ELSEVIER

Statistics & Probability Letters 48 (2000) 335–345

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Kernel-based functional principal components ☆

Graciela Boente[a,b,*], Ricardo Fraiman[a,c]

[a] Universidad de Buenos Aires, Argentina
[b] CONICET, Argentina
[c] Universidad de San Andrés, Argentina

Received November 1998; received in revised form November 1999

Abstract

In this paper, we propose kernel-based smooth estimates of the functional principal components when data are continuous trajectories of stochastic processes. Strong consistency and the asymptotic distribution are derived under mild conditions.

MSC: primary 62G07; 62H25

Keywords: Functional principal components; Kernel methods; Hilbert–Schmidt operators; Eigenfunctions

1. Introduction

In many situations, the individual observed responses are curves rather than finite-dimensional vectors as, for instance, in some growth curve models. In this context, the observable response of each individual may be modeled as a sampling path $X(t,\omega)$, $\omega \in \Omega$, of a stochastic process with expected value $\mu(t)$ and covariance function $\gamma(t,s)$, for t, s in a finite interval $\mathscr{I}$.

Ramsay (1982), Hart and Wehrly (1986), Rice and Silverman (1991), Ramsay and Dalzell (1991) and Fraiman and Pérez Iribarren (1991), among others, discussed further examples and applications in this general setting. Some of these works focussed the problem of estimating the mean curve and the covariance function of the underlying process. While some others went further on and also analyzed the covariance structure through the so-called functional principal component analysis.

The functional principal component problem is mainly principal components analysis when data are curves, instead of finite-dimensional vectors. This problem was analyzed by Dauxois et al. (1982) where asymptotic

☆ This research was partially supported by Grants EX-038 and TX-49 from the Universidad de Buenos Aires, PIP #4186 from the CONICET and PICT # 03-00000-00576 from ANPCYT at Buenos Aires, Argentina.

* Correspondence address: Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón 1, Buenos Aires, 1428, Argentina.
E-mail address: gboente@mate.dm.uba.ar (G. Boente)

0167-7152/00/$ - see front matter
PII: S0167-7152(00)00014-6

properties of non-smooth principal components of functional data were derived. Further analysis of this problem has been developed by Besse and Ramsay (1986), Rice and Silverman (1991), Ramsay and Dalzell (1991), Pezzulli and Silverman (1993), Silverman (1996) and Ramsay and Silverman (1997), where, in particular, smooth principal components for functional data – based on roughness penalty methods – were considered. Several examples and applications can be found in these references.

In Ramsay and Silverman (1997), it is pointed out that principal components analysis of functional data is a key technique to be considered in functional analysis, in order to explore the data to see features characterizing typical functions. There, it is stated that “some indication of the complexity of the data is also required, in the sense of how many types of curves and characteristics are to be found. Principal components analysis serves these ends admirably,...”. They also argue for smoothness properties: “A second issue is that of regularization; for many data sets, PCA of functional data is more revealing if some type of smoothness is required to the principal components themselves”.

In this paper, we propose kernel-based principal components for functional data, and studied their asymptotic properties. There are two common ways of performing smooth principal component analysis. The first is to smooth the functional data and then perform PCA. The second is to directly define smoothed the principal components. This can be achieved, for example, by adding a penalty term to the sample variance and maximizing the penalized sample variance, as described in Ramsay and Silverman (1997). If a kernel-based smoothing method is used, it will be shown that both approaches are the same. On the other hand, the kernel-based approach allows to derive the asymptotic distribution of the smooth principal components, which is unknown for penalized methods as in other non-parametric settings. It is also shown that the degree of regularity of kernel-based principal components is given by that of the kernel function used. In Sections 3 and 4, strong consistency and the asymptotic distribution are derived under mild conditions.

2. Notation and background

Let $\{X(t): t \in \mathscr{I}\}$ be a stochastic process defined in $(\Omega, \mathscr{A}, P)$ with continuous trajectories, zero mean and finite second moment, i.e.

$$E(X(t)) = 0 \qquad E(X^2(t)) < \infty \quad \text{for } t \in \mathscr{I}, \tag{1}$$

where $\mathscr{I} \subset \mathbb{R}$ is a finite interval. Without loss of generality, we may assume that $\mathscr{I} = [0,1]$. We will denote by

$$\gamma(t,s) = E(X(t)X(s))$$

its covariance function, which is just the functional version of the variance–covariance matrix in the classical multivariate analysis. As in the finite-dimensional case, the covariance function has an associated linear operator $\boldsymbol{\Gamma}: L^2(0,1) \to L^2(0,1)$ defined as

$$(\boldsymbol{\Gamma} u)(t) = \int_0^1 \gamma(t,s)u(s)\,\mathrm{d}s \quad \forall u \in L^2(0,1). \tag{2}$$

