The correction term in a Small–Ball Probability factorization for random curves

Jean-Baptiste Aubin[1], Enea G. Bongiorno[2,3], Aldo Goia[2,3]
jean-baptiste.aubin@insa-lyon.fr, enea.bongiorno@uniupo.it, aldo.goia@uniupo.it
[1]INSA-Lyon, France
[2]DiSEI and AI@UPO, Università del Piemonte Orientale, Italy
[3]Gnampa, INdAM

February 17, 2021

Abstract

In this work we propose an analysis of the correction term appearing in a Small-Ball Probability factorization for random elements taking values in a separable Hilbert space. Its local nature, its meaning and behavior are discussed also through the derivation of some bounds. Nonparametric kernel–type estimators of the considered statistics are introduced and some asymptotic properties are provided. Finally, in the context of reconstructing a sample of curves by truncated Karhunen–Loève expansion, a local approach to select the dimensionality is illustrated through numerical and real data examples.

Keywords: Hilbert random elements, Karhunen–Loève expansion, nonparametric estimation, asymptotic theorems

1 Introduction

In the last two decades, functional statistic has consolidated, reaching increasingly deeper levels of maturity along different directions. As evidence of this, there are several monographs Ramsay and Silverman [2005], Ferraty and Vieu [2006], Horváth and Kokoszka [2012], Hsing and Eubank [2015], Kokoszka and Reimherr [2017], many specialized conference proceedings Bongiorno et al. [2014], Aneiros et al. [2017, 2020] and special issues Goia and Vieu [2016], Ferraty et al. [2019], Aneiros et al. [2019b,a] that especially in recent years have enriched the literature on this topic. Nevertheless, in the frenzied rush to explore the enormous potential of functional statistics in the applications, different aspects has been left underdeveloped. Among these, an overlooked tool is the small–ball probability (SmBP) that, for a random element X of a metric space H, is $\varphi(x,h) = \mathbb{P}(X \in B(x,h))$ as $h \to 0$ where $B(x,h)$ is the ball centered at $x \in H$ with radius h. The main difficulty, which has limited the exploitation of SmBP in applications, consists in transforming a purely theoretical concept into a useful operative instrument. In fact, until recently, the use of the SmBP has been limited to the evaluation of the convergence rates of the estimators emerging from functional regression problems (see Ferraty and Vieu [2006] and references therein) and, only in the last decade, few works have given to the SmBP an active role in applications as well. The efforts

made so far are closely related to the possibility of factorizing the SmBP, for example assuming that, for any $x \in H$ and as $h \to 0$,

$$\varphi(x,h) \sim \psi(x)\phi(h), \tag{1}$$

with the identifiability constraint $\mathbb{E}[\psi(X)] = 1$. The convenience of the above factorization resides in the fact that it isolates the way the SmBP depends upon x and h (the space and volumetric variables respectively) and allows to interpret ψ and ϕ as a surrogate of the probability density of X (that can not be straightforwardly defined in a infinite dimensional space) and a volumetric term respectively. This interpretation of the factorization (1) strongly resembles the multivariate case and has led to define a concept of mode (see Gasser et al. [1998]), to introduce a surrogate-density (see Ferraty et al. [2012]) and to evaluate the intrinsic complexity or dimension of the underlying process X (see Bongiorno et al. [2017, 2019, 2020]).

Although the factorization (1) appears very attractive from both a theoretical and a practical point of view, it is known that it holds true only for some processes (see Li and Shao [2001], Lifshits [2012] and references therein) and there is no global theory that assesses its validity for a general process X. To compensate for this lack, some alternatives have been proposed (see Delaigle and Hall [2010], Bongiorno and Goia [2017]).

In this paper, a three terms factorization proposed in Bongiorno and Goia [2017] is considered: given $x \in H$, a separable Hilbert space, as $h \to 0$,

$$\varphi(x,h) \sim f_d(\Pi_d x) V_d(h) C_d(x,h), \tag{2}$$

where $f_d(\Pi_d x)$ is the pdf of the first d principal components, V_d is the volume of the d–dimensional ball of radius h and $C_d(x,h) \in (0,1]$, depending on x, h and d denotes a suitable correction factor. While f_d and V_d maintain the same meaning of ψ and ϕ respectively in (1), making possible the use of f_d in density based techniques (e.g. Bongiorno and Goia [2016], Jacques and Preda [2014] and references therein), the factor C_d has never been analyzed or used for statistical purposes.

The aim of this work is to fill this gap deepening the knowledge of $C_d(x,h)$ to understand what role it plays in factorizing the SmBP, how it can modify the assessments made on the process and/or on its SmBP and, finally, how it can be exploited in the applications in identifying a local dimension to parsimoniously represent x. At the first glance, looking at the factorization (2), C_d provides a compensation for the use of the finite dimensional factorization $f_d V_d$; in particular, the closer the correction factor and zero are, the worse the factorization $f_d V_d$ is in approximating the SmBP. This implies that, instead of using the sole f_d in the density based statistical applications, one should employ its adjusted version $\psi_d = f_d C_d$. Going deeper, this work studies the behaviour of $C_d(x,h)$ varying x by means of some bounds. For some processes, it is proven that $C_d(x,h)$ reaches its maximum over all the points in the space H_d generated by the first d eigenfunctions of the covariance operator of X. These latter results allow to interpret $C_d(x,h)$ as a local measure of the quality of the representation of x as an element of H_d. As a consequence, in this paper the correction factor is used to customize the dimension used in the truncated Karhunen–Loève representation of each curve of a given sample defining a novel local reduction dimensionality principle that can be seen as an alternative to the well known fraction of explained variance which leads to a unique dimension for all the curves. Further, the average behaviour of C_d is also studied, leading in a natural way to charaterize finite dimensional processes.

To make C_d usable in practice, a kernel–type nonparametric estimate is provided exploiting the fact that, by its own definition, C_d is a conditioned mean. For this estimator, it has been proved that the rate of convergence in quadratic mean is the optimal one. As a by–product,

similar results are derived for the surrogate density f_d and its adjusted version ψ_d. By using the proposed nonparametric estimate of the correction factor, a local reduction dimensionality algorithm is proposed, implemented and its performances are analysed by means of a Monte Carlo simulation study. Finally, the new technique is applied to two real datasets.

The outline of the paper goes as follows. Section 2 formally introduces factorization (2). Section 3 discusses various aspects of the correction factor proposing some interpretations, properties and potential uses. Section 4 studies nonparametric estimators for f_d, C_d and ψ_d. Section 5 introduces the local reduction dimensionality algorithm which is tested by means of simulations. Applications to two real cases are illustrated in Section 6. Finally, proof of theoretical results are collected in Appendix.

2 Preliminaries: a factorization result

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $H = \mathcal{L}^2_{[0,1]}$ be the Hilbert space of square integrable real functions on $[0,1]$, endowed with the standard inner product $\langle g, h\rangle = \int_0^1 g(t)\, h(t)\, dt$ and the induced norm $\|g\|^2 = \langle g, g\rangle$. Consider a measurable map X defined on $(\Omega, \mathcal{F})$ taking values in $(H, \mathcal{B})$, where $\mathcal{B}$ denotes the Borel sigma–algebra induced by $\|\cdot\|$. Assume that the mean function and covariance operator of X, defined by $\mu_X = \{\mathbb{E}[X(t)], t \in [0,1]\}$ and $\Sigma[\cdot] = \mathbb{E}[\langle X - \mu_X, \cdot\rangle (X - \mu_X)]$ respectively, are well posed. Consider the Karhunen–Loève expansion associated to X (see e.g. Bosq [2000]): denoting by $\{\lambda_j, \xi_j\}_{j=1}^{\infty}$ the decreasing to zero sequence of positive eigenvalues and the associated orthonormal eigenfunctions of the covariance operator Σ, the random curve X admits the representation $X(t) = \mu_X(t) + \sum_{j\geq 1} \theta_j \xi_j(t)$, $0 \leq t \leq 1$, where $\theta_j = \langle X - \mu_X, \xi_j\rangle$ are the so–called principal components (PCs in the sequel) of X satisfying $\mathbb{E}[\theta_j] = 0$, $Var(\theta_j) = \lambda_j$ and $\mathbb{E}[\theta_j \theta_{j'}] = 0$, $j \neq j'$. From now on and without loss of generality, assume that $\mu_X = 0$.
Consider the following assumptions:

A.1 $\varphi(x, h) > 0$, for any $x \in H$, and $h > 0$.

A.2 The center $x \in H$ of the ball $B(x,h) = \{\|X - x\| \leq h\}$ satisfies $x_j^2 \leq c_1 \lambda_j$ for any $j \geq 1$, where $x_j = \langle x, \xi_j\rangle$ for some strictly positive constant c_1; that means that projections x_j of x decay to zero at most with the same rate of $\sqrt{\lambda_j}$ and in some sense x is sufficiently close to the process along all directions ξ_j.

A.3 Denote by Π_d the projector onto H_d the d–dimensional space spanned by $\{\xi_j\}_{j=1}^d$, and by $\Pi_d^{\perp}$ its orthogonal projector.

(a) The first d PCs, $\Pi_d X = (\theta_1, \ldots, \theta_d)'$, admit a joint strictly positive probability density, $\vartheta \in \mathbb{R}^d \mapsto f_d(\vartheta)$. Moreover, f_d is twice differentiable at $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_d)' \in \mathbb{R}^d$, and there exists a strictly positive constant c_2 (not depending on d) for which

$$\left| \frac{\partial^2 f_d}{\partial \vartheta_i \partial \vartheta_j}(\boldsymbol{\vartheta}) \right| \leq \frac{c_2}{\sqrt{\lambda_i \lambda_j}} f_d(\Pi_d x)$$

for any $d \in \mathbb{N}$, $i, j \leq d$ and ϑ such that $\sum_{j\leq d}(\vartheta_j - x_j)^2 \leq \rho^2$ for some $\rho \geq h$.

(b) The r.v. $\|\Pi_d^{\perp}(X - x)\|^2 / h^2$ admits a strictly positive density whose support includes $(0,1)$.

It is worth noting that the last assumption includes, for instance, the case of Gaussian Hilbert–valued processes.

The following result, proved in [Bongiorno and Goia, 2017, Theorem 1], provides a three terms factorization of the SmBP introducing the correction term that plays a main role along all the paper.

Proposition 1 *If A.1–A.3 hold, then, for a finite strictly positive integer d and a given point $x \in \mathcal{L}^2_{[0,1]}$,*

$$\varphi(x,h) \sim f_d(\Pi_d x) V_d(h) C_d(x,h) \qquad \text{for } h \to 0 \tag{2}$$

where $V_d(h) = h^d \pi^{d/2} / \Gamma(d/2+1)$ is the volume of the d–dimensional ball with radius h,

$$C_d(x,h) = \mathbb{E}\left[\left(\left(1 - \frac{\|\Pi_d^{\perp}(X-x)\|^2}{h^2} \right) \mathbb{I}_{\left\{\|\Pi_d^{\perp}(X-x)\|^2 \le h^2\right\}} \right)^{d/2} \middle| \Pi_d x \right] \in (0,1] \tag{3}$$

and $\Pi_d^{\perp}$ denotes the orthogonal projector of Π_d.

In other words, for a fixed d and as $h \to 0$, the SmBP $\varphi(x,h)$ behaves as $f_d(\Pi_d x) V_d(h) C_d(x,h)$, the usual first order approximation of the SmBP in a d–dimensional space $f_d(\Pi_d x) V_d(h)$ up to the scale factor $C_d(x,h)$. Changing d affects all the terms in the factorization but asymptotic in (2) still remains true. A sketch of the proof of Proposition 1 is proposed in Appendix, more details can be found in Bongiorno and Goia [2017] and Bongiorno and Goia [2018].

3 The correction factor C_d

This section focuses on the scale factor $C_d(x,h)$, defined by (3), showing how it provides a local index of dimensionality and a characterization for finite dimensional processes.

In order to understand the role of $C_d(x,h)$ in the infinite dimensional setting, it is useful to preliminarily consider what happens in the finite dimensional case. In fact, if $H = \mathbb{R}^D$, $D > 1$, and the probability law of X is dominated by the Lebesgue measure on $\mathbb{R}^D$, then, as $0 < d < D$,

$$\begin{aligned} \varphi(x,h) &\sim f_D(x) V_D(h) \\ &= f_d(\Pi_d x) V_d(h) \underbrace{\frac{V_D(h)}{V_d(h)} f_{D-d}(\Pi_d^{\perp} x | \Pi_d x)}_{C_d(x,h)} \end{aligned}$$

where $f_D(x)$ and $f_{D-d}(\Pi_d^{\perp} x | \Pi_d x)$ denote the joint distribution of X and the conditional one of $\Pi_d^{\perp} X$ given $\Pi_d X = \Pi_d x$ respectively. In other words, $C_d(x,h)$ combines the information about the conditional probability law of the last components given the first ones with a suitable volumetric ratio and compensates for the information lost when one tries to approximate a D–dimensional distribution with a d–dimensional one. Clearly if the probability law of X can be represented in a lower dimension (say $d < D$), then correction factor can be dropped.

Moving back to the infinite dimensional setting, since the densities cannot be defined in general, the conditional mean $C_d(x,h)$ surrogates the conditional law of $\Pi_d^{\perp} X$ given $\Pi_d X = \Pi_d x$ multiplied by a volumetric part, acting as a correction factor which shrinks the overestimate of the SmBP at x provided by $f_d(\Pi_d x) V_d(h)$. In fact, $f_d(\Pi_d x)$ reflects the probability behavior of X only on

the subspace H_d neglecting the probability effects of the process on the orthogonal subspace $H_d^{\perp}$ which are then captured by $C_d(x,h)$. To fix the idea, consider $x \in H$ for which $x_j \neq 0$ for some $j > d$, that, in the following, is named as *d–high frequency* curve. The larger $|x_j| / \sqrt{\lambda_j}$ are for some $j > d$, the smaller is the probability of $\left\{\|\Pi_d^{\perp}(X-x)\|^2 \leq h^2\right\}$ and the closer to zero is $C_d(x,h)$. This pushes SmBP to zero even if the corresponding surrogate density $f_d(\Pi_d x)$ is large. As a consequence, the use of f_d as the pseudo-density in the statistical applications (such as clustering density based approaches, discriminant methods, or mode computation) could lead to biased results and interpretations because the pseudo–density f_d could associate too much importance to the *d–high frequency* curves. So, whenever one needs the evaluation of the pseudo-density, a good practice in applications would be to use the adjusted pseudo-density $\psi_d(x,h) = f_d(\Pi_d x) C_d(x,h)$ instead of $f_d(\Pi_d x)$ even if $C_d(x,h)$ implicitly includes a volumetric part of the SmBP. In fact, since this volumetric part is common to all the $x \in H$ its effects on ψ_d may be assumed to be homogeneous over all x. In this perspective, the correction factor may be seen as a local measure at x of the goodness of the SmBP approximation via $f_d(\Pi_d x) V_d(h)$.

3.1 Some bounds

At this point, it is clear the importance of exploring what is the range of C_d as a function of x, h and d and, to do this, some bounds are provided in the following proposition whose proof can be found in the Appendix A.2.

