}_1)^2],$$

with C an absolute constant, and

$$\|u^T\tilde{\varepsilon}_1\|_{\psi_2}^2 \le CE[(u^T\tilde{\varepsilon}_1)^2].$$

The equalities (17) *are obvious.*

Results of the previous lemma are useful for the following result.

Lemma 6 *We denote $X = (X_{\lambda\lambda'})_{\lambda,\lambda'\in\Lambda_D}$ the matrix whose entries are*

$$X_{\lambda\lambda'} = \widetilde{Y}_{1,\lambda}\widetilde{Y}_{1,\lambda'} - E[\widetilde{Y}_{1,\lambda}\widetilde{Y}_{1,\lambda'}].$$

Setting

$$M_D := \sum_{\lambda\in\Lambda_D} s_\lambda^2 + \sum_{\lambda\in\Lambda_D} \sigma_\lambda^2 = \sum_{\lambda\in\Lambda_D} \left(\frac{1}{p^2}\sum_{h,h'=0}^{p-1} K(t_h,t_{h'})\phi_\lambda(t_h)\phi_\lambda(t_{h'}) + \frac{\sigma^2}{p^2}\sum_{h=0}^{p-1}\phi_\lambda^2(t_h) \right),$$

under Assumption 2, there existe an absolute constant $\bar{C}$ such that for any $t \ge \bar{C}(C_2+1)M_D$,

$$E\left[\exp(t^{-1}\|X\|_F)\right] \le \exp(1).$$

Proof (of Lemma 6) *We have*

$$\begin{aligned}\|X\|_F &= \|\widetilde{Y}_1\widetilde{Y}_1^T - E[\widetilde{Y}_1\widetilde{Y}_1^T]\|_F \\ &\leq \|(\tilde{z}_1+\tilde{\varepsilon}_1)(\tilde{z}_1+\tilde{\varepsilon}_1)^T\|_F + \|E\big[(\tilde{z}_1+\tilde{\varepsilon}_1)(\tilde{z}_1+\tilde{\varepsilon}_1)^T\big]\|_F \\ &\leq \|\tilde{z}_1\tilde{z}_1^T\|_F + \|\tilde{\varepsilon}_1\tilde{\varepsilon}_1^T\|_F + 2\|\tilde{z}_1\tilde{\varepsilon}_1^T\|_F + \|E\big[\tilde{z}_1\tilde{z}_1^T\big]\|_F + \|E\big[\tilde{\varepsilon}_1\tilde{\varepsilon}_1^T\big]\|_F \\ &\leq \|\tilde{z}_1\|_{\ell_2}^2 + \|\tilde{\varepsilon}_1\|_{\ell_2}^2 + 2\|\tilde{z}_1\|_{\ell_2}\|\tilde{\varepsilon}_1\|_{\ell_2} + \|E\big[\tilde{z}_1\tilde{z}_1^T\big]\|_F + \|E\big[\tilde{\varepsilon}_1\tilde{\varepsilon}_1^T\big]\|_F.\end{aligned}$$

We also have

$$\begin{aligned}\|E\big[\tilde{z}_1\tilde{z}_1^T\big]\|_F^2 &= \sum_{\lambda,\lambda'\in\Lambda_D}\big(E[\tilde{z}_{1,\lambda}\tilde{z}_{1,\lambda'}]\big)^2 \\ &\leq \sum_{\lambda\in\Lambda_D}\sum_{\lambda'\in\Lambda_D} E[\tilde{z}_{1,\lambda}^2]E[\tilde{z}_{1,\lambda'}^2] \\ &\leq \Big(\sum_{\lambda\in\Lambda_D} s_\lambda^2\Big)^2.\end{aligned}$$

Therefore

$$\|E\big[\tilde{z}_1\tilde{z}_1^T\big]\|_F \leq \sum_{\lambda\in\Lambda_D} s_\lambda^2$$

and similarly,

$$\|E\big[\tilde{\varepsilon}_1\tilde{\varepsilon}_1^T\big]\|_F \leq \sum_{\lambda\in\Lambda_D} \sigma_\lambda^2.$$

We finally obtain

$$\|X\|_F \leq 2\|\tilde{z}_1\|_{\ell_2}^2 + 2\|\tilde{\varepsilon}_1\|_{\ell_2}^2 + M_D$$

and we have

$$E\left[\exp(t^{-1}\|X\|_F)\right] \leq E\left[\exp(2t^{-1}\|\tilde{z}_1\|_{\ell_2}^2)\right] \times E\left[\exp(2t^{-1}\|\tilde{\varepsilon}_1\|_{\ell_2}^2)\right] \times \exp(t^{-1}M_D).$$