Throughout all the paper, we will assume that

$$\int_0^1 \int_0^1 \gamma^2(t,s)\,\mathrm{d}t\,\mathrm{d}s < \infty. \tag{3}$$

Cauchy–Schwartz inequality implies that $|\boldsymbol{\Gamma} u|^2 \leqslant \|\gamma\|^2 |u|^2$, where $|u|$ stands for the usual norm in the space $L^2[0,1]$, while $\|\gamma\|$ denotes the norm in the space $L^2([0,1]\times[0,1])$. Therefore, $\boldsymbol{\Gamma}$ is a self-adjoint continuous linear operator. Moreover, (3) implies that $\boldsymbol{\Gamma}$ is a Hilbert–Schmidt operator.

$\mathscr{F}$ will stand for the Hilbert space of such operators with inner product defined by

$$(\boldsymbol{\Gamma}_1,\boldsymbol{\Gamma}_2)_{\mathscr{F}} = \mathrm{trace}(\boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2) = \sum_{j=1}^{\infty}(\boldsymbol{\Gamma}_1 u_j, \boldsymbol{\Gamma}_2 u_j),$$

where $\{u_j: j\geqslant 1\}$ is any orthonormal basis of $L^2(0,1)$ and (u,v) denotes the usual inner product in $L^2(0,1)$.

Choosing a basis $\{\phi_j: j\geqslant 1\}$ of eigenfunctions of $\boldsymbol{\Gamma}$ we have that

$$\|\boldsymbol{\Gamma}\|_{\mathscr{F}}^2 = \sum_{j=1}^{\infty}\lambda_j^2 = \int_0^1\int_0^1 \gamma^2(t,s)\,\mathrm{d}t\,\mathrm{d}s < \infty,$$

where $\{\lambda_j: j\geqslant 1\}$ are the eigenvalues of $\boldsymbol{\Gamma}$. (The last equality is a consequence of Schmidt Theorem.)

As in the classical multivariate case, the linear operator $\boldsymbol{\Gamma}$ can be difficult to interpret and does not always give a fully comprehensive presentation of the structure of the variability in the observed data directly. A principal component analysis provides a way of looking at covariance structure which can be much more informative and can complement the direct examination of the variance operator. So, following Dauxois et al. (1982) we define the population functional principal components as follows.

For any random variable, Y, defined through a linear combination of the process $\{X(s)\}$, i.e.

$$Y = \int_0^1 \alpha(t)X(t)\,\mathrm{d}t = (\alpha, X), \quad \alpha\in L^2(0,1),$$

we have that

$$\mathrm{var}(Y) = E(Y^2) = \int_0^1\int_0^1 \alpha(t)\gamma(t,s)\alpha(s)\,\mathrm{d}s\,\mathrm{d}t = (\alpha, \boldsymbol{\Gamma}\alpha).$$

The first principal component is defined as the random variable $Y_0 = (\alpha_0, X)$ such that

$$\mathrm{var}(Y_0) = \sup_{\{\alpha:|\alpha|=1\}} \mathrm{var}((\alpha,X)) = \sup_{\{\alpha:|\alpha|=1\}} (\alpha, \boldsymbol{\Gamma}\alpha). \tag{4}$$

Therefore, if $\lambda_j\geqslant\lambda_{j+1}$, Riesz Theorem (Riesz and Nagy, 1965, p. 230) entails that the solution of (4) is related to an eigenfunction associated to the largest eigenvalue of the operator $\boldsymbol{\Gamma}$, i.e., $\alpha_0=\phi_1$ and $\mathrm{var}(Y_0)=\lambda_1$.

Moreover, if $\mathscr{A}_k = \{\alpha\in L^2(0,1): |\alpha|=1, (\alpha,\phi_i)=0,\ 1\leqslant i\leqslant k-1\}$ and since

$$\sup_{\alpha\in\mathscr{A}_k}\mathrm{var}((\alpha,X)) = \sup_{\alpha\in\mathscr{A}_k}(\alpha,\boldsymbol{\Gamma}\alpha) = (\phi_k, \boldsymbol{\Gamma}\phi_k) = \lambda_k, \tag{5}$$

the kth populational functional principal component is just (ϕ_k, X).

If all the eigenvalues have multiplicity one the solution is uniquely defined. As in the finite-dimensional case, from (5) we get the following geometrical interpretation. For any fixed integer k, let $\mathscr{H}\subset L^2[0,1]$ be a linear subspace of dimension k and denote by $X^*_{\mathscr{H}}$ the orthogonal projection of X on $\mathscr{H}$. Then, the linear space spanned by $\phi_1,\ldots,\phi_k$, minimizes $E(|X(t)-X^*_{\mathscr{H}}(t)|^2)$.