Proposition 2 *Given A.1–A.3 and denoting* $A_{d,h} = \left\{\|\Pi_d^{\perp}(X-x)\|^2 \leq h^2\right\}$*, then for any* x *and* $h > 0$*,*

$$0 < C_d(x,h) \leq \left(1 - \frac{1}{h^2}\mathbb{E}\left[\|\Pi_d^{\perp}(X-x)\|^2 \middle| \mathbb{I}_{A_{d,h}} = 1, \Pi_d x\right]\right)^{1/2} \qquad d = 1 \tag{4}$$

$$1 - \frac{d+2}{2h^2}\mathbb{E}\left[\|\Pi_d^{\perp}(X-x)\|^2 \middle| \mathbb{I}_{A_{d,h}} = 1, \Pi_d x\right] \leq C_d(x,h) \leq 1 - \frac{1}{h^2}\mathbb{E}\left[\|\Pi_d^{\perp}(X-x)\|^2 \middle| \mathbb{I}_{A_{d,h}} = 1, \Pi_d x\right] \qquad d > 1. \tag{5}$$

Such a bounds confirm what said above on the values of $C_d(x,h)$ when x is a *d–high frequency* curve. On the other hand, given that

$$\mathbb{E}\left[\|\Pi_d^{\perp}(X-x)\|^2 \middle| \mathbb{I}_{A_{d,h}} = 1, \Pi_d x\right] \leq \|\Pi_d^{\perp} x\|^2 + \mathbb{E}\left[\|\Pi_d^{\perp} X\|^2 \middle| \mathbb{I}_{A_{d,h}} = 1, \Pi_d x\right],$$

whenever $x \in H_d$ (a *d–low frequency* curve), then $x_j = 0$ for any $j > d$ and the inferior bound (5) increases leading greater values than if x is a *d–high frequency* curve. Anyway, thanks to A.3.b, the expectations in Proposition 2 are strictly positive and, consequently, all the upper bounds are strictly smaller than 1.

In some cases, it is possible to prove that $C_d(x,h)$ admits a maximum as shown in the next proposition whose proof is in Appendix A.2.

Proposition 3 *Given A.1–A.3. Fix* $h > 0$ *and a strictly positive integer* d*. Assume that* $((1 - \|\Pi_d^{\perp}(X-x)\|^2/h^2)\mathbb{I}_{\left\{\|\Pi_d^{\perp}(X-x)\|^2 \leq h^2\right\}})^{d/2}$ *is uncorrelated with* $\{\Pi_d X = \Pi_d x\}$*. Then,* $C_d(x,h)$ *admits a maximum* $M_d(h)$ *over* H *and it is achieved for any* $x \in H_d$*.*

Roughly speaking, the maximum of $C_d(x,h)$ is reached at any point $x \in H_d$ such that $x_j = 0$ for any $j > d$. Note that the uncorrelation hypothesis holds true for all the processes with independent PCs as, for example, the Gaussian processes.

As a consequence the results illustrated in this section, $C_d(x,h)$ can be of help in identifying the *d–low frequency* curves leading to a parsimonious representation of them. In fact, the closer $C_d(x,h)$ and $M_d(h)$ are, the more accurate the representation of x over the subspace H_d is: additional dimensions do not substantially improve the quality of the approximation. In this view, the correction factor provides a local index of dimensionality that could replace the use of the Fraction of Explained Variance (FEV) which instead furnishes a global measure of dimensionality. This idea is the starting point for a new local reduction dimensionality approach that is presented in Section 5.

3.2 A finite dimensionality characterization result

Results above are stated in the case of a infinite dimensional process for which the factorization (2) holds true. For a finite dimensional process, there exists a positive integer d_0 such that for every $d > d_0$ assumption A.3.b fall since $\left\|\Pi_d^{\perp}(X-x)\right\|^2/h^2 = 0$ a.s.. In this situation, the factorization still holds true for each $d < d_0$, while for $d = d_0$ the factorization simplifies to $f_{d_0}(\Pi_{d_0}x)\,V_{d_0}(h)$. This means that $C_d(x,h)$ (whose definition is well posed regardless of the factorization) allows to characterize the finite dimensionality of a process as shown in the next result.

Proposition 4 *Let X' be an independent copy of X, d a strictly positive integer and $h > 0$. Then the following statements are equivalent:*

i) $\mathbb{E}[C_d(X',h)] = 1$*;*

ii) $C_d(X',h) = 1$ *a.s.;*

iii) $\lambda_{d+1} = 0$*;*

iv) the process admits the following finite dimensional representation $X = \sum_{j=1}^{d}\theta_j\xi_j$ *a.s..*

The smallest d for which one of the above statements holds true is d_0, that is the dimension of the process X. If such a minimum does not exists then X is a infinite dimensional process.

Suppose that one wants to test the null hypothesis that X has dimension d_0. Then one could test that $\mathbb{E}[C_{d_0}(X',h)]$ equals one against less than one, but this is equivalent to test that $\lambda_{d_0+1} = 0$ for which some statistical procedures have been already explored (see Bathia et al. [2010] and Hall and Vial [2006]).

4 Nonparametric Estimate

Consider a sample $X_1,\ldots,X_n$ drawn from X, nonparametric estimates of $f_d(x)$ and $C_d(x,h)$ are provided: combining them, one gets also a nonparametric estimate of the adjusted pseudo–density $\psi_d(x,h)$.

For what concerns the first, consider the Parzen–Rosenblatt estimator

$$f_{d,n}(x) = \frac{1}{nb_1^d}\sum_{i=1}^{n} K_1\left(\frac{\|\Pi_d(X_i - x)\|}{b_1}\right) \tag{6}$$

where b_1 is a bandwidth (in general depending on n) and K is a suitable kernel.

For the second term, consider X' an independent copy of X and define the real r.v.

$$Y = \left(\left(1 - \frac{\|\Pi_d^{\perp}(X - X')\|^2}{h^2} \right) \mathbb{I}_{\left\{ \|\Pi_d^{\perp}(X-X')\|^2 \leq h^2 \right\}} \right)^{d/2}, \tag{7}$$

then $C_d(x,h)$ can be seen as the regression function of Y, given the d–dimensional random vector $Z = \Pi_d X'$ at $X' = x$. This interpretation suggests a way to estimate nonparametrically C_d by means of the classical Nadaraya–Watson approach:

$$C_{d,n}(x,h) = \sum_{i=1}^{n} w_i(x, b_2) \left(\left(1 - \frac{\|\Pi_d^{\perp}(X_i - x)\|^2}{h^2} \right) \mathbb{I}_{\left\{ \|\Pi_d^{\perp}(X_i-x)\|^2 \leq h^2 \right\}} \right)^{d/2} \tag{8}$$

where

$$w_i(x, b_2) = \frac{K_2(\|\Pi_d(X_i - x)\|/b_2)}{\sum_j K_2(\|\Pi_d(X_j - x)\|/b_2)} \tag{9}$$

with b_2 being a bandwidth (in general depending on n) and K_2 a suitable kernel.

Combining (6), (8) and (9) one also gets the following

$$\psi_{d,n}(x,h) = f_{d,n}(x)\, C_{d,n}(x,h) \tag{10}$$

that, whenever $K_1 = K_2 = K$ and $b_1 = b_2 = b$, simplifies to

$$\psi_{d,n}(x,h) = \frac{1}{nb^d} \sum_{i=1}^{n} \left(\left(1 - \frac{\|\Pi_d^{\perp}(X_i - x)\|^2}{h^2} \right) \mathbb{I}_{\left\{ \|\Pi_d^{\perp}(X_i-x)\|^2 \leq h^2 \right\}} \right)^{d/2} K\left(\frac{\|\Pi_d(X_i - x)\|}{b} \right). \tag{11}$$

Note that the latter is nothing but the $f_{d,n}$ whose series arguments are adjusted by a factor depending on $\left\|\Pi_d^{\perp}(X_i - x)\right\|$. This means that $\psi_{d,n}$ depends on $\|\Pi_d^{\perp}(X_i - x)\|$ and $\|\Pi_d(X_i - x)\|$ (whose sum gives $\|X_i - x\|$), differently from $f_{d,n}$ that depends only on $\|\Pi_d(X_i - x)\|$.

Now, to derive some consistency results for the introduced estimators, consider the following additional assumptions:

A.4 $f_d(x)$ is a bounded function, p times differentiable at $x \in \mathbb{R}^d$, with $p \geq 2$;

A.5 the sequences $\{b_j = b_{j,n}\}$, $j = 1, 2$, satisfy: $b_j \to 0$ and $nb_j^d / \log n \to \infty$ as $n \to \infty$;

A.6 the kernels K_1, K_2 are Lipschitz, bounded, integrable density functions with compact support $[0, 1]$;

A.7 as $t \to \infty$, $\mathbb{P}(\|X - x\| > t) = o(\exp\{-rt\})$ for any strictly positive integer r.

Assumptions A.4–A.6 are standard hypothesis in the nonparametric framework, while A.7 implies that all the moments of $\|X - x\|$ are bounded and holds for a wide family of processes, including the Gaussian ones.

Theorem 5 *Under assumptions A.1–A.7, choosing*

$$c_1 n^{-1/(2p+d)} \leq b_1 \leq c_2 n^{-1/(2p+d)}, \tag{12}$$

$$c_3 n^{-1/(2p+d)} \leq b_2 \leq c_4 n^{-1/(2p+d)}, \tag{13}$$

for some strictly positive constants c_1, c_2, c_3, c_4, and for a given $h > 0$, then $f_{d,n}$, $C_{d,n}(x,h)$ and $\psi_{d,n}(x,h)$ are L^2–consistent estimator of $f_d(x)$, $C_d(x,h)$ and $\psi_d(x,h)$ with the optimal rates

(i) $\mathbb{E}\left[\left(f_{d,n}\left(x\right)-f_{d}\left(x\right)\right)^{2}\right]=O\left(n^{-2p/(2p+d)}\right)$,

(ii) $\mathbb{E}\left[\left(C_{d,n}\left(x,h\right)-C_{d}\left(x,h\right)\right)^{2}\right]=O\left(n^{-2p/(2p+d)}\right)$,

(iii) $\mathbb{E}\left[\left(\psi_{d,n}(x,h)-\psi_{d}(x,h)\right)^{2}\right]=O\left(n^{-2p/(2p+d)}\right)$.

It is worth noticing that the proposed estimators involve the projectors Π_d and $\Pi_d^{\perp}$ that, in practice, are not available and must be estimated from data. This leads to define new estimators plugging the empirical versions of the projectors in (6), (8) and (10). Hence, the resulting estimators are

$$\widehat{\psi}_{d,n}(x,h)=\widehat{f}_{d,n}\left(x\right)\widehat{C}_{d,n}\left(x,h\right) \tag{14}$$

$$\widehat{f}_{d,n}\left(x\right)=\frac{1}{nb_{1}^{d}}\sum_{i=1}^{n}K_{1}\left(\frac{\|\widehat{\Pi}_{d}(X_{i}-x)\|}{b_{1}}\right)$$

$$\widehat{C}_{d,n}\left(x,h\right)=\sum_{i=1}^{n}\widehat{w}_{i}\left(x,b_{2}\right)\left(\left(1-\frac{\|\widehat{\Pi}_{d}^{\perp}(X_{i}-x)\|^{2}}{h^{2}}\right)\mathbb{I}_{\left\{\|\widehat{\Pi}_{d}^{\perp}(X_{i}-x)\|^{2}\leq h^{2}\right\}}\right)^{d/2} \tag{15}$$

with

$$\widehat{w}_{i}\left(x,b_{2}\right)=\frac{K_{2}(\|\widehat{\Pi}_{d}(X_{i}-x)\|/b_{2})}{\sum_{j}K_{2}(\|\widehat{\Pi}_{d}(X_{j}-x)\|/b_{2})}.$$

In the special case $K_1=K_2=K$ and $b_1=b_2=b$, (14) is estimated by

$$\widehat{\psi}_{d,n}(x,h)=\frac{1}{nb^{d}}\sum_{i=1}^{n}\left(\left(1-\frac{\|\widehat{\Pi}_{d}^{\perp}(X_{i}-x)\|^{2}}{h^{2}}\right)\mathbb{I}_{\left\{\|\widehat{\Pi}_{d}^{\perp}(X_{i}-x)\|^{2}\leq h^{2}\right\}}\right)^{d/2}K\left(\frac{\|\widehat{\Pi}_{d}(X_{i}-x)\|}{b}\right).$$

The following results hold for the introduced estimators $\widehat{f}_{d,n}$, $\widehat{C}_{d,n}$ and $\widehat{\psi}_{d,n}$.

Theorem 6 *Under assumptions A.1-A.7 $\widehat{f}_{d,n}\left(x\right)$, $\widehat{C}_{d,n}\left(x,h\right)$ and $\widehat{\psi}_{d,n}(x,h)$ are L^2–consistent estimators of $f_d\left(x\right)$, $C_d\left(x,h\right)$ and $\psi_d(x,h)$ for any $d>2$ and $h>0$. Moreover, if one chooses the optimal bandwidths as in equations (12) and (13) the optimal rates are achieved:*

(i) $\mathbb{E}\left[\left(\widehat{f}_{d,n}\left(x\right)-f_{d}\left(x\right)\right)^{2}\right]=O\left(n^{-2p/(2p+d)}\right)$,

(ii) $\mathbb{E}\left[\left(\widehat{C}_{d,n}\left(x,h\right)-C_{d}\left(x,h\right)\right)^{2}\right]=O\left(n^{-2p/(2p+d)}\right)$,

(iii) $\mathbb{E}\left[\left(\widehat{\psi}_{d,n}(x,h)-\psi_{d}(x,h)\right)^{2}\right]=O\left(n^{-2p/(2p+d)}\right)$.

5 Local dimension selection criterion

Given a sample of curves a typical criterion to approximate them is to use a truncated version of the Karhunen–Loève decomposition. The selection of the dimension of the subspace on which the curves are projected is often based on the well-known fraction of explained variance (FEV), that is

$$FEV\left(d\right) = \frac{\sum_{j=1}^{d} \lambda_j}{\sum_{j=1}^{\infty} \lambda_j} 100\%$$

where λ_j are the eigenvalues of the covariance operator of the process that, since unknown in practice, are usually estimated by their empirical versions. The chosen dimension is heuristically selected in order to obtain a large enough ammount of the FEV (for instance, a reasonable threshold, could be about 90%). This criterium is global in the sense that it selects the same dimension for all the curves of the sample which could be too large (or too small, resp.) for some curves producing inefficient (or inadequate, resp.) representations. Moreover, if the sample size is small, the estimation of eigenfunctions associated to the lowest eigenvalues suffers of problems that worsens the representation quality.

Intuitively, the quality of reconstructions can be improved customizing the choice of dimension for each curve through a local–based methodology. To do this, one can exploit the nature of the correction factor C_d and, in particular, its ability in distinguishing *d–low* and *d–high frequency* curves (see Section 3). What follows illustrates an algorithm that, based on estimates of C_d, provides a local–based reconstruction of a set of curves.