Then, using Lemma 5 and Proposition A.1. of Bunea and Xiao (2015), we obtain for C_* and c_* two absolute positive constants, if $t > c_*(4C_2+1)\sum_{\lambda\in\Lambda_D} s_\lambda^2$,

$$\begin{aligned}E\left[\exp(2t^{-1}\|\tilde{z}_1\|_{\ell_2}^2)\right] &\leq E\left[\exp\left(2t^{-1}\Big(\|\tilde{z}_1\|_{\ell_2}^2 - \sum_{\lambda\in\Lambda_D} s_\lambda^2\Big)\right)\right] \times \exp\left(2t^{-1}\sum_{\lambda\in\Lambda_D} s_\lambda^2\right) \\ &\leq \exp\left(C_*\left(\frac{(4C_2+1)\sum_{\lambda\in\Lambda_D} s_\lambda^2}{t}\right)^2 + 2t^{-1}\sum_{\lambda\in\Lambda_D} s_\lambda^2\right).\end{aligned}$$

Similarly, for t larger than $\sum_{\lambda\in\Lambda_D}\sigma_\lambda^2$ up to a multiplicative absolute constant,

$$E\left[\exp(2t^{-1}\|\tilde{\varepsilon}_1\|_{\ell_2}^2)\right] \leq \exp\left(C_{**}\left(\frac{\sum_{\lambda\in\Lambda_D}\sigma_\lambda^2}{t}\right)^2 + 2t^{-1}\sum_{\lambda\in\Lambda_D}\sigma_\lambda^2\right),$$

where C_{**} is an absolute constant. This ends the proof of the lemma.

The following proposition controls the term $\|\widehat{G}_\phi - G_\phi\|_F$ as required to complete the proof of Theorem 3 .

Proposition 5 *We assume that Assumption 2 is satisfied. For $\gamma > 0$, with probability larger than $1 - 2\exp(-1/64\min(\gamma^2, 16\gamma\sqrt{n}))$,*

$$\|\widehat{G}_\phi - G_\phi\|_F \leq \frac{\bar{C}(e^{1/2} + \gamma)(C_2 + 1)}{\sqrt{n}} \sum_{\lambda \in \Lambda_D} \Big[\sigma_\lambda^2 + s_\lambda^2\Big],$$

where $\bar{C}$ is an absolute constant.

Proof (of Proposition 5) *We apply Theorem 4.1 of Juditsky and Nemiroski (2008) with $\alpha = 1$, since $(\mathbb{R}^{|\Lambda_D|^2}, \|\cdot\|_F)$ is 1-smooth (see Definition 2.1 of Juditsky and Nemiroski (2008)). Since Lemma 6 gives for $t \geq \bar{C}(C_2 + 1)\sum_{\lambda\in\Lambda_D}\Big[\sigma_\lambda^2 + s_\lambda^2\Big]$,*

$$E\left[\exp(t^{-1}\|X\|_F)\right] \leq \exp(1),$$

for $\gamma > 0$, with probability larger than $1 - 2\exp(-1/64\min(\gamma^2, 16\gamma\sqrt{n}))$,

$$\|\widehat{G}_\phi - G_\phi\|_F \leq \frac{\bar{C}(e^{1/2} + \gamma)(C_2 + 1)}{\sqrt{n}} \sum_{\lambda \in \Lambda_D} \Big[\sigma_\lambda^2 + s_\lambda^2\Big].$$

Proposition 5 is proved.

Plugging the upper bound of Proposition 5 in (12) provides the stated result of Theorem 3 .

C.4 Proof of Proposition 1

We control each deterministic term of the bound obtained in Theorems 2 and 3. Using (14), we first have for any $f \in \mathbb{L}_2$,

$$\begin{aligned}
\|\Pi_D\Gamma\Pi_D(f) - \Gamma(f)\|^2 &= \int_0^1 \Big(\Pi_D\Gamma\Pi_D(f)(t) - \Gamma(f)(t)\Big)^2 dt \\
&= \int_0^1 \Big(\int_0^1 \Pi_{S_D^2}K(s,t)f(s)ds - \int_0^1 K(s,t)f(s)ds\Big)^2 dt \\
&= \int_0^1 \Big(\int_0^1 (\Pi_{S_D^2}K(s,t) - K(s,t))f(s)ds\Big)^2 dt \\
&\leq \int_0^1 \left[\int_0^1 (\Pi_{S_D^2}K(s,t) - K(s,t))^2 ds \int_0^1 f^2(s)ds\right] dt \\
&\leq \|\Pi_{S_D^2}K - K\|^2\|f\|^2
\end{aligned}$$

and then

$$|\!|\!|\Pi_D\Gamma\Pi_D - \Gamma|\!|\!|^2 \leq \|\Pi_{S_D^2}K - K\|^2.$$

Now, we take $(s,t) \in [0,1]^2$. Then there exists a unique couple $(\lambda, \lambda') \in \Lambda_D^2$ such that $s \in I_\lambda$ and $t \in I_{\lambda'}$. Therefore, $\phi_{\lambda''}(s) = 0$ for $\lambda'' \neq \lambda$ and $\phi_{\lambda'''}(t) = 0$ for $\lambda''' \neq \lambda'$ and then,