Non-smooth estimators of the eigenfunctions and eigenvalues of $\boldsymbol{\Gamma}$ were considered by Dauxois et al. (1982), in a natural way through the empirical covariance operator.

More precisely, let $\gamma_n(t,s)$ denote the empirical covariance function, i.e.,

$$\gamma_n(t,s) = \frac{1}{n}\sum_{i=1}^{n} X_i(t)X_i(s)$$

and $\boldsymbol{\Gamma}_n$ the related linear operator. Then, if $\boldsymbol{V}_i$ stands for the linear operator given by

$$(\boldsymbol{V}_i u)(t) = \int_0^1 X_i(t)X_i(s)u(s)\,\mathrm{d}s,$$

we have that, for $1 \leqslant i \leqslant n$,

$$\boldsymbol{\Gamma}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{V}_i \quad \text{and} \quad E(\boldsymbol{V}_i) = \boldsymbol{\Gamma}. \tag{6}$$

Dauxois et al. (1982) defined non-smooth estimators of the population functional principal component ϕ_k as the eigenfunction $\hat{\phi}_k$ related to the kth largest eigenvalue $\hat{\lambda}_k$ of the random operator $\boldsymbol{\Gamma}_n$.

There, they derived strong consistency results for the eigenvalues and their associated eigenmanifolds from the fact that $\|\boldsymbol{\Gamma}_n - \boldsymbol{\Gamma}\|_{\mathscr{F}} \to 0$ almost surely, which follows directly from the strong law of large numbers in the space $\mathscr{F}$.

Using the Central Limit Theorem in Hilbert spaces, they have also shown that $\sqrt{n}(\boldsymbol{\Gamma}_n - \boldsymbol{\Gamma})$ converges in distribution to a zero mean Gaussian random element, $\boldsymbol{U}$, of $\mathscr{F}$ with covariance operator $\boldsymbol{\Sigma}$ and derived from it the asymptotic distribution of the non-smooth estimates of the eigenvalues and of the associated eigenmanifolds of the linear operator $\boldsymbol{\Gamma}$.

Smooth versions of the previous estimates have been defined, through roughness penalties on the sample variance or on the L^2-norm, respectively, by Rice and Silverman (1991) and by Silverman (1996), where consistency results were obtained. See also Ramsay and Dalzell (1991) and Ramsay and Silverman (1997).

3. A kernel-based smooth approach

Our aim is to define smooth estimates of the principal components using a kernel method. This approach is equivalent to smoothing the functional data and then performing the PCA analysis. We begin by defining a smooth version of the estimated covariance operator.

Let $K_h(\cdot) = h^{-1}K(\cdot/h)$ be a kernel nonnegative function with smoothing factor h, such that $\int K(u)\,\mathrm{d}u = 1$ and $\int K^2(u)\,\mathrm{d}u < \infty$.

Given a sample $\{X_1(t),\dots,X_n(t)\}$, $0 \leqslant t \leqslant 1$, of i.i.d. trajectories of the stochastic process $\{X(t),\ 0 \leqslant t \leqslant 1\}$ define smoothed trajectories, via convolution, as

$$X_{ih}(t) = \int K_h(t-s)X_i(s)\,\mathrm{d}s \quad \text{for } 0 \leqslant t \leqslant 1, \tag{7}$$

where we extend $X_i(s)$ as $X_i(0)$ or $X_i(1)$ for $s < 0$ or $s > 1$, respectively.

Define also the empirical covariance function of the smoothed trajectories

$$\gamma_{n,h}(t,s) = \frac{1}{n}\sum_{i=1}^{n} X_{ih}(t)X_{ih}(s).$$

Remark 1. It is worthwhile noting that $\gamma_{n,h}(t,s)$ is a smooth version of the empirical covariance function of the current process, $\gamma_n(t,s)$, since

$$\gamma_{n,h}(t,s) = \int\int \bar{K}_h(t-u, s-v)\gamma_n(u,v)\,\mathrm{d}u\,\mathrm{d}v$$

with $\bar{K}_h(t-u, s-v) = K_h(u)K_h(v)$.

Similarly, we can define the smoothed process

$$X_h(t) = (K_h * X)(t) = \int K_h(t-s)X(s)\,\mathrm{d}s \quad \text{for} \quad 0 \leqslant t \leqslant 1 \tag{8}$$

and

$$\gamma_h(t,s) = E(X_h(t)X_h(s)),$$

which is well defined since $X_h(t) \in L^2(P)$.