Consider a sample of curves $\{X_i, i = 1, \ldots, n\}$ and a set of centers $\{\chi_j, j = 1, \ldots, N\}$ (that may coincide with the sample). The idea is to reconstruct χ_j over the subspace $H_{d_j^\star} = span\left\{\xi_1, \ldots, \xi_{d_j^\star}\right\}$ where $d_j^\star$ is chosen as the smallest d for which $C_d(\chi_j, h)$ is close enough to the upper bound $M_d(h)$ given in Section 3. Such proximity is heuristically quantified by considering if $\left(M_d(h) - C_d(\chi_j, h)\right)/M_d(h)$ is larger or smaller than α, with $\alpha \in (0, 1)$ being a given threshold. In practice, for all $j = 1, \ldots, N$, the chosen dimension for χ_j is

$$d_j^\star = \min\left\{d \in \{1, \ldots, d_{max}\} : \widehat{C}_{d,n}\left(\chi_j, h\right) \geq (1 - \alpha)\, \widehat{M}_{d,n}(h)\right\}, \tag{16}$$

where d_{max} is a suitable strictly positive integer, $\widehat{C}_{d,n}$ is defined by (15) and the empirical maximum over the sample $\widehat{M}_{d,n}\left(h\right) = \max_i \widehat{C}_{d,n}\left(X_i, h\right)$ provides an estimate of $M_d(h)$. If, for a given curve, an optimal dimension is not found, the procedure assignes to it the dimension $d_{max} + 1$.

If d_{max} can be easily selected by using the FEV, the choice of the parameter α is instead a more delicate matter. Indeed, if α were too small, there would be a risk of systematically chose too large dimensions, producing conservative results and less parsimonious reconstructions. Conversely, if α were too large, the dimensions $d_j^\star$ would be too small, providing imprecise reconstructions. Therefore the choice of α must balance the trade-off between the variability of $d_j^\star$ (over the sample) and the estimation error of the reconstructions. The impact of the choice of α is illustrated in the numerical examples in the following.

5.1 An introductory example

To understand how the algorithm works in practice and what kind of results produces, an illustrative example is here detailed.

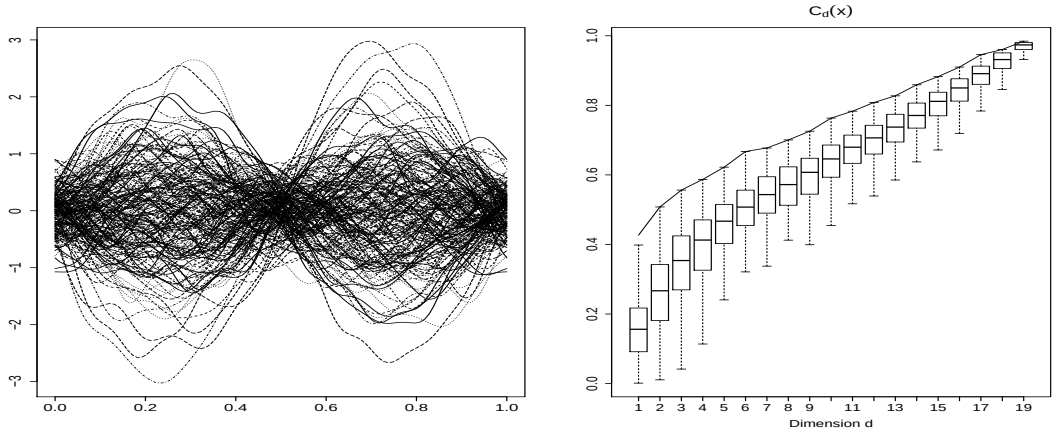

Figure 1: Left: sample of curves. Right: distributions of $\widehat{C}_{d,n}$ varying d and their maxima.

Consider a sample of $n = 200$ curves generated according to

$$X_i(t) = \sum_{q=1}^{Q} \sqrt{\lambda_q} Z_{qi} \phi_q(t) \qquad t \in [0,1],\ i = 1, \ldots, n \tag{17}$$

where $Q = 20$, Z_{qi} are i.i.d. standard Gaussian r.v.s, ϕ_q is the q–th term of the Fourier basis

$$\phi_q(t) = \begin{cases} \sqrt{2}\sin(2\pi m t - \pi) & q = 2m-1 \\ \sqrt{2}\cos(2\pi m t - \pi) & q = 2m \end{cases}$$

and $\lambda_q = 2^{(2q-1)\pi/2}$. Each trajectory is discretized over a grid of 100 equispaced points. The left panel in Figure 1 visualizes that curves.

The first step is to estimate the bounds $M_d(h)$ over this sample used as a training–set. To do this C_d is estimated for $d = 1, \ldots, 19$ by using the nonparametric estimator introduced in Section 4 and then the maxima $\widehat{M}_{d,n}(h)$ are calculated. In this preliminary study, the radius h is chosen as the 10%–quantile of the estimated norms $\left\|\widehat{\Pi}_d(X_i - X_j)\right\|$, whereas the bandwidth b_2 is selected as a percentage (not smaller than 30%) of the range of the same quantitites, following a data–driven approach in order to guarantee that the weights $w_i(\chi_j, b_2)$ are well defined. In the right panel of Figure 1 the distributions of $\widehat{C}_{d,n}(X_i, h)$ for any d and their maxima are depicted.

The second step is to evaluate the algorithm performance in reconstructing curves. Therefore, consider a test–set of $N = 300$ curves $\chi_j(t)$ generated according to (17) with:

- $Q = 2$ for $j = 1, \ldots, 100$,
- $Q = 4$ for $j = 101, \ldots, 200$,
- $Q = 7$ for $j = 201, \ldots, 300$.

Using (16) with $1 - \alpha = 0.99$, the local dimension $d_j^\star$ is estimated for each of these curves, obtaining the frequency distribution plotted in the left panel of Figure 2 with average 3.52 and

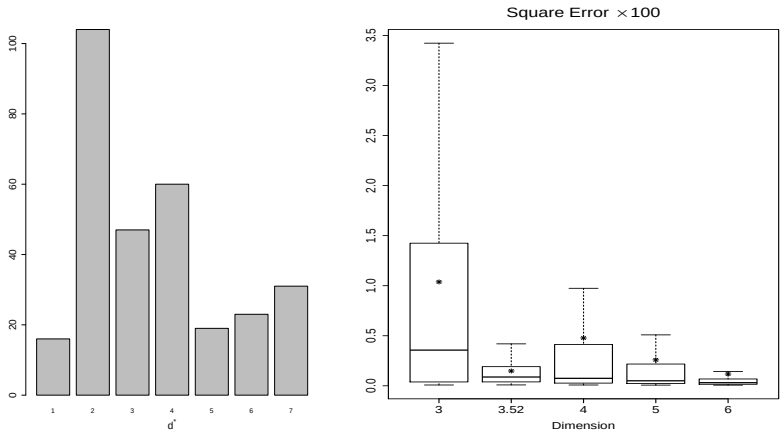

Figure 2: Left: distribution of estimated dimensions $d_j^\star$. Right: distribution of ISEs varying the dimensions ($*$ denotes the mean).

$\downarrow d_j^\star \quad Q \rightarrow$	2	4	7
1	16	0	
2	82	21	1
3	2	34	11
4		45	15
5			19
6			23
7			31
	100	100	100

Table 1: Joint frequency distribution of the selected $d_j^\star$ against the true dimensions Q.

standard deviation 1.76. To appreciate the ability of the algorithm in detecting parsimoniously the dimension of a curve, the joint frequency distribution of the selected $d_j^\star$ and the true dimensions Q of the curves χ_j is reported in Table 1. Naturally, it is reasonable that a curve χ_j generated with a given Q could be well represented in lower dimension than Q, whereas it is desirable that the algorithm does not pick an higher dimension than Q.

To conclude the experiment, the quality of the approximation have to be evaluated. Then, the classical approximations by the Karhunen–Loève expansion $\chi_j^d = \widehat{\Pi}_d \chi_j$ are computed for $d = d_j^\star$ and with a fixed common d for all the curves ($d = 1, \ldots, 19$). The goodness of the approximations is measured by the classical quadratic L^2-norm, namely the Integrated Square Error:

$$ISE_j(d) = \int_0^1 \left(\chi_j(t) - \chi_j^d(t)\right)^2 dt \tag{18}$$

where the integral is approximated by summation. Table 2 collects the means and standard deviations of errors (18) obtained when one uses $d_j^\star$ or a global d (only results with $d \leq 9$ are reported); the values have been multiplied by 100 to improve the readability. For the sake of completeness, the $FEV(d)$, estimated on the data, are provided. It arises that the approximations based on

d	$d^\star$	1	2	3	4	5	6	7	8	9
$FEV\,(d)$	3.52	80.4%	90.1%	93.7%	95.6%	96.9%	97.7%	98.2%	98.5%	98.8%
Mean	0.148	7.055	2.169	1.038	0.477	0.259	0.117	0.027	0.022	0.021
St. dev.	0.152	7.512	2.899	1.461	0.862	0.493	0.266	0.026	0.021	0.020

Table 2: Means and standard deviations of $ISE(d)$ (results are multiplied by 100) obtained when one uses the proposed algorithm or global dimensions.

local dimensions provides a mean error of a comparable order as that corresponding to a global dimension 6, for which the FEV is greater than 97%. To better understand the behaviour of these errors varying the selected dimension, the right panel in Figure 2 visualizes their distributions for $d = 3, 4, 5, 6$ and when $d_j^\star$ are used (in the plot, the average dimension is reported on the horizontal axis). The box–plots highlight the good performances resulting from the employ of the local–based algorithm for dimension selection. Note that if one takes as global dimension $d = 3$ or $d = 4$, that are the closest dimensions to the mean of $d_j^\star$ that is 3.52, the mean error would be much bigger.

5.2 Simulation study

In the following simulation study, an experiment similar to the one conducted in the previous section is replicated many times under different experimental conditions, in order to assess the stability of the introduced procedure and to show what happens when one modifies the parameter α, the radious h of the ball, the size n of the training–set sample and the nature of the process.

More in detail, let consider training–sets of curves generated according to (17) with ϕ_q the q–th term of the Fourier basis, $\lambda_q = 2^{(2q-1)\pi/2}$ and $Q = 20$. The random coefficients Z_{qi} in the expansion are choosen i.i.d. standard Gaussian r.v.s and i.i.d. standardized Student t with 5 degree of freedom: in the first case the PCs are independent, while in the second are only uncorrelated. Samples of small ($n = 50$), medium ($n = 100$) and relatively large sizes ($n = 200$) are used. For what concerns the parameters in the algorithm, $1 - \alpha = 0.9, 0.95, 0.97, 0.99$, $d_{max} = 15$, whereas h should be small enough to guarantee that the SmBP factorization makes sense but avoiding that the balls $\{\|X_i - X_j\| \leq h\}$ are sistematically empty; a reasonable choice is to consider the quantile of the estimated norms $\left\|\widehat{\Pi}_d\left(X_i - X_j\right)\right\|$ of order 5% and 10%. The test–set consists of 10 blocks of 100 curves, generated according to (17), each one with fixed dimension $Q = 1, \ldots, 10$, for a total of $N = 1000$ curves.

For each experimental setting, 1000 Monte Carlo replications are conducted and, for each of them, the mean of the selected dimensions (denoted by $\overline{d}_m^\star$, $m = 1, \ldots, 1000$) and the ISEs (18) are computed.

The box–plots in Figure 3 show the empirical distributions of the means of the errors (values are multiplied by 100) calculated in the Gaussian and Student t case with $n = 200$, when both the global dimensions $d = 3, \ldots, 7$ and the local ones provided by the algorithm are exploited (h is set as the 10%–quantile of the estimated norms and $1 - \alpha = 0.99$). In the graphics, the values on the horizontal axis associated to the box-plots obtained from the proposed algorithm (highlighted by a grey box) are the averages of the means $\overline{d}_m^\star$. Both plots show the good quality of the approximations gotten in reconstructing the curves with customized dimensions. For instance, in the Gaussian case, with a mean dimension of 4.26, results comparable to those obtained using $d = 6$ are obtained ($FEV\,(6)$ is about the 98%): the reconstruction appears choerent with the

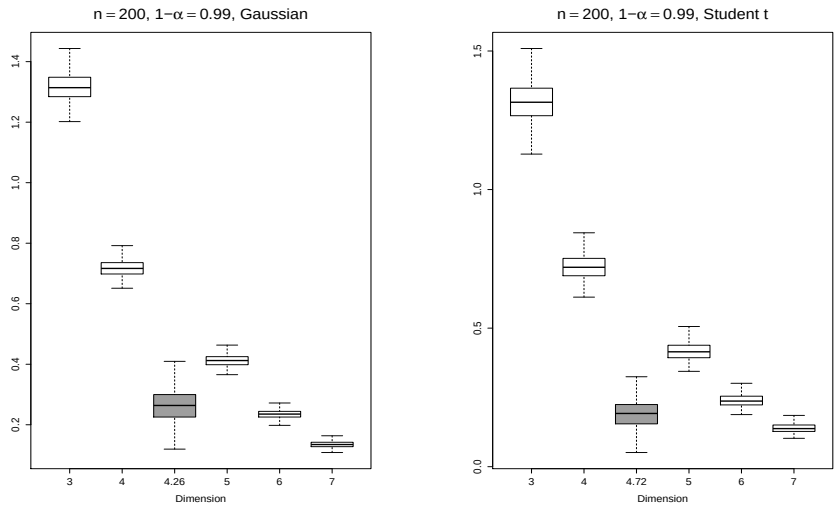

Figure 3: Empirical distributions from 1000 Monte Carlo replications of the means of the ISEs (multiplied by 100) for the Gaussian (left) and the Student t (right) cases. In each panel, the third box–plot from the left refers to the introduced algorithm; on the horizontal axis it is printed the mean of the dimension averages obtained by each Monte Carlo replication.

chosen mixture of the test–set curves and globally parsimonious.

To allow a general overview on the whole experiment, tables 3 and 4 collect a summary of results under the different experimental conditions; the first table belongs to the Gaussian case, whereas the second to the Student t one. The first 8 lines of each table contain the results obtained by applying the new algorithm varying $1-\alpha$, and the remaining 8 the ones with global dimensions $d = 3, 4, 5, 6$. Since, when one uses the algorithm for the different sample sizes n and choices of h, each Monte Carlo replication generates a mean dimension and a mean of the ISEs, in the higher part of the tables a synthesis is provided by the means and the standard deviations of these quantities: the first columns refer to the dimension, the second ones to the error; the standard deviations are in brackets. In the lower part of the table, since the dimensions are fixed, only the means and the standard deviations of the mean of ISEs from each repetition are reported. To allow a more effective reading, errors are multiplied by 100.