$$\begin{aligned}
\Pi_{S_D^2}K(s,t) - K(s,t) &= \sum_{\lambda'',\lambda'''\in\Lambda_D} \int_0^1\int_0^1 K(s',t')\phi_{\lambda''}(s')\phi_{\lambda'''}(t')ds'dt'\phi_{\lambda''}(s)\phi_{\lambda'''}(t) - K(s,t) \\
&= \int_0^1\int_0^1 K(s',t')\phi_\lambda(s')\phi_{\lambda'}(t')ds'dt'\phi_\lambda(s)\phi_{\lambda'}(t) - K(s,t) \\
&= D^2\int_{I_\lambda}\int_{I_{\lambda'}} (K(s',t') - K(s,t))ds'dt'.
\end{aligned}$$

Then, Eq. (4) gives

$$\begin{aligned}\Big|\Pi_{S_D^2}K(s,t)-K(s,t)\Big| &\le D^2\sqrt{L\|K\|_\infty}\int_{I_\lambda}\int_{I_{\lambda'}}\Big[|s'-s|^\alpha+|t-t'|^\alpha\Big]ds'dt'\\ &\le \frac{4\sqrt{L\|K\|_\infty}}{\alpha+1}D^{-\alpha},\end{aligned}$$

meaning that

$$|||\Pi_D\Gamma\Pi_D-\Gamma|||^2\le\frac{16L\|K\|_\infty}{(\alpha+1)^2}D^{-2\alpha}.$$

For studying the terms $A_p^{(K)}(\phi,D)$ and $A_p^{(\sigma)}(\phi,D)$, we set for any $h=0,\ldots,p-1$, $b_h=h/p$. Observe that $t_h=h/(p-1)\in[b_h,b_{h+1}]$. We also set for any $\lambda=0,\ldots,D-1$,

$$J_\lambda=\{h=0,\ldots,p-1:\ Leb([b_h,b_{h+1}]\cap I_\lambda)\neq 0\}.$$

Remember that $m:=p/D$ is an integer, so that, $J_\lambda=\{m\lambda,\ldots,m\lambda+m-1\}$ and

$$I_\lambda=\Big[\frac{m\lambda}{p},\frac{m\lambda+m}{p}\Big]=\bigcup_{h\in J_\lambda}[b_h,b_{h+1}].$$

Then, since $\phi_\lambda(x)=\sqrt{D}1_{I_\lambda}(x)$ and $\mathrm{card}(J_\lambda)=m=p/D$, for any $\lambda,\lambda'=0,\ldots,D-1$,

$$\begin{aligned}G^{(K)}_{\lambda,\lambda'} &:= \frac{1}{p^2}\sum_{h,h'=0}^{p-1}K(t_h,t_{h'})\phi_\lambda(t_h)\phi_{\lambda'}(t_{h'})-\int_0^1\int_0^1K(s,t)\phi_\lambda(s)\phi_{\lambda'}(t)dsdt\\ &= \sum_{h,h'=0}^{p-1}\int_{b_h}^{b_{h+1}}\int_{b_{h'}}^{b_{h'+1}}\Big[K(t_h,t_{h'})-K(s,t)\Big]\phi_\lambda(s)\phi_{\lambda'}(t)dsdt\\ &\quad+\sum_{h,h'=0}^{p-1}\int_{b_h}^{b_{h+1}}\int_{b_{h'}}^{b_{h'+1}}K(t_h,t_{h'})\Big[\phi_\lambda(t_h)\phi_{\lambda'}(t_{h'})-\phi_\lambda(s)\phi_{\lambda'}(t)\Big]dsdt\\ &= D\sum_{h\in J_\lambda}\sum_{h'\in J_{\lambda'}}\int_{b_h}^{b_{h+1}}\int_{b_{h'}}^{b_{h'+1}}\Big[K(t_h,t_{h'})-K(s,t)\Big]dsdt.\end{aligned}$$

Therefore,

$$\begin{aligned}|G^{(K)}_{\lambda,\lambda'}| &\le D\sum_{h\in J_\lambda}\sum_{h'\in J_{\lambda'}}\int_{b_h}^{b_{h+1}}\int_{b_{h'}}^{b_{h'+1}}\sqrt{\|K\|_\infty L}\Big(|s-t_h|^\alpha+|t-t_{h'}|^\alpha\Big)dsdt\\ &\le 2\sqrt{\|K\|_\infty L}\times Dp^{-1}\mathrm{card}(J_{\lambda'})\sum_{h\in J_\lambda}\int_{b_h}^{b_{h+1}}|s-t_h|^\alpha ds\\ &\le 2\sqrt{\|K\|_\infty L}\times Dp^{-1}\mathrm{card}(J_{\lambda'})\mathrm{card}(J_\lambda)\times\frac{2}{\alpha+1}p^{-\alpha-1}\\ &\le \frac{4\sqrt{\|K\|_\infty L}}{\alpha+1}D^{-1}p^{-\alpha}.\end{aligned}$$