Thus, we have that

$$E(\gamma_{n,h}(t,s)) = \gamma_h(t,s) \quad \text{and} \quad \gamma_h(t,s) = (\bar{K}_h * \gamma)(t,s)$$

and so the covariance function of the smoothed process $X_h(t)$ is just a smooth version of the covariance function of the original process.

Define $\boldsymbol{\Gamma}_{nh}$ as the random linear operator associated to γ_{nh}, i.e.

$$(\boldsymbol{\Gamma}_{nh}u)(t) = \int_0^1 \gamma_{nh}(t,s)u(s)\,\mathrm{d}s \quad \text{for} \quad u \in L^2(0,1) \tag{9}$$

and its expected value $\boldsymbol{\Gamma}_h$ defined by

$$(\boldsymbol{\Gamma}_h u)(t) = \int_0^1 \gamma_h(t,s)u(s)\,\mathrm{d}s \quad \text{for} \quad u \in L^2(0,1). \tag{10}$$

Then, $\boldsymbol{\Gamma}_{nh}$ which is a smooth estimate of the covariance operator, can be written as

$$\boldsymbol{\Gamma}_{nh} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{V}_{ih}, \tag{11}$$

where $\boldsymbol{V}_{ih}$ is the linear operator defined through

$$(\boldsymbol{V}_{ih}u)(t) = \int_0^1 X_{ih}(t)X_{ih}(s)u(s)\,\mathrm{d}s.$$

Note that V_{ih} has only one non-null eigenvalue, $\eta_{ih}=|X_{ih}|^2$ with related eigenfunction $X_{ih}/|X_{ih}|^2$ and $E(\boldsymbol{V}_{ih})=\boldsymbol{\Gamma}_h$ for $1 \leqslant i \leqslant n$.

Natural smooth estimates of the eigenfunctions and eigenvalues defining the functional principal components, i.e., the eigenfunctions and eigenvalues of $\boldsymbol{\Gamma}$, are then the eigenfunctions $\hat{\phi}_{jh}$ and the eigenvalues $\hat{\lambda}_{jh}$ of the random operator $\boldsymbol{\Gamma}_{nh}$. Then, using a kernel approach it is equivalent to first smooth the data and then perform a principal component analysis or to consider the covariance operator related to the raw data, then smooth it and then perform the PCA. One can also think in directly smoothing the principal functions but in this approach, the orthogonality between the smooth estimates of the eigenfunctions is lost.

3.1. Smoothness properties

The following lemma shows that smoothness properties are attained. More regularity conditions can be obtained from a more regular behavior of the kernel K.

Lemma 1. *Let $\hat{\phi}_{jh}$ be an eigenfunction associated to a non-null eigenvalue $\hat{\lambda}_{jh}$ of the operator $\boldsymbol{\Gamma}_{nh}$:*

(a) *If K is continuous, then the eigenfunction $\hat{\phi}_{jh}$ can be chosen to be continuous.*

(b) *If K is Lipschitz continuous, then the eigenfunction $\hat{\phi}_{jh}$ can be chosen to be Lipschitz continuous.*

Proof. It is easy to see that there exist eigenfunctions $\hat{\phi}_{jh}$ satisfying

$$\hat{\lambda}_{jh}\hat{\phi}_{jh}(t) = \int_0^1 \gamma_{nh}(t,s)\hat{\phi}_{jh}(s)\,\mathrm{d}s = (\boldsymbol{\Gamma}_{nh}\hat{\phi}_{jh})(t) \quad \text{for all } t \in [0,1]. \tag{12}$$

Effectively, if $\psi_{jh}(t)$ is an eigenfunction of $\boldsymbol{\Gamma}_{nh}$, ψ_{jh} satisfies (12) a.s. Define $\hat{\phi}_{jh}(t)=(\boldsymbol{\Gamma}_{nh}\psi_{jh})(t)$, then, $\hat{\phi}_{jh}$ satisfies (12) for all t.

(a) The continuity of K entails the continuity of γ_{nh} and therefore, $M_1 = \sup_{[0,1]\times[0,1]}|\gamma_{nh}(t,s)| < \infty$. Let $\{t_l, l \geqslant 1\}$ be a sequence converging to t, then $\gamma_{nh}(t_l,s) \to \gamma_{nh}(t,s)$ and $|\gamma_{nh}(t_l,s)\hat{\phi}_{jh}(s)| \leqslant M_1|\hat{\phi}_{jh}(s)|$.