As a general comment, the algorithm produces parsimonious and precise reconstructions of the curves in a more efficient way than the FEV selector can do, for almost all the proposed parameter constellations, also for rather small sample sizes. More in detail, one observes that the best performances on the errors are achieved when h is the 10%–quantile of estimated norms and $1-\alpha = 0.99$: in such cases, the mean errors are comparables with those obtained by globally using $d = 6$. Clearly, the employ of the 5%–quantile for h can be reasonable only for large sample sizes. The expected trade-off between α and the selected dimensions appears evident: trying to stay as close as possible to the maximum M_d leads to a more effective representation in terms of error, but less efficient in terms of parsimony ($\overline{d}^{\star}_m$ decreases when $1-\alpha$ increases). Note that, since in this controlled experiment the mean dimension of the test-set is 5.5, the choice $1-\alpha = 0.99$ could be seen as a good compromise for $n = 100$ and $n = 200$, whereas $1-\alpha = 0.95$ can be used for $n = 50$.

6 Applications

In this section, the methodology proposed and illustrated in Section 5 is applied to two real datasets. For both examples, after a brief description of the dataset, the approximation by means of local and global dimensions is performed and its quality is evaluated by using the ISEs, following a cross-validation approach.

6.1 Tecator dataset

The first example illustrated involves the well-known Tecator dataset. It consists of 215 spectra in the near infra-red (NIR) wavelength range from 852 to 1050 nm, discretized on a grid of 100 equispaced points, corresponding to the same number of finely chopped pork samples. The sample of spectrometric curves is visualized in the top panel of Figure 4. The curves appear rather smooth and it is well known that a shift is present: in fact, the empirical $FEV\,(d)$ reaches 99.58% and 99.9% for $d = 2$ and $d = 3$ respectively, and a good representation of the curves can be obtained by using $d = 3$.

One might wonder if the proposed algorithm provides more parsimonious representations mantaining a comparable precision. To do this, the dataset is split randomly in two parts: the first one, containing 200 curves, is used to estimate the bounds $M_d(h)$ for $d = 1, \ldots, 5$, whereas the remaining part $\{\chi_j, j = 1, \ldots, 15\}$ is used to evaluate the local dimensions $d_j^\star$ and the correspondent ISEs. For what concerns the algorithm parameters, h is the 10%-quantile of the estimated norms coherently with the link between this parameter and the sample size (see simulations), $1 - \alpha = 0.85, 0.86, \ldots, 0.99$, and $d_{max} = 5$. Given the small size of the test-set, the procedure is repeated 100 times: in each replication, the means of ISEs obtained using both a global dimension d and the local one are computed, as well as the mean dimension $\overline{d}_m^\star$.

To select a reasonable threshold $1 - \alpha$, which is able to balance parsimony and accuracy, it could be useful to observe the behaviour of the mean errors with respect to the means of selected dimensions and to relate it to the mean errors obtained when a global dimension d is used (see the bottom left panel of Figure 4, where the latter errors are visualized by horizontal lines). From the graphic it emerges that it is better to select $1 - \alpha$ rather small, corresponding to a mean dimension smaller than 2. In fact, choosing for instance $1 - \alpha = 0.87$, the mean of the mean dimensions $\overline{d}_m^\star$ is 1.91 to which corresponds a mean error 0.068: the customization of dimensions produces an efficient representation of the curves, when compared with that which would occur when global dimensions are adopted. This is made evident by looking at the distributions of the mean ISEs (whose values multiplied by 100) plotted in the bottom right panel of Figure 4.

6.2 Neuronal dataset

The second example proposed deals with a dataset coming from a behavioural experiment performed at the Andrew Schwartz motorlab at University of Pittsburgh. In that experiment, a macaque monkey performed a centre-out and out-centre target reaching task with 26 targets in a virtual $3D$ environment while its neural activity was recorded (see Todorova et al. [2014] for details). The dataset collects 1000 curves representing the voltage of neurons versus the time, discretized over a grid of 32 equispaced points normalized between 0 and 1. A sample of 30 of such a curves selected randomly, together with the empirical mean, is shown in the left panel of Figure 5. In the right panel of the same figure, the first three estimated eigenfunctions are shown; note that the process

$n \rightarrow$	50				100				200			
$h \rightarrow$	5%		10%		5%		10%		5%		10%	
$1-\alpha$	$\overline{d}^\star$	ISE	$\overline{d}^\star$	ISE	$\overline{d}^\star$	ISE	$\overline{d}^\star$	ISE	$\overline{d}^\star$	ISE	$\overline{d}^\star$	ISE
0.9	4.032	0.812	3.539	0.868	3.476	0.691	3.166	0.770	3.387	0.590	3.090	0.694
	(0.618)	(0.243)	(0.507)	(0.209)	(0.294)	(0.165)	(0.261)	(0.148)	(0.183)	(0.107)	(0.166)	(0.102)
0.95	4.823	0.646	4.426	0.621	3.973	0.518	3.762	0.519	3.810	0.415	3.638	0.436
	(0.768)	(0.214)	(0.677)	(0.171)	(0.359)	(0.138)	(0.321)	(0.114)	(0.217)	(0.088)	(0.199)	(0.077)
0.97	5.267	0.584	4.984	0.530	4.262	0.455	4.110	0.429	4.014	0.352	3.910	0.346
	(0.846)	(0.204)	(0.773)	(0.156)	(0.423)	(0.129)	(0.375)	(0.102)	(0.237)	(0.080)	(0.220)	(0.068)
0.99	5.808	0.525	5.749	0.443	4.667	0.394	4.628	0.344	4.275	0.294	4.256	0.265
	(0.937)	(0.194)	(0.897)	(0.142)	(0.522)	(0.119)	(0.495)	(0.089)	(0.282)	(0.073)	(0.266)	(0.058)
$\#PC$	ISE				ISE				ISE			
3	1.538				1.387				1.317			
	(0.137)				(0.070)				(0.045)			
4	0.912				0.781				0.718			
	(0.090)				(0.047)				(0.028)			
5	0.587				0.470				0.414			
	(0.068)				(0.037)				(0.023)			
6	0.391				0.285				0.236			
	(0.054)				(0.026)				(0.014)			

Table 3: Gaussian case – Means and standard deviations (in brackets) of mean dimensions and mean ISEs obtained by each Monte Carlo replication varying the sample size, the choice of h and $1-\alpha$, compared with the means and standard deviations of mean ISEs when a global dimension d is used.

$n \rightarrow$	50				100				200			
$h \rightarrow$	5%		10%		5%		10%		5%		10%	
$1-\alpha$	$\overline{d}^\star$	ISE	$\overline{d}^\star$	ISE	$\overline{d}^\star$	ISE	$\overline{d}^\star$	ISE	$\overline{d}^\star$	ISE	$\overline{d}^\star$	ISE
0.9	4.508	0.598	3.951	0.667	3.758	0.502	3.402	0.588	3.571	0.428	3.255	0.523
	(0.727)	(0.196)	(0.638)	(0.188)	(0.609)	(0.127)	(0.505)	(0.137)	(0.490)	(0.096)	(0.415)	(0.109)
0.95	5.377	0.469	4.957	0.470	4.380	0.369	4.126	0.390	4.086	0.296	3.885	0.324
	(0.893)	(0.172)	(0.830)	(0.152)	(0.797)	(0.106)	(0.771)	(0.104)	(0.671)	(0.076)	(0.641)	(0.077)
0.97	5.847	0.421	5.571	0.397	4.742	0.320	4.565	0.317	4.362	0.249	4.233	0.254
	(0.976)	(0.163)	(0.938)	(0.137)	(0.884)	(0.097)	(0.871)	(0.090)	(0.743)	(0.068)	(0.726)	(0.066)
0.99	6.416	0.375	6.392	0.328	5.239	0.273	5.216	0.249	4.737	0.204	4.721	0.190
	(1.075)	(0.154)	(1.079)	(0.124)	(1.011)	(0.088)	(1.025)	(0.077)	(0.829)	(0.060)	(0.837)	(0.054)
$\#PC$	ISE				ISE				ISE			
3	1.591				1.412				1.327			
	(0.217)				(0.148)				(0.118)			
4	0.958				0.794				0.727			
	(0.143)				(0.082)				(0.059)			
5	0.620				0.484				0.419			
	(0.104)				(0.062)				(0.038)			
6	0.416				0.299				0.242			
	(0.078)				(0.045)				(0.029)			

Table 4: Student t case – Means and standard deviations (in brackets) of mean dimensions and mean ISEs obtained by each Monte Carlo replication varying the sample size, the choice of h and $1-\alpha$, compared with the means and standard deviations of mean ISEs when a global dimension d is used.

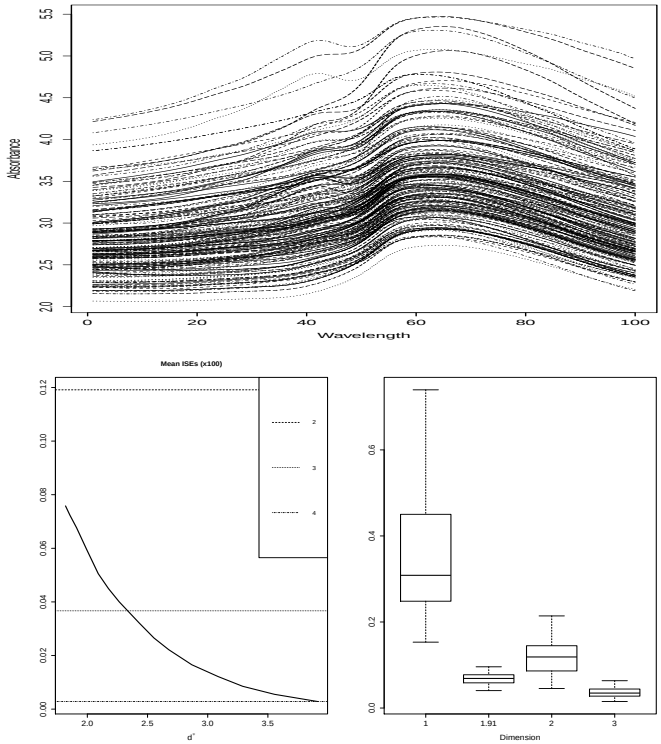

Figure 4: Top: Tecator dataset. Bottom Left: behaviour of mean dimension and mean error. Horizontal lines refer to mean errors for global dimensions. Bottom Right: empirical distributions of the means of the ISEs.

underlying the data appears rather concentrated: the first 3 estimated PCs explain the 97.3% of the total variance, whereas using 6 PCs allows to achieve a FEV equal to 99%.

A two-fold cross-validation is performed: the dataset is randomly split in half, obtaining a training–set of 500 curves used to estimate M_d and a test–set of as many of curves on which the local dimensions and the corresponding approximation errors are computed, as well as the errors got when a global dimension is used. The parameter $1-\alpha$ is set as in the first example whereas, because of the large size of the training–set, h has been selected as the 5%–quantile of the estimated norms. Finally, d_{max} is fixed to 15.

Observing the behaviour of the mean errors varying the means of selected dimensions and comparing it with the mean of ISEs computed when a global dimension is assigned (see the left panel in Figure 6), one can note that with a mean dimension 10.03 (obtained when $1-\alpha = 0.91$) one has the same mean error as with a fixed dimension 11, whereas with a mean dimension 11.02 (corresponding to $1-\alpha = 0.94$) one reaches performances equivalent to those with global dimension 12.

From a parsimonious perspective, it is reasonable to choose $1-\alpha = 0.91$; Figure 6 shows the distributions of selected dimensions $d_j^\star$ (middle panel) and of ISEs for global and local dimensions

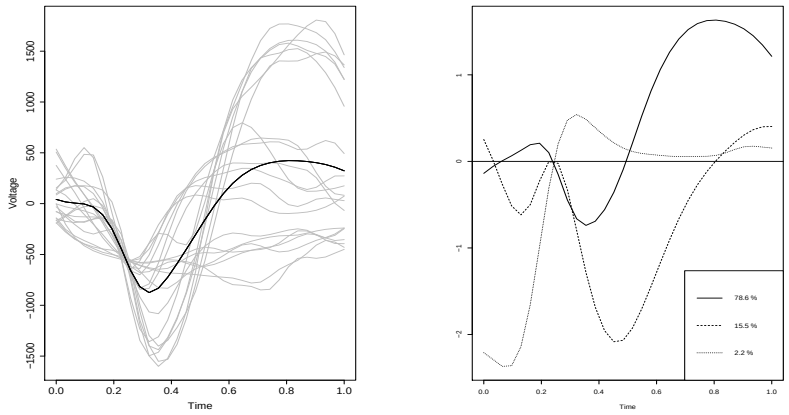

Figure 5: Neuronal experiment. Left: a random sample of 30 curves (in grey) and the empirical mean (in black). Right: the first three estimated eigenfuntions with the corresponding explained variance.

(right panel with the grey boxplot corresponding to the proposed algorithm). Thanks to the curve dimension customization procedure, it is then possible to achieve a better quality of the approximation (both in terms of mean and variability) than that which would be obtained by assigning a global dimension. In particular, the method suggests that few curves could be well represented in low dimension, and to obtain a good approximation quality, one needs from 9 to 12 PCs to approximate the most part of the test–set.