Finally,

$$A_p^{(K)}(\phi,D)=\|G^{(K)}\|_F^2=\sum_{\lambda,\lambda'\in\Lambda_D}\Big(G^{(K)}_{\lambda,\lambda'}\Big)^2\le\frac{16\|K\|_\infty L}{(\alpha+1)^2}p^{-2\alpha}.$$

Similarly, for any $\lambda, \lambda' = 0, \ldots, D-1$, observing that for $\lambda \neq \lambda'$, $J_\lambda \cap J_{\lambda'} = \emptyset$,

$$
\begin{aligned}
G^{(\sigma)}_{\lambda,\lambda'} &:= \frac{\sigma^2}{p}\left(\frac{1}{p}\sum_{h=0}^{p-1}\phi_\lambda(t_h)\phi_{\lambda'}(t_h) - \langle\phi_\lambda, \phi_{\lambda'}\rangle\right) \\
&= \frac{\sigma^2}{p}\left(\frac{1}{p}\sum_{h \in J_\lambda \cap J_{\lambda'}} D - \mathbf{1}_{\{\lambda=\lambda'\}}\right) = 0
\end{aligned}
$$

and

$$
A^{(\sigma)}_p(\phi, D) = \|G^{(\sigma)}\|_F^2 = \sum_{\lambda,\lambda' \in \Lambda_D}\left(G^{(\sigma)}_{\lambda,\lambda'}\right)^2 = 0.
$$

Finally, for any $\lambda = 0, \ldots, D-1$,

$$
\begin{aligned}
\sigma_\lambda^2 + s_\lambda^2 &= \frac{\sigma^2}{p^2}\sum_{h=0}^{p-1}\phi_\lambda^2(t_h) + \frac{1}{p^2}\sum_{h,h'=0}^{p-1}K(t_h, t_{h'})\phi_\lambda(t_h)\phi_\lambda(t_{h'}) \\
&\leq \frac{D\sigma^2}{p^2}\mathrm{card}(J_\lambda) + \frac{\|K\|_\infty D}{p^2}(\mathrm{card}(J_\lambda))^2 \\
&\leq \frac{\sigma^2}{p} + \frac{\|K\|_\infty}{D}
\end{aligned}
$$

and

$$
\sum_{\lambda \in \Lambda_D}\left[\sigma_\lambda^2 + s_\lambda^2\right] \leq \|K\|_\infty + \frac{\sigma^2 D}{p}.
$$

This ends the proof of Proposition 1.

C.5 Proof of Lemma 1

To prove Lemma 1, we just need to prove that for all $v = (v_h)_{h=0,\ldots,p} \in \mathbb{R}^p$,

$$
E[\exp(tv^T\mathbf{Z})] \leq \exp\left(ct^2M^2E[(v^T\mathbf{Z})^2]\right), \quad \forall\, t \in \mathbb{R}, \tag{18}
$$

with c an absolute constant (see Proposition 2.5.2 of Vershynin [2018]). For this purpose, we denote

$$
u_{v,d} := \sum_{h=0}^{p-1} v_h \eta_d^*(t_h) \in \mathbb{R}.
$$

Then, using (3), we have

$$
E[(v^T\mathbf{Z})^2] = \sum_{d \in \mathbb{N}^*}\mu_d^* u_{v,d}^2
$$

and, still by using Proposition 2.5.2 of Vershynin [2018], $\forall\, t \in \mathbb{R}$,

$$
\begin{aligned}
E[\exp(tv^T\mathbf{Z})] &= \prod_{d \in \mathbb{N}^*} E\Big[\exp\left(t\zeta_d^*\mu_d^{*1/2}u_{v,d}\right)\Big] \\
&\leq \prod_{d \in \mathbb{N}^*} E\Big[\exp\left(ct^2\mu_d^* u_{v,d}^2\|\zeta_d^*\|_{\psi_2}^2\right)\Big],
\end{aligned}
$$

where c is an absolute constant. By using (6), we obtain

$$
\begin{aligned}
E[\exp(tv^T\mathbf{Z})] &\le \prod_{d\in\mathbb{N}^*} E\Big[\exp\big(ct^2\mu_d^* u_{v,d}^2 M^2\big)\Big] \\
&\le \exp\big(ct^2 M^2 E[(v^T\mathbf{Z})^2]\big)
\end{aligned}
$$

and (18) is satisfied.