Since $\hat{\phi}_{jh}(s)$ is integrable, applying the Dominated Convergence Theorem in (12) we get that $\hat{\phi}_{jh}(t_l)$ converges to $\hat{\phi}_{jh}(t)$.

(b) Denote by C the Lipschitz constant of K, hence, we have

$$|\gamma_{nh}(t,s) - \gamma_{nh}(t',s)| \leqslant \frac{C}{h^2}|t-t'|M_1 \int_0^1 |K_h(v-s)|\,\mathrm{d}v = A|t-t'| \tag{13}$$

for a suitable constant A. Therefore, γ_{nh} is Lipschitz in each variable.

Finally, using (12) and (13) we obtain

$$|\hat{\phi}_{jh}(t) - \hat{\phi}_{jh}(t')| \leqslant \frac{A}{\hat{\lambda}_{jh}}|t-t'| \int_0^1 |\hat{\phi}_{jh}(s)|\,\mathrm{d}s,$$

which concludes the proof. □

4. Asymptotic results

4.1. Consistency

In order to get the strong consistency of the eigenvalues and their associated eigenmanifolds, we have that from Propositions 2 and 4 of Dauxois et al. (1982), it will be enough to show that

$$\|\boldsymbol{\Gamma}_{nh} - \boldsymbol{\Gamma}\|_{\mathscr{F}} \to 0 \quad \text{a.s.} \tag{14}$$

In the non-smooth case, an analogous result to (14) follows directly from the Strong Law of Large Numbers in the space $\mathscr{F}$. In the smooth case, we will use a Bernstein inequality for Hilbert valued random elements due to Yurinskii (1976), which we include for the sake of completeness.

In a similar way, we will derive strong rates of convergence for the estimates of eigenvalues and their related eigenmanifolds.

Proposition 1 (Yurinskii, 1976). *Let ξ_i be independent random elements taking values on a Hilbert space $\mathscr{F}$. Assume that $E(\xi_i)=0$ and that*

$$E\|\xi_i\|_{\mathscr{F}}^m \leqslant \frac{m!}{2} b_i^2 A^{m-2} \quad \textit{for all } m \geqslant 2.$$

Then, if $B_n^2 = \sum_{i=1}^n b_i^2$, we have that

$$P\left(\left\|\sum_{i=1}^n \xi_i\right\|_{\mathscr{F}} > xB_n\right) \leqslant 2\exp\left(-x^2\left[2+3.24\frac{xA}{B_n}\right]^{-1}\right). \tag{15}$$

Proposition 2. *Assume that K is a non-negative kernel function with $\int K(u)\,\mathrm{d}u = 1$ and that*

$$E(|X_1|^{2m}) \leqslant \frac{m!}{2} b^2 A^{m-2}$$

with $b^2 = E(|X_1|^2)$ and $A>0$, for $m \geqslant 2$. Then, for any sequence $\{\beta_n\}_{n\geqslant 1}$ such that $\beta_n = \mathrm{o}(n/\log n)$, we have

$$\sqrt{\beta_n}\|\boldsymbol{\Gamma}_{nh} - \boldsymbol{\Gamma}_h\|_{\mathscr{F}} \to 0 \quad \textit{completely}. \tag{16}$$

Proof. Since $E(\boldsymbol{\Gamma}_{nh}) = E(\boldsymbol{V}_{ih}) = \boldsymbol{\Gamma}_h$ for all i, we have that

$$\|\boldsymbol{\Gamma}_{nh} - \boldsymbol{\Gamma}_h\|_{\mathscr{F}} = \frac{1}{n}\left\|\sum_{i=1}^n (\boldsymbol{V}_{ih} - E(\boldsymbol{V}_{ih}))\right\|_{\mathscr{F}}.$$

On the other hand, Young's inequality entail that

$$\|\boldsymbol{V}_{ih}\|_{\mathscr{F}} = \eta_{ih} = |X_{ih}|^2 \leqslant |X_i|^2 \left(\int_0^1 |K_h(t)|\, \mathrm{d}t \right)^2 \leqslant |X_i|^2, \tag{17}$$

which implies

$$\begin{aligned} E(\|\boldsymbol{V}_{ih} - E(\boldsymbol{V}_{ih})\|_{\mathscr{F}}^m) &\leqslant 2^{m+1} E(|X_i|^{2m}) \\ &\leqslant 2^m m! b^2 A^{m-2} = \frac{m!}{2} b_1^2 A_1^{m-2}, \end{aligned}$$

where $b_1 = 2\sqrt{2}b$, $A_1 = 2A$ and thus $B_n^2 = 8b^2 n$. Therefore, using (15) with $x = \varepsilon\sqrt{n/\beta_n}$, one gets

$$\begin{aligned} P\