A Appendix

A.1 Sketch of the proof of Proposition 1

Fix $d \in \mathbb{N}$, $x \in H$ and define

$$S_1 = S_1(d,x) = \sum_{j \leq d} \left(\theta_j - \langle x, \xi_j \rangle\right)^2, \qquad S = S(d,h,x) = \frac{1}{h^2} \sum_{j \geq d+1} \left(\theta_j - \langle x, \xi_j \rangle\right)^2,$$

then applying the Parseval's identity,

$$\begin{aligned}\varphi(x,h) &= \mathbb{P}\left(\|X - x\|^2 \leq h^2\right) = \mathbb{P}\left(S_1 + h^2 S \leq h^2\right) = \mathbb{P}\left(\left\{S_1 \leq h^2\left(1 - S\right)\right\} \cap \left\{0 \leq S \leq 1\right\}\right) \\ &= \int_0^1 \mathbb{P}\left(S_1 \leq (1-S)h^2 \middle| S = s\right) f_S\left(s\right) ds\end{aligned}$$

where f_S is the pdf of S that is strictly positive on $(0,1)$ thanks to A.3.b. Now, the first order approximation gives, for $h \to 0$,

$$\mathbb{P}\left(S_1 \leq (1-S)h^2 \middle| S = s\right) \sim f_{d|S}\left(\Pi_d x | s\right) V_d\left(h\sqrt{1-s}\,\right),$$

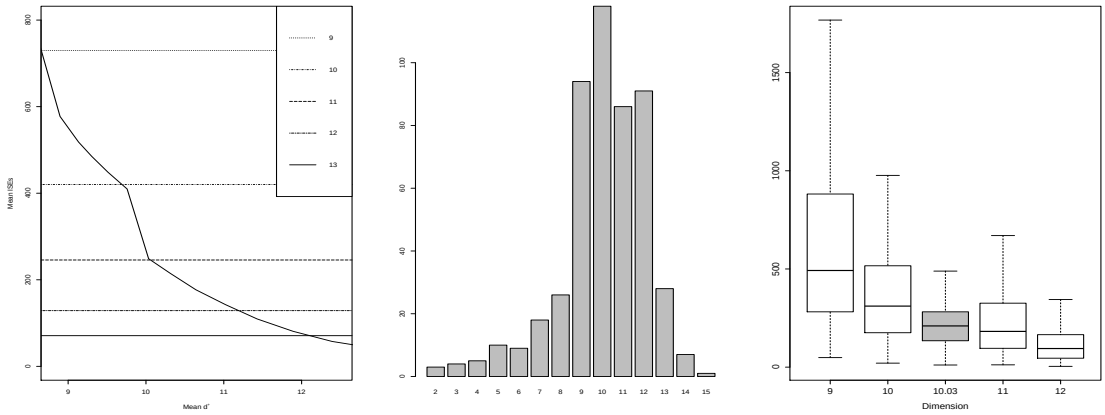

Figure 6: Neuronal experiment. Left: behaviour of mean dimension and mean error. Middle: the distribution of selected dimensions. Right: distributions of ISEs in the test–sample.

where $f_{d|S}$ is the pdf of $(\boldsymbol{\theta}|S)$ that is $f_d f_{S|d}/f_S$ and $V_d(h)$ is the volume of $B_d(0,h) \subset \mathbb{R}^d$. Note that both $f_{d|S}$ $f_{S|d}$ are well defined and strictly positive thanks to A.1 and A.3. Finally, Bayes's theorem applied to $f_{d|S}$ provides the considered factorization, as $h \to 0$,

$$\begin{aligned}\varphi(h,x) \sim & f_d\left(\Pi_d x\right) V_d(h) \int_0^1 f_{S|d}\left(s|\Pi_d x\right)(1-s)^{d/2}\, ds \\ = & f_d\left(\Pi_d x\right) V_d(h) \mathbb{E}\left[\left((1-S)\,\mathbb{I}_{\{0\le S\le 1\}}\right)^{d/2}\middle|\,\Pi_d x\right].\end{aligned}$$

A.2 Proofs of propositions of Section 3

Proof of Proposition 2. Consider Equation (4). The lower bound is a consequence of the definition of C_1. For what concerns the upper bound, by Jensen inequality one gets

$$\begin{aligned}C_1(x,h) = & \mathbb{E}\left[\left(\left(1-\frac{\|\Pi_1^{\perp}(X-x)\|^2}{h^2}\right)\mathbb{I}_{A_{1,h}}\right)^{1/2}\middle|\, x_1\right] \\ \le & \left(\mathbb{E}\left[\left(1-\frac{\|\Pi_1^{\perp}(X-x)\|^2}{h^2}\right)\mathbb{I}_{A_{1,h}}\middle|\, x_1\right]\right)^{1/2}\end{aligned}$$

where

$$\begin{aligned}\mathbb{E}\left[\left(1-\frac{\|\Pi_1^{\perp}(X-x)\|^2}{h^2}\right)\mathbb{I}_{A_{1,h}}\middle|\, x_1\right] = & \mathbb{E}\left[1-\frac{\|\Pi_1^{\perp}(X-x)\|^2}{h^2}\middle|\,\mathbb{I}_{A_{1,h}}=1, x_1\right]\mathbb{P}\left(A_{1,h}|x_1\right) \\ \le & \mathbb{E}\left[1-\frac{\|\Pi_1^{\perp}(X-x)\|^2}{h^2}\middle|\,\mathbb{I}_{A_{1,h}}=1, x_1\right].\end{aligned}$$

Consider now Equation (5). In order to prove the lower bound, consider the Bernoulli inequality

$(1-s)^r \geq 1-rs$ for any $s \in [0,1]$ and $r \geq 1$ to get

$$\begin{aligned}
C_d(x,h) &= \mathbb{E}\left[\left(\left(1-\frac{\|\Pi_d^{\perp}(X-x)\|^2}{h^2}\right)\mathbb{I}_{A_{d,h}}\right)^{d/2}\middle|\,\Pi_d x\right] \\
&\geq \mathbb{E}\left[\left(1-\frac{d}{2h^2}\|\Pi_d^{\perp}(X-x)\|^2\right)\mathbb{I}_{A_{d,h}}\middle|\,\Pi_d x\right] \\
&= \mathbb{E}\left[1-\frac{d}{2h^2}\|\Pi_d^{\perp}(X-x)\|^2\middle|\,\mathbb{I}_{A_{d,h}}=1,\Pi_d x\right]\mathbb{P}(A_{d,h}|\Pi_d x) \\
&= \mathbb{P}(A_{d,h}|\Pi_d x) - \frac{d}{2h^2}\mathbb{E}\left[\|\Pi_d^{\perp}(X-x)\|^2\middle|\,\mathbb{I}_{A_{d,h}}=1,\Pi_d x\right]\mathbb{P}(A_{d,h}|\Pi_d x) \\
&\geq \mathbb{P}(A_{d,h}|\Pi_d x) - \frac{d}{2h^2}\mathbb{E}\left[\|\Pi_d^{\perp}(X-x)\|^2\middle|\,\mathbb{I}_{A_{d,h}}=1,\Pi_d x\right].
\end{aligned}$$

Applying Markov inequality it holds

$$\begin{aligned}
C_d(x,h) &\geq 1-\frac{1}{h^2}\mathbb{E}\left[\|\Pi_d^{\perp}(X-x)\|^2\middle|\,\mathbb{I}_{A_{d,h}}=1,\Pi_d x\right]-\frac{d}{2h^2}\mathbb{E}\left[\|\Pi_d^{\perp}(X-x)\|^2\middle|\,\mathbb{I}_{A_{d,h}}=1,\Pi_d x\right] \\
&= 1-\frac{d+2}{2h^2}\mathbb{E}\left[\|\Pi_d^{\perp}(X-x)\|^2\middle|\,\mathbb{I}_{A_{d,h}}=1,\Pi_d x\right].
\end{aligned}$$

For what concerns the upper bound, noting that $(1-s)^{r/2} \leq 1-s$ for any $s \in [0,1]$ and $r \geq 2$, one has

$$\begin{aligned}
C_d(x,h) &= \mathbb{E}\left[\left(\left(1-\frac{\|\Pi_d^{\perp}(X-x)\|^2}{h^2}\right)\mathbb{I}_{A_{d,h}}\right)^{d/2}\middle|\,\Pi_d x\right] \\
&\leq \mathbb{E}\left[\left(1-\frac{\|\Pi_d^{\perp}(X-x)\|^2}{h^2}\right)\mathbb{I}_{A_{d,h}}\middle|\,\Pi_d x\right] \\
&= \mathbb{E}\left[1-\frac{\|\Pi_d^{\perp}(X-x)\|^2}{h^2}\middle|\,\mathbb{I}_{A_{d,h}}=1,\Pi_d x\right]\mathbb{P}(A_{d,h}) \\
&\leq 1-\frac{1}{h^2}\mathbb{E}\left[\|\Pi_d^{\perp}(X-x)\|^2\middle|\,\mathbb{I}_{A_{d,h}}=1,\Pi_d x\right].
\end{aligned}$$

■

Proof of Proposition 3. In this proof, the following notations are used

$$x_{(d)} = \Pi_d x, \qquad x_{(d^{\perp})} = \Pi_d^{\perp} x, \qquad \|\cdot\|_{d^{\perp}} = \|\Pi_d^{\perp}\cdot\|.$$

$$C_d(x,h) = C_d\left(x_{(d)}+x_{(d^{\perp})},h\right) = \mathbb{E}\left[\left(\left(1-\frac{\|X-x\|_{d^{\perp}}^2}{h^2}\right)\mathbb{I}_{\left\{\|X-x\|_{d^{\perp}}^2 \leq h^2\right\}}\right)^{d/2}\middle|\,x_{(d)}\right]. \tag{19}$$

Clearly, $x_{(d)} \in H_d = span\{\xi_1,\ldots,\xi_d\}$, $x_{(d^{\perp})} \in H_d^{\perp}$, $x = x_{(d)}+x_{(d^{\perp})}$, $\|(X-x)_{(d^{\perp})}\|_{d^{\perp}}^2 = \|X-x\|_{d^{\perp}}^2$ and

$$M_d(h) = \sup_{x\in H} C_d(x,h) = \sup_{x_{(d)}\in H_d}\sup_{x_{(d^{\perp})}\in H_d^{\perp}} C_d\left(x_{(d)}+x_{(d^{\perp})},h\right).$$

Given the fact that $\left(\left(1-\frac{\|X-x\|_{d^\perp}^2}{h^2}\right)\mathbb{I}_{\left\{\|X-x\|_{d^\perp}^2\leq h^2\right\}}\right)^{d/2}$ is uncorrelated with $\left\{X_{(d)}=x_{(d)}\right\}$, the conditioning in (19) can be dropped leading to

$$C_d\left(x,h\right)=C_d\left(x_{(d^\perp)},h\right)=\mathbb{E}\left[\left(\left(1-\frac{\|X-x\|_{d^\perp}^2}{h^2}\right)\mathbb{I}_{\left\{\|X-x\|_{d^\perp}^2\leq h^2\right\}}\right)^{d/2}\right] \tag{20}$$

$$=\int_{H_d^\perp}\left(\left(1-\frac{\|y-x\|_{d^\perp}^2}{h^2}\right)\mathbb{I}_{\left\{\|y-x\|_{d^\perp}^2\leq h^2\right\}}\right)^{d/2}d\mathbb{P}_{X_{(d^\perp)}}\left(y\right) \tag{21}$$

where $\mathbb{P}_{X_{(d^\perp)}}$ is the probability law of $X_{(d^\perp)}$. Hence, $M_d\left(h\right)=\sup_{x_{(d^\perp)}\in H_d^\perp}C_d\left(x_{(d^\perp)},h\right)$ and to get the thesis, it is enough to prove that $M_d\left(h\right)=C_d\left(0,h\right)$ where 0 here denotes the null element in $H_d^\perp$.

Note that $M_d\left(h\right)$ is strictly positive thanks to A.3.b and since the expectation argument in (20) is strictly positive over $\left\{\|X-x\|_{d^\perp}^2<h^2\right\}$ and null otherwise. Moreover, by definition of supremum, consider a sequence $\left\{x\left(n\right)\right\}_{n\in\mathbb{N}}\subset H_d^\perp$ such that $C_d\left(x\left(n\right),h\right)\to M_d\left(h\right)$ as $n\to\infty$. Such a sequence must be bounded (i.e. $\|x\left(n\right)\|_{d^\perp}\leq c$) otherwise there will be a subsequence $\left\{x\left(n_k\right)\right\}$ for which $\|x\left(n_k\right)\|_{d^\perp}\to\infty$ as $k\to\infty$ and, by reverse triangular inequality $\|X-x\left(n_k\right)\|_{d^\perp}^2\geq\left(\|X\|_{d^\perp}-\|x\left(n_k\right)\|_{d^\perp}\right)^2$, it holds $\mathbb{I}_{\left\{\|X-x(n_k)\|_{d^\perp}^2\leq h^2\right\}}\leq\mathbb{I}_{\left\{\left(\|X\|_{d^\perp}-\|x(n_k)\|_{d^\perp}\right)^2\leq h^2\right\}}$ and

$$C_d\left(x\left(n_k\right),h\right)\leq\mathbb{E}\left[\mathbb{I}_{\left\{\left(\|X\|_{d^\perp}-\|x(n_k)\|_{d^\perp}\right)^2\leq h^2\right\}}\right]\to 0<M_d\left(h\right),\qquad k\to\infty,$$

that contradicts the fact that $M_d\left(h\right)>0$. Thus, since $\left\{x\left(n\right)\right\}$ is bounded, there exists a subsequence that, without loss of generality is still denoted by $\left\{x\left(n\right)\right\}$ and, weakly converges to an element $m\in H_d^\perp$ (i.e. $\left\langle x\left(n\right),y\right\rangle\to\left\langle m,y\right\rangle$ for any $y\in H_d^\perp$ as $n\to\infty$).

The element m is a maximizer of $C_d\left(x_{(d^\perp)},h\right)$ over $H_d^\perp$ if $\left\{x\left(n\right)\right\}$ strongly converges to it. Indeed in this case, by the dominate convergence theorem, it is possible to exchange the limit and the integral obtaining the following chain of equalities

$$\begin{aligned}M_d\left(h\right)&=\lim_{n\to\infty}C_d\left(x\left(n\right),h\right)=\lim_{n\to\infty}\int_{H_d^\perp}\left(\left(1-\frac{\|y-x\left(n\right)\|_{d^\perp}^2}{h^2}\right)\mathbb{I}_{\left\{\|y-x(n)\|_{d^\perp}^2\leq h^2\right\}}\right)^{d/2}d\mathbb{P}_{X_{(d^\perp)}}\left(y\right)\\&=\int_{H_d^\perp}\lim_{n\to\infty}\left(\left(1-\frac{\|y-x\left(n\right)\|_{d^\perp}^2}{h^2}\right)\mathbb{I}_{\left\{\|y-x(n)\|_{d^\perp}^2\leq h^2\right\}}\right)^{d/2}d\mathbb{P}_{X_{(d^\perp)}}\left(y\right)\\&=C_d\left(m,h\right).\end{aligned}$$

In what follows it is shown that $\left\{x\left(n\right)\right\}$ strongly converges to m in $H_d^\perp$. To do this, consider $\delta>0$, $n\in\mathbb{N}$ and the subset of

$$\begin{aligned}A_\delta&=\left\{y\in H_d^\perp:\liminf\|y-x\left(n\right)\|_{d^\perp}^2>\|y-m\|_{d^\perp}^2+\delta\right\},\\A_\delta^n&=\left\{y\in H_d^\perp:\|y-x\left(k\right)\|_{d^\perp}^2>\|y-m\|_{d^\perp}^2+\delta,\forall k\geq n\right\}\end{aligned}$$

for which $\mathbb{P}_{X_{(d^\perp)}}\left(A_\delta^n\right)\to\mathbb{P}_{X_{(d^\perp)}}\left(A_\delta\right)$ as $n\to\infty$. By contradiction assume that $\left\{x\left(n\right)\right\}$ does not strongly converge to m, then it is known that $\liminf\|y-x\left(n\right)\|_{d^\perp}^2>\|y-m\|_{d^\perp}^2$ and $\lim_{\delta\to0^+}\mathbb{P}_{X_{(d^\perp)}}\left(A_\delta\right)=$

1. Fix $\delta_0 < h^2$ such that $\mathbb{P}_{X_{(d^\perp)}}(A_\delta) > 0$. By limit definition for any $\varepsilon > 0$, there exists $\delta < \delta_0$ such that $\mathbb{P}(A_\delta) > 1-\varepsilon$ or equivalently $\mathbb{P}_{X_{(d^\perp)}}\left(A_\delta^C\right) < \varepsilon$. Thus, denoting by $f(X-x) = \left(\left(1-\|X-x\|_{d^\perp}^2/h^2\right)\mathbb{I}_{\left\{\|X-x\|_{d^\perp}^2 \le h^2\right\}}\right)^{d/2}$, it holds

$$\begin{aligned}\lim_{n\to\infty}\left(C_d(x(n),h) - C_d(m,h)\right) &= \lim_{n\to\infty}\int_{H_d^\perp}\left(f(y-x(n)) - f(y-m)\right)d\mathbb{P}_{X_{(d^\perp)}}(y)\\ &= \lim_{n\to\infty}\left(\int_{A_{\delta_0}^n}\left(f(y-x(n)) - f(y-m)\right)d\mathbb{P}_{X_{(d^\perp)}}(y) + \right.\\ &\int_{\left(A_{\delta_0}^n\right)^C\cap A_\delta}\left(f(y-x(n)) - f(y-m)\right)d\mathbb{P}_{X_{(d^\perp)}}(y) +\\ &\left.\int_{\left(A_{\delta_0}^n\right)^C\cap (A_\delta)^C}\left(f(y-x(n)) - f(y-m)\right)d\mathbb{P}_{X_{(d^\perp)}}(y)\right). \qquad (22)\end{aligned}$$

By definition of $A_{\delta_0}^n$, one has

$$\begin{aligned}&\int_{A_{\delta_0}^n}\left(f(y-x(n)) - f(y-m)\right)d\mathbb{P}_{X_{(d^\perp)}}(y)\\ &= \int_{A_{\delta_0}^n}\left[\left(1-\frac{\|y-x(n)\|_{d^\perp}^2}{h^2}\right)^{d/2}\mathbb{I}_{\left\{\|y-x(n)\|_{d^\perp}^2\le h^2\right\}} - \left(1-\frac{\|y-m\|_{d^\perp}^2}{h^2}\right)^{d/2}\mathbb{I}_{\left\{\|y-m\|_{d^\perp}^2\le h^2\right\}}\right]d\mathbb{P}_{X_{(d^\perp)}}(y)\\ &\le \int_{A_{\delta_0}^n}\left[\left(1-\frac{\|y-m\|_{d^\perp}^2}{h^2}-\frac{\delta_0}{h^2}\right)^{d/2} - \left(1-\frac{\|y-m\|_{d^\perp}^2}{h^2}\right)^{d/2}\right]\mathbb{I}_{\left\{\|y-m\|_{d^\perp}^2+\delta_0\le h^2\right\}}d\mathbb{P}_{X_{(d^\perp)}}(y)\end{aligned}$$

and, because $\left(1-\|y-m\|_{d^\perp}^2/h^2-\delta_0/h^2\right)^{d/2} < \left(1-\|y-m\|_{d^\perp}^2/h^2\right)^{d/2}$, there exists $\widetilde{\delta}_0 > 0$ such that

$$\int_{A_{\delta_0}^n}\left(f(y-x(n)) - f(y-m)\right)d\mathbb{P}_{X_{(d^\perp)}}(y) \le -\widetilde{\delta}_0\mathbb{P}_{X_{(d^\perp)}}\left(\left\{\|X-m\|_{d^\perp}^2 \le h^2-\delta_0\right\}\cap A_{\delta_0}^n\right) = -\eta_0.$$

Because $A_{\delta_0}^n \supset A_\delta$, $\left(A_{\delta_0}^n\right)^C \cap A_\delta = \varnothing$ and the second addend in (22) is null. Moreover, since $\left(A_{\delta_0}^n\right)^C \cap (A_\delta)^C \subset (A_\delta)^C$ one has

$$\begin{aligned}\int_{\left(A_{\delta_0}^n\right)^C\cap(A_\delta)^C}\left(f(y-x(n)) - f(y-m)\right)d\mathbb{P}_{X_{(d^\perp)}}(y) &\le \int_{(A_\delta)^C}\left(f(y-x(n)) - f(y-m)\right)d\mathbb{P}_{X_{(d^\perp)}}(y)\\ &\le \mathbb{P}_{X_{(d^\perp)}}\left((A_\delta)^C\right) < \varepsilon\end{aligned}$$

and, thanks to the arbitrariness of ε, the third addend in (22) is not positive. Thus,

$$M_d(h) = \lim_{n\to\infty} C_d(x(n),h) \le C_d(m,h) - \eta_0$$

that is in contrast with the supremum definition of $M_d(h)$ and thus $\{x(n)\}$ must strongly converges to m.

It remains to prove that $m = 0$ (the zero element in $H_d^{\perp}$) or equivalently that $\|m\|_{d^{\perp}} = 0$. The proof procedes by contradiction using similar arguments as above and for this reason some details are skipped. In particular, given a sequence $\{x(n)\}$ strongly convergent to m, the following sets are considered

$$B_{\delta} = \left\{ y \in H_d^{\perp} : \liminf \|y - x(n)\|_{d^{\perp}}^2 > \|y\|_{d^{\perp}}^2 + \delta \right\},$$
$$B_{\delta}^n = \left\{ y \in H_d^{\perp} : \|y - x(k)\|_{d^{\perp}}^2 > \|y\|_{d^{\perp}}^2 + \delta, \forall k \geq n \right\}$$

for which $\mathbb{P}_{X_{(d^{\perp})}}(B_{\delta}^n) \to \mathbb{P}_{X_{(d^{\perp})}}(B_{\delta})$ as $n \to \infty$. By contradiction assume that $\|m\|_{d^{\perp}} > 0$, then $\liminf \|y - x(n)\|_{d^{\perp}}^2 > \|y\|_{d^{\perp}}^2$ and $\lim_{\delta \to 0^+} \mathbb{P}_{X_{(d^{\perp})}}(B_{\delta}) = 1$. Fix $\delta_0 < h^2$ such that $\mathbb{P}_{X_{(d^{\perp})}}(B_{\delta}) > 0$. By limit definition, for any $\varepsilon > 0$ there exists $\delta < \delta_0$ such that $\mathbb{P}(B_{\delta}) > 1 - \varepsilon$ or equivalently $\mathbb{P}_{X_{(d^{\perp})}}\left(B_{\delta}^C\right) < \varepsilon$. Thus, using similar arguments as above,

$$\lim_{n \to \infty} \left(C_d(x(n), h) - C_d(0, h) \right) \leq \eta_0$$

and, then

$$M_d(h) = \lim_{n \to \infty} C_d(x(n), h) \leq C_d(0, h) - \eta_0$$

that, once again, contrasts with the supremum definition of $M_d(h)$ and implies that $m = 0 \in H_d^{\perp}$. ■

Proof of Proposition 4. Note that

$$\begin{aligned}
\mathbb{E}\left[C_d(X', h)\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(\left(1 - \frac{\|\Pi_d^{\perp}(X - X')\|^2}{h^2}\right)\mathbb{I}_{\left\{\|\Pi_d^{\perp}(X - X')\|^2 \leq h^2\right\}}\right)^{d/2} \middle| \Pi_d X'\right]\right] \\
&= \mathbb{E}\left[\left(\left(1 - \frac{\|\Pi_d^{\perp}(X - X')\|^2}{h^2}\right)\mathbb{I}_{\left\{\|\Pi_d^{\perp}(X - X')\|^2 \leq h^2\right\}}\right)^{d/2}\right].
\end{aligned}$$

Because $C_d(X', h) \leq 1$, then $\mathbb{E}\left[C_d(X', h)\right] = 1$ is equivalent to $C_d(X', h) = 1$ a.s. to $\|\Pi_d^{\perp}(X - X')\|^2 = 0$ a.s. and to $\mathbb{E}\left[\|\Pi_d^{\perp}(X - X')\|^2\right] = 0$ (see e.g. [Shiryayev, 1996, p.185]) that proves $i) \Leftrightarrow ii)$. Now, using the explicit expression of $\Pi_d^{\perp} X$, thanks to the orthonormality of the eigenfunctions and the independence of X and X' (and consequently of their PCs), then

$$\begin{aligned}
\mathbb{E}\left[\|\Pi_d^{\perp}(X - X')\|^2\right] &= \mathbb{E}\left[\|\Pi_d^{\perp} X\|^2\right] + \mathbb{E}\left[\|\Pi_d^{\perp} X'\|^2\right] - 2\mathbb{E}\left[\left\langle \Pi_d^{\perp} X, \Pi_d^{\perp} X' \right\rangle\right] \\
&= 2\sum_{j \geq d+1} \lambda_j - 2\mathbb{E}\left[\left\langle \sum_{j \geq d+1} \theta_j \xi_j, \sum_{j \geq d+1} \theta_j' \xi_j \right\rangle\right] \\
&= 2\sum_{j \geq d+1} \lambda_j - 2\sum_{j \geq d+1} \mathbb{E}\left[\theta_j \theta_j'\right] = 2\sum_{j \geq d+1} \lambda_j.
\end{aligned}$$

Hence, $ii) \Leftrightarrow iii)$ if and only if $\lambda_j = 0$ for any $j \geq d + 1$. Finally, $iii) \Leftrightarrow iv)$ is a consequence of the fact that, for a zero mean process, the Karhunen–Loève expansion is $X(t) = \sum_{j \geq 1} \theta_j \xi_j(t)$, $0 \leq t \leq 1$, with $\mathbb{E}[\theta_j] = 0$ and $Var(\theta_j) = \lambda_j$ with $j \geq 1$. ■

A.3 Proofs of theorems in Section 4

Proof of Theorem 5. The statement (i) is a consequence of classical results on kernel density estimation (see Stone [1980]) which guarantees that, for a fixed x, d and h, if one chooses the optimal bandwidth (12), the optimal rate is $n^{-2p/(2p+d)}$. On the other hand, because $0 \leq Y \leq 1$ (see definition (7)), classical results on kernel regression estimators applies (see Stone [1982]) and then, choosing the bandwidth (13), the optimal rate (ii) is achieved. For what concerns (iii), it holds

$$\begin{aligned}\mathbb{E}\left[\left(\psi_{d,n}(x,h)-\psi_d(x,h)\right)^2\right] \leq\ & 2\mathbb{E}\left[\left(f_{d,n}\left(x\right)C_{d,n}\left(x,h\right)-f_d\left(x\right)C_{d,n}\left(x,h\right)\right)^2\right] \\ & +2\mathbb{E}\left[\left(f_d\left(x\right)C_{d,n}\left(x,h\right)-f_d\left(x\right)C_d\left(x,h\right)\right)^2\right].\end{aligned}$$

Since $C_{d,n}$, $C_d \leq 1$ and $f_d \leq m$ (a strictly positive constant),

$$\begin{aligned}\mathbb{E}\left[\left(\psi_{d,n}(x,h)-\psi_d(x,h)\right)^2\right] \leq\ & 2\mathbb{E}\left[\left(f_{d,n}\left(x\right)-f_d\left(x\right)\right)^2\right] \\ & +2m\mathbb{E}\left[\left(C_{d,n}\left(x,h\right)-C_d\left(x,h\right)\right)^2\right]\end{aligned}$$

and finally thanks to (i) and (ii) the thesis is completed. ■

Next proof invokes some technical Lemmas that are stated and proved in the next section.

Proof of Theorem 6. For what concerns (i), note that

$$\mathbb{E}\left[\left(\widehat{f}_{d,n}\left(x\right)-f_d\left(x\right)\right)^2\right] \leq 2\mathbb{E}\left[\left(\widehat{f}_{d,n}\left(x\right)-f_{d,n}\left(x\right)\right)^2\right]+2\mathbb{E}\left[\left(f_{d,n}\left(x\right)-f_d\left(x\right)\right)^2\right]$$

where, thanks to Lemma 7,

$$\mathbb{E}\left[\left(\widehat{f}_{d,n}\left(x\right)-f_{d,n}\left(x\right)\right)^2\right]=o\left(n^{-2p/(2p+d)}\right)$$

while Theorem 5 guarantees that

$$\mathbb{E}\left[\left(f_{d,n}\left(x\right)-f_d\left(x\right)\right)^2\right]=O\left(n^{-2p/(2p+d)}\right).$$

For the second statement (ii), consider

$$\mathbb{E}\left[\left(\widehat{C}_{d,n}\left(x,h\right)-C_d\left(x,h\right)\right)^2\right] \leq 2\mathbb{E}\left[\left(\widehat{C}_{d,n}\left(x,h\right)-C_{d,n}\left(x,h\right)\right)^2\right]+2\mathbb{E}\left[\left(C_{d,n}\left(x,h\right)-C_d\left(x,h\right)\right)^2\right].$$

The second addend is $O\left(n^{-2p/(2p+d)}\right)$ (see Theorem 5), while for the first one it holds

$$\mathbb{E}\left[\left(\widehat{C}_{d,n}\left(x,h\right)-C_{d,n}\left(x,h\right)\right)^2\right] \leq 2\mathbb{E}\left[\left(\widehat{C}_{d,n}\left(x,h\right)-\widetilde{C}_{d,n}\left(x,h\right)\right)^2\right]+2\mathbb{E}\left[\left(\widetilde{C}_{d,n}\left(x,h\right)-C_{d,n}\left(x,h\right)\right)^2\right]$$

where

$$\widetilde{C}_{d,n}\left(x,h\right)=\sum_{i=1}^{n}\left(\left(1-\frac{\|\Pi_d^{\perp}(X_i-x)\|^2}{h^2}\right)\mathbb{I}_{\left\{\|\Pi_d^{\perp}(X_i-x)\|^2\leq h^2\right\}}\right)^{d/2}\frac{K_2(\|\widehat{\Pi}_d(X_i-x)\|/b_2)}{\sum_j K_2(\|\widehat{\Pi}_d(X_j-x)\|/b_2)}.$$

Note that, differently from $\widehat{C}_{d,n}$, in $\widetilde{C}_{d,n}$ the projector $\Pi_d^{\perp}$ is not estimated as if Y (see Equation (7)) was fully observed without errors. Lemma 8 provides that, for $d > 2$,

$$\mathbb{E}\left[\left(\widetilde{C}_{d,n}\left(x,h\right) - C_{d,n}\left(x,h\right)\right)^2\right] = o\left(n^{-2p/(2p+d)}\right),$$

whereas Lemma 9 gives

$$\mathbb{E}\left[\left(\widehat{C}_{d,n}\left(x,h\right) - \widetilde{C}_{d,n}\left(x,h\right)\right)^2\right] = o\left(n^{-2p/(2p+d)}\right).$$

Finally, to prove (iii), consider

$$\begin{aligned}\mathbb{E}\left[\left(\widehat{\psi}_{d,n}(x,h) - \psi_d(x,h)\right)^2\right] \leq\ & 2\mathbb{E}\left[\left(\widehat{\psi}_{d,n}(x,h) - \psi_{d,n}(x,h)\right)^2\right] \\ & + 2\mathbb{E}\left[\left(\psi_{d,n}(x,h) - \psi_d(x,h)\right)^2\right].\end{aligned}$$

Properties of the second addend are explored in Theorem 5, while for the first addend, because $0 \leq \widehat{C}_{d,n} \leq 1$ and $f_{d,n} \leq m$, it holds

$$\begin{aligned}\mathbb{E}\left[\left(\widehat{\psi}_{d,n}(x,h) - \psi_{d,n}(x,h)\right)^2\right] &\leq 2\mathbb{E}\left[\left(\widehat{f}_{d,n}\left(x\right)\widehat{C}_{d,n}\left(x,h\right) - f_{d,n}\left(x\right)\widehat{C}_{d,n}\left(x,h\right)\right)^2\right] + \\ &\quad 2\mathbb{E}\left[\left(f_{d,n}\left(x\right)\widehat{C}_{d,n}\left(x,h\right) - f_{d,n}\left(x\right)C_{d,n}\left(x,h\right)\right)^2\right] \\ &\leq 2\mathbb{E}\left[\left(\widehat{f}_{d,n}\left(x\right) - f_{d,n}\left(x\right)\right)^2\right] + 2m\mathbb{E}\left[\left(\widehat{C}_{d,n}\left(x,h\right) - C_{d,n}\left(x,h\right)\right)^2\right].\end{aligned}$$

Using (i) and (ii) in the latter the thesis is achieved. ■

A.4 Technical Lemmas

In this section, the following notations are considered

$$S_n\left(x\right) = \sum_{i=1}^{n} K\left(\frac{\left\|\Pi_d\left(X_i - x\right)\right\|}{b_n}\right), \qquad \widehat{S}_n\left(x\right) = \sum_{i=1}^{n} K\left(\frac{\left\|\widehat{\Pi}_d\left(X_i - x\right)\right\|}{b_n}\right) \tag{23}$$

$$Z_n\left(x\right) = \sum_{i=1}^{n} Y_i K\left(\frac{\left\|\Pi_d\left(X_i - x\right)\right\|}{b_n}\right), \qquad \widehat{Z}_n\left(x\right) = \sum_{i=1}^{n} Y_i K\left(\frac{\left\|\widehat{\Pi}_d\left(X_i - x\right)\right\|}{b_n}\right) \tag{24}$$

so that

$$\begin{aligned} f_{d,n}\left(x\right) &= \frac{S_n\left(x\right)}{nb_n^d}, \qquad \widehat{f}_{d,n}\left(x\right) = \frac{\widehat{S}_n\left(x\right)}{nb_n^d} \\ C_{d,n}\left(x,h\right) &= \frac{Z_n\left(x\right)}{S_n\left(x\right)}, \qquad \widetilde{C}_{d,n}\left(x,h\right) = \frac{\widehat{Z}_n\left(x\right)}{\widehat{S}_n\left(x\right)}\end{aligned}$$

and the following events

$$A_i = \{V_i \leq b_n\} \qquad\qquad B_i = \left\{\widehat{V}_i \leq b_n\right\}$$

where Y_i is defined in (7), $V_i = \|\Pi_d (X_i - x)\|$, $\widehat{V}_i = \left\|\widehat{\Pi}_d (X_i - x)\right\|$ and, to unburden notations, b_n denotes $b_{1,n}$ or $b_{2,n}$ according to the setting.

Lemma 7 *Assume A.1–A.7, then for $d > 2$, if one chooses the optimal bandwidth like in (12), then*

$$\mathbb{E}\left[\left(\widehat{f}_{d,n}(x) - f_{d,n}(x)\right)^2\right] = o\left(n^{-2p/(2p+d)}\right).$$

Proof. The proof follows similar arguments of those in Biau and Mas [2012] where the same thesis is achieved under the assumption that X is bounded and only giving some hints for a more general setting. In that paper, it is shown that

$$\begin{aligned}
\mathbb{E}\left[\left(\widehat{f}_{d,n}(x) - f_{d,n}(x)\right)^2\right] &= \frac{1}{\left(nb_n^d\right)^2}\mathbb{E}\left[\left(S_n(x) - \widehat{S}_n(x)\right)^2\right] \\
&\leq \frac{c}{\left(nb_n^d\right)^2}\mathbb{E}\left[\left(\frac{\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty}{b_n}\left(\sum_{i=1}^n \|X_i - x\|\,\mathbb{I}_{A_i}\right)\right)^2\right] \\
&\quad + \frac{c}{\left(nb_n^d\right)^2}\mathbb{E}\left[\left(\sum_{i=1}^n \left(\mathbb{I}_{A_i \cap \overline{B}_i} + \mathbb{I}_{\overline{A}_i \cap B_i}\right)\right)^2\right] \qquad (25)
\end{aligned}$$

where c denotes a general strictly positive constant.

FIRST STEP. Consider the first addend of the right hand side of (25); for $i = 1, \ldots, n$, let $U_i = \|X_i - x\|\,\mathbb{I}_{A_i} - \mathbb{E}\left[\|X_i - x\|\,\mathbb{I}_{A_i}\right]$ and write

$$\begin{aligned}
\mathbb{E}\left[\left(\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty \left(\sum_{i=1}^n \|X_i - x\|\,\mathbb{I}_{A_i}\right)\right)^2\right] &\leq 2\mathbb{E}\left[\left(\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty \left(\sum_{i=1}^n U_i\right)\right)^2\right] \\
&\quad + 2n^2\left\{\mathbb{E}\left[\|X_1 - x\|\,\mathbb{I}_{A_1}\right]\right\}^2 \mathbb{E}\left[\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty^2\right]. \qquad (26)
\end{aligned}$$

Results in Biau and Mas [2012] guarantee that

$$\mathbb{E}\left[\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty^2\right] = O\left(\frac{1}{n}\right), \qquad (27)$$

while the fact that $Cov\left(\|X_1 - x\|, \mathbb{I}_{A_1}\right) \leq 0$, Assumption A.7 and $\mathbb{E}\left[\mathbb{I}_{A_i}\right] \sim b_n^d$ give

$$\left\{\mathbb{E}\left[\|X_1 - x\|\,\mathbb{I}_{A_1}\right]\right\}^2 \leq \left(\mathbb{E}\left[\|X_1 - x\|\right]\right)^2 \left(\mathbb{E}\left[\mathbb{I}_{A_1}\right]\right)^2 = O\left(b_n^{2d}\right). \qquad (28)$$

Hence, (27) and (28) provide

$$2n^2\left\{\mathbb{E}\left[\|X_i - x\|\,\mathbb{I}_{A_i}\right]\right\}^2 \mathbb{E}\left[\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty^2\right] = O\left(nb_n^{2d}\right). \qquad (29)$$

For what concerns the first addend in the right hand side of (26), applying Chauchy–Schwarz inequality one gets

$$\mathbb{E}\left[\left(\left\|\Pi_d-\widehat{\Pi}_d\right\|_\infty\left(\sum_{i=1}^n U_i\right)\right)^2\right]\leq\sqrt{\mathbb{E}\left[\left\|\Pi_d-\widehat{\Pi}_d\right\|_\infty^4\right]\mathbb{E}\left[\left(\sum_{i=1}^n U_i\right)^4\right]}. \tag{30}$$

Results in Biau and Mas [2012] guarantee

$$\mathbb{E}\left[\left\|\Pi_d-\widehat{\Pi}_d\right\|_\infty^4\right]=O\left(\frac{1}{n^2}\right), \tag{31}$$

while, denoting by $W_i=\|X_i-x\|\,\mathbb{I}_{A_i}$ and by $\overline{W}_n$ the mean of $\{W_i\}$,

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^n U_i\right)^4\right]&=\mathbb{E}\left[\left(\sum_{i=1}^n\left(W_i-\mathbb{E}\left[W_i\right]\right)\right)^4\right]=\mathbb{E}\left[\left(n\overline{W}_n-n\mathbb{E}\left[W_1\right]\right)^4\right]\\&=n^4\mathbb{E}\left[\left(\overline{W}_n-\mathbb{E}\left[W_1\right]\right)^4\right].\end{aligned}$$

The latter fourth moment can be bounded by using the second centered moment of $W=\|X-x\|\,\mathbb{I}_A$ (see, as instance, Dodge and Rousson [1999]) as follows

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^n U_i\right)^4\right]&=n^4\left(\frac{\mathbb{E}\left[\left(W-\mathbb{E}\left[W\right]\right)^4\right]}{n^3}+\frac{3\left(n-1\right)}{n^3}\left(\mathbb{E}\left[\left(W-\mathbb{E}\left[W\right]\right)^2\right]\right)^2\right)\\&=O\left(n^2\left(\mathbb{E}\left[W^2\right]-\left(\mathbb{E}\left[W\right]\right)^2\right)^2\right)=O\left(n^2\left(\mathbb{E}\left[W^2\right]\right)^2\right)+O\left(n^2\left(\mathbb{E}\left[W\right]\right)^4\right),\end{aligned}$$

that, using similar arguments to get (28), gives

$$\mathbb{E}\left[\left(\sum_{i=1}^n U_i\right)^4\right]=O\left(n^2b^{2d}\right). \tag{32}$$

Combining (31) and (32) in (30) one gets

$$\mathbb{E}\left[\left(\left\|\Pi_d-\widehat{\Pi}_d\right\|_\infty\left(\sum_{i=1}^n U_i\right)\right)^2\right]=O\left(b_n^d\right). \tag{33}$$

Hence, using (33), (29) in (26) one gets that the first addend of the right hand side of (25) can be bounded as follows

$$\frac{c}{\left(nb_n^{d+1}\right)^2}\mathbb{E}\left[\left(\left\|\Pi_d-\widehat{\Pi}_d\right\|_\infty\left(\sum_{i=1}^n\|X_i-x\|\,\mathbb{I}_{A_i}\right)\right)^2\right]=\frac{c}{\left(nb_n^{d+1}\right)^2}\left(O\left(b_n^d\right)+O\left(nb_n^{2d}\right)\right)=O\left(\frac{1}{nb_n^2}\right).$$

Finally, choosing the optimal bandwidth like in (12), it holds that $1/\left(nb_n^2\right)=o\left(n^{-2p/(2p+d)}\right)$.

SECOND STEP. For what concerns the second addend of the right hand side of (25) note that

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\left(\mathbb{I}_{A_i\cap\overline{B}_i}+\mathbb{I}_{\overline{A}_i\cap B_i}\right)\right)^2\right]\leq 2\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{A_i\cap\overline{B}_i}\right)^2\right]+2\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{\overline{A}_i\cap B_i}\right)^2\right]$$

and, because the two addend can be treated similarly, it is sufficient to focus on the first addend. To do this, define the sequence κ_n, with $\kappa_n\to 0$ as $n\to\infty$, and using the same argument as in [Biau and Mas, 2012, Lemma 5.5] it holds

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{A_i\cap\overline{B}_i}\right)^2\right]\leq 2\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{\{b_n(1-\kappa_n)<V_i\leq b_n\}}\right)^2\right]+2\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{\left\{\left\|\widehat{\Pi}_d-\Pi_d\right\|\|X_i-x\|>\kappa_n b_n\right\}}\mathbb{I}_{\{V_i\leq b_n\}}\right)^2\right]. \tag{34}$$

and

$$\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{\{b_n(1-\kappa_n)<V_i\leq b_n\}}\right)^2\right]=O\left(n^2b_n^{2d}\kappa_n^2\right). \tag{35}$$

About the second addend in (34), define the sequence t_n, with $t_n\to\infty$ as $n\to\infty$ and consider

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{\left\{\left\|\widehat{\Pi}_d-\Pi_d\right\|\|X_i-x\|>\kappa_n b_n\right\}}\mathbb{I}_{\{V_i\leq b_n\}}\right)^2\right]\\
&=\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{\left\{\left\|\widehat{\Pi}_d-\Pi_d\right\|\|X_i-x\|>\kappa_n b_n\right\}}\mathbb{I}_{\{V_i\leq b_n\}}\left(\mathbb{I}_{\{\|X_i-x\|\leq t_n\}}+\mathbb{I}_{\{\|X_i-x\|>t_n\}}\right)\right)^2\right]\\
&\leq 2\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{\left\{\left\|\widehat{\Pi}_d-\Pi_d\right\|\|X_i-x\|>\kappa_n b_n\right\}}\mathbb{I}_{\{\|X_i-x\|\leq t_n\}}\mathbb{I}_{\{V_i\leq b_n\}}\right)^2\right]\\
&\quad+2\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{\left\{\left\|\widehat{\Pi}_d-\Pi_d\right\|\|X_i-x\|>\kappa_n b_n\right\}}\mathbb{I}_{\{\|X_i-x\|>t_n\}}\mathbb{I}_{\{V_i\leq b_n\}}\right)^2\right]\\
&\leq 2\mathbb{E}\left[\mathbb{I}_{\left\{\left\|\widehat{\Pi}_d-\Pi_d\right\|>\kappa_n b_n/t_n\right\}}\left(\sum_{i=1}^{n}\mathbb{I}_{\{V_i\leq b_n\}}\right)^2\right]\\
&\quad+2\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{\{\|X_i-x\|>t_n\}}\mathbb{I}_{\{V_i\leq b_n\}}\right)^2\right]
\end{aligned} \tag{36}$$

Results in [Biau and Mas, 2012, Lemma 5.5] give

$$\mathbb{E}\left[\mathbb{I}_{\left\{\left\|\widehat{\Pi}_d-\Pi_d\right\|>\kappa_n b_n/t_n\right\}}\left(\sum_{i=1}^{n}\mathbb{I}_{\{V_i\leq b_n\}}\right)^2\right]=n^2O\left(\exp\left(-cn\frac{\kappa_n^2b_n^2}{t_n^2}\right)\right). \tag{37}$$

For what concerns the second addend in the right hand side of (36), as a consequence of Jensen inequality, the fact that $Cov\left(\mathbb{I}_{\{\|X_1-x\|>t_n\}},\mathbb{I}_{\{\|\Pi_d(X_1-x)\|\leq b_n\}}\right)\leq 0$, $\mathbb{E}\left[\mathbb{I}_{\{\|\Pi_d(X_1-x)\|\leq b_n\}}\right]\sim b_n^d$ and

Assumption A.7, then one has, for any strictly positive integer r,

$$\begin{aligned}
\mathbb{E}\left[\left(\sum_{i=1}^{n}\mathbb{I}_{\{\|X_i-x\|>t_n\}}\mathbb{I}_{\{V_i\leq b_n\}}\right)^2\right] &= n^2\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{\{\|X_i-x\|>t_n\}}\mathbb{I}_{\{V_i\leq b_n\}}\right)^2\right] \\
&\leq n^2\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{\{\|X_i-x\|>t_n\}}\mathbb{I}_{\{V_i\leq b_n\}}\right] \\
&= n^2\mathbb{E}\left[\mathbb{I}_{\{\|X_1-x\|>t_n\}}\mathbb{I}_{\{\|\Pi_d(X_1-x)\|\leq b_n\}}\right] \\
&\leq n^2\mathbb{E}\left[\mathbb{I}_{\{\|X_1-x\|>t_n\}}\right]\mathbb{E}\left[\mathbb{I}_{\{\|\Pi_d(X_1-x)\|\leq b_n\}}\right] \\
&\sim n^2\mathbb{P}\left(\|X_1-x\|>t_n\right)b_n^d \\
&= o\left(n^2b_n^de^{-rt_n}\right). \qquad (38)
\end{aligned}$$

Thus, combining the bounds (34), (35), (36), (37) and (38) one has

$$\frac{c}{\left(nb_n^d\right)^2}\mathbb{E}\left[\left(\sum_{i=1}^{n}\left(\mathbb{I}_{A_i\cap\overline{B}_i}+\mathbb{I}_{\overline{A}_i\cap B_i}\right)\right)^2\right] = O\left(\kappa_n^2\right)+O\left(\frac{\exp\left(-cn\kappa_n^2b_n^2/t_n^2\right)}{b_n^{2d}}\right)+o\left(\frac{\exp\left(-rt_n\right)}{b_n^d}\right). \qquad (39)$$

Finally, to prove that the latter bound is $o\left(n^{-2p/(2p+d)}\right)$ we have to choose b_n, κ_n and t_n. Once again b_n is chosen to be the optimal bandwidth like in (12), whereas for κ_n let consider the following choice,

$$c_5\frac{\log^8\left(nb_n^2\right)}{nb_n^2}\leq\kappa_n^2\leq c_6\frac{\log^8\left(nb_n^2\right)}{nb_n^2},\qquad\text{for some } 0<c_5<c_6<+\infty$$

which guarantees that the first addend of the right hand side of (39) and κ_n^2 are $o\left(n^{-2p/(2p+d)}\right)$, for any $d>2$. For what concerns t_n, note that, because of the last two addends of the right hand side of (39), it should satisfy $n\kappa_n^2b_n^2/t_n^2\to\infty$ and $t_n\to\infty$. Hence

$$t_n^2=o\left(n\kappa_n^2b_n^2\right)=o\left(\log^8\left(nb_n^2\right)\right)$$

so that a suitable choice for t_n might be

$$c_7\log^4\left(nb_n^2\right)\leq t_n^2\leq c_8\log^4\left(nb_n^2\right),\qquad\text{for some } 0<c_7<c_8<+\infty$$

for which $b_n^{-2d}\exp\left(-Cn\kappa_n^2b_n^2/t_n^2\right)$ and e^{-rt_n}/b_n^d are $o\left(n^{-2p/(2p+d)}\right)$. This concludes the proof. ■

Lemma 8 *Assume A.1–A.7, then for $d>2$,*

$$\mathbb{E}\left[\left(\widetilde{C}_{d,n}\left(x,h\right)-C_{d,n}\left(x,h\right)\right)^2\right]=o\left(n^{-2p/(2p+d)}\right).$$

Proof. Consider the decomposition

$$\widetilde{C}_{d,n}\left(x,h\right)-C_{d,n}\left(x,h\right)=\frac{\widehat{Z}_n\left(x\right)}{\widehat{S}_n\left(x\right)}\frac{S_n\left(x\right)-\widehat{S}_n\left(x\right)}{S_n\left(x\right)}+\frac{1}{S_n\left(x\right)}\left(\widehat{Z}_n\left(x\right)-Z_n\left(x\right)\right)$$

where the involved objects are defined in (23) and (24). Since $0 \leq Y_i \leq 1$, similar arguments as in Biau and Mas [2012] lead to write

$$\mathbb{E}\left[\left(\widetilde{C}_{d,n}(x,h) - C_{d,n}(x,h)\right)^2\right] \leq c\frac{\mathbb{E}\left[\left(S_n(x) - \widehat{S}_n(x)\right)^2\right]}{\left(nb_n^d\right)^2} + c\frac{\mathbb{E}\left[\left(\widehat{Z}_n(x) - Z_n(x)\right)^2\right]}{\left(nb_n^d\right)^2}$$

where

$$\frac{\mathbb{E}\left[\left(\widehat{Z}_n(x) - Z_n(x)\right)^2\right]}{\left(nb_n^d\right)^2} \leq \frac{\mathbb{E}\left[\left(S_n(x) - \widehat{S}_n(x)\right)^2\right]}{\left(nb_n^d\right)^2} = \mathbb{E}\left[\left(f_{d,n}(x) - \widehat{f}_{d,n}(x)\right)^2\right].$$

Lemma 7 gives the thesis. ■

Lemma 9 *Assume A.1–A.7, then for $d > 2$,*

$$\mathbb{E}\left[\left(\widehat{C}_{d,n}(x,h) - \widetilde{C}_{d,n}(x,h)\right)^2\right] = o\left(n^{-2p/(2p+d)}\right)$$

Proof. Consider

$$\left|\widehat{C}_{d,n}(x,h) - \widetilde{C}_{d,n}(x,h)\right| \leq \frac{\sum_{i=1}^n \left|\widehat{Y}_i - Y_i\right| K_2(\|\widehat{\Pi}_d(X_i - x)\|/b_2)}{\sum_j K_2(\|\widehat{\Pi}_d(X_j - x)\|/b_2)}. \tag{40}$$

About the numerator in (40), note that because $\left(\left(1 - \frac{t}{h^2}\right)\mathbb{I}_{\{t \leq h^2\}}\right)^{d/2}$ is a Lipschitz function with respect to t on $[0, \infty)$, then there exists a strictly positive constant L' such that

$$\begin{aligned}\left|\widehat{Y}_i - Y_i\right| &= \left|\left(\left(1 - \frac{\left\|\Pi_d^\perp(X_i - x)\right\|^2}{h^2}\right)\mathbb{I}_{\left\{\|\Pi_d^\perp(X_i - x)\|^2 \leq h^2\right\}}\right)^{d/2} - \left(\left(1 - \frac{\left\|\widehat{\Pi}_d^\perp(X_i - x)\right\|^2}{h^2}\right)\mathbb{I}_{\left\{\|\widehat{\Pi}_d^\perp(X_i - x)\|^2 \leq h^2\right\}}\right)^{d/2}\right| \\ &\leq L'\left|\left\|\Pi_d^\perp(X_i - x)\right\|^2 - \left\|\widehat{\Pi}_d^\perp(X_i - x)\right\|^2\right|.\end{aligned}$$

Now, Parseval identity allows to write:

$$\begin{aligned}\left\|\Pi_d^\perp(X_i - x)\right\|^2 - \left\|\widehat{\Pi}_d^\perp(X_i - x)\right\|^2 &= \|X_i - x\|^2 - \|\Pi_d(X_i - x)\|^2 - \left(\|X_i - x\|^2 - \left\|\widehat{\Pi}_d(X_i - x)\right\|^2\right) \\ &= \left\|\widehat{\Pi}_d(X_i - x)\right\|^2 - \|\Pi_d(X_i - x)\|^2\end{aligned}$$

and hence,

$$\begin{aligned}\left|\widehat{Y}_i - Y_i\right| &\leq L'\left|\left\|\widehat{\Pi}_d(X_i - x)\right\|^2 - \|\Pi_d(X_i - x)\|^2\right| \\ &\leq L\left|\|\widehat{\Pi}_d(X_i - x)\| - \|\Pi_d(X_i - x)\|\right|\end{aligned}$$

where the latter inequality is a consequence of $g(t) = t^2$ being Lipschitz.
The reverse triangle inequality, the definition of operatorial norm and the boundedness of the projectors give

$$\begin{aligned}\left|\|\widehat{\Pi}_d(X_i - x)\| - \|\Pi_d(X_i - x)\|\right| &\leq \left\|\widehat{\Pi}_d(X_i - x) - \Pi_d(X_i - x)\right\| \\ &\leq \left\|\widehat{\Pi}_d - \Pi_d\right\|_\infty \|X_i - x\|,\end{aligned}$$

and thus, because the strictly positiveness of K_2, the numerator of the left hand side of (40) can be bounded as follows

$$\sum_{i=1}^{n} \left|\widehat{Y}_i - Y_i\right| K_2\left(\frac{\left\|\widehat{\Pi}_d(X_i - x)\right\|}{b_2}\right) \leq L\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty \sum_{i=1}^{n} \|X_i - x\| K_2\left(\frac{\left\|\widehat{\Pi}_d(X_i - x)\right\|}{b_2}\right). \quad (41)$$

Since K_2 is a decreasing function, the algebraic Chebychev inequality (see, for instance, [Mitrinović et al., 1993, page 243]) gives

$$\frac{1}{n}\sum_{i=1}^{n} \|X_i - x\| K_2\left(\frac{\left\|\widehat{\Pi}_d(X_i - x)\right\|}{b_2}\right) \leq \frac{1}{n}\sum_{i=1}^{n} \|X_i - x\| \frac{1}{n}\sum_{i=1}^{n} K_2\left(\frac{\left\|\widehat{\Pi}_d(X_i - x)\right\|}{b_2}\right) \quad (42)$$

and so, combining equations (40), (41) and (42), it follows

$$\left|\frac{\sum_{i=1}^{n}\left(\widehat{Y}_i - Y_i\right) K_2(\|\widehat{\Pi}_d(X_i - x)\|/b_2)}{\sum_{i=1}^{n} K_2\left(\left\|\widehat{\Pi}_d(X_i - x)\right\|/b_2\right)}\right| \leq L\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty \left(\frac{1}{n}\sum_{i=1}^{n}\|X_i - x\|\right).$$

Thus

$$\begin{aligned}\mathbb{E}\left[\left(\widehat{C}_{d,n}(x,h) - \widetilde{C}_{d,n}(x,h)\right)^2\right] &\leq L^2\mathbb{E}\left[\left(\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty \left(\frac{1}{n}\sum_{i=1}^{n}\|X_i - x\|\right)\right)^2\right] \\ &\leq L^2\sqrt{\mathbb{E}\left[\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty^4\right]\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\|X_i - x\|\right)^4\right]}\end{aligned}$$

where the latter follows from the Chauchy–Schwarz inequality. Results in [Biau and Mas, 2012, Theorem 2.1 (iii)] provides

$$\mathbb{E}\left[\left\|\Pi_d - \widehat{\Pi}_d\right\|_\infty^4\right] = O\left(\frac{1}{n^2}\right)$$

while the strong law of large numbers and Assumption A.7 guarantee that

$$\frac{1}{n}\sum_{i=1}^{n}\|X_i - x\| = \frac{1}{n}\sum_{i=1}^{n}\|X_i - x\| - \mathbb{E}[\|X - x\|] + \mathbb{E}[\|X - x\|] = O_P(1).$$

Finally, the last two bounds give

$$\mathbb{E}\left[\left(\widehat{C}_{d,n}(x,h) - \widetilde{C}_{d,n}(x,h)\right)^2\right] = O\left(\frac{1}{n}\right) = o\left(n^{-2p/(2p+d)}\right).$$

■

References

G. Aneiros, E. G. Bongiorno, R. Cao, and P. Vieu. *Functional Statistics and Related Fields*. Springer, 2017.

G. Aneiros, R. Cao, R. Fraiman, C. Genest, and P. Vieu. Recent advances in functional data analysis and high-dimensional statistics. *Journal of Multivariate Analysis*, 170:3–9, 2019a.

G. Aneiros, R. Cao, and P. Vieu. Editorial on the special issue on functional data analysis and related topics. *Computat. Stat.*, 34:447–450, 2019b.

G. Aneiros, I. Horová, M. Hušková, and P. Vieu. *Functional and High-Dimensional Statistics and Related Fields*. Springer, 2020.

N. Bathia, Q. Yao, F. Ziegelmann, et al. Identifying the finite dimensionality of curve time series. *The Annals of Statistics*, 38(6):3352–3386, 2010.

G. Biau and A. Mas. PCA-kernel estimation. *Stat. Risk Model.*, 29(1):19–46, 2012.

E. G. Bongiorno and A. Goia. Classification methods for hilbert data based on surrogate density. *Comput. Statist. Data Anal.*, 99:204 – 222, 2016. ISSN 0167-9473. doi: http://dx.doi.org/10.1016/j.csda.2016.01.019.

E. G. Bongiorno and A. Goia. Some insights about the small ball probability factorization for Hilbert random elements. *Statist. Sinica*, 27:1949–1965, 2017.

E. G. Bongiorno and A. Goia. Corrections on some insights about the small ball probability factorization for hilbert random elements. *Statistica Sinica*, pages 549–549, 2018.

E. G. Bongiorno, A. Goia, E. Salinelli, and P. Vieu, editors. *Contributions in infinite-dimensional statistics and related topics*, 2014. Società Editrice Esculapio.

E. G. Bongiorno, A. Goia, and P. Vieu. Evaluating the complexity of functional data. *Preprint*, 2017.

E. G. Bongiorno, A. Goia, and P. Vieu. Modeling functional data: a test procedure. *Computational Statistics*, 34(2):451–468, 2019.

E. G. Bongiorno, A. Goia, and P. Vieu. Estimating the complexity index of functional data: Some asymptotics. *Statistics & Probability Letters*, 161:108731, 2020.

D. Bosq. *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2000.

A. Delaigle and P. Hall. Defining probability density for a distribution of random functions. *Ann. Statist.*, 38(2):1171–1193, 2010.

Y. Dodge and V. Rousson. The complications of the fourth central moment. *The American Statistician*, 53(3):267–269, 1999.

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

F. Ferraty, N. Kudraszow, and P. Vieu. Nonparametric estimation of a surrogate density function in infinite-dimensional spaces. *J. Nonparametr. Stat.*, 24(2):447–464, 2012.

F. Ferraty, P. Kokoszka, J.-L. Wang, and Y. Wu. Editorial for the special issue on high-dimensional and functional data analysis. *Comput. Statist. Data Anal.*, 131:10–11, 2019.

T. Gasser, P. Hall, and B. Presnell. Nonparametric estimation of the mode of a distribution of random curves. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(4):681–691, 1998.

A. Goia and P. Vieu. An introduction to recent advances in high/infinite dimensional statistics [Editorial]. *J. Multivariate Anal.*, 146:1–6, 2016. ISSN 0047-259X.

P. Hall and C. Vial. Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(4):689–705, 2006.

L. Horváth and P. Kokoszka. *Inference for functional data with applications*. Springer Series in Statistics. Springer, New York, 2012. ISBN 978-1-4614-3654-6.

T. Hsing and R. Eubank. *Theoretical foundations of functional data analysis, with an introduction to linear operators*, volume 997. John Wiley & Sons, 2015.

J. Jacques and C. Preda. Model-based clustering for multivariate functional data. *Comput. Statist. Data Anal.*, 71:92–106, 2014.

P. Kokoszka and M. Reimherr. *Introduction to functional data analysis*. CRC Press, 2017.

W. V. Li and Q.-M. Shao. Gaussian processes: inequalities, small ball probabilities and applications. In *Stochastic processes: theory and methods*, volume 19 of *Handbook of Statist.*, pages 533–597. North-Holland, Amsterdam, 2001.

M. A. Lifshits. *Lectures on Gaussian processes*. Springer Briefs in Mathematics. Springer, Heidelberg, 2012.

D. S. Mitrinović, J. E. Pečarić, and A. M. Fink. *Classical and new inequalities in analysis*, volume 61 of *Mathematics and its Applications (East European Series)*. Kluwer Academic Publishers Group, Dordrecht, 1993. ISBN 0-7923-2064-6. doi: 10.1007/978-94-017-1043-5.

J. O. Ramsay and B. W. Silverman. *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition, 2005.

A. N. Shiryayev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996.

C. J. Stone. Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, 8(6):1348–1360, 1980. ISSN 0090-5364.

C. J. Stone. Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, 10 (4):1040–1053, 1982. ISSN 0090-5364.

S. Todorova, P. Sadtler, A. Batista, S. Chase, and V. Ventura. To sort or not to sort: the impact of spike-sorting on neural decoding performance. *J. Neural. Eng.*, 11(5):056005, 2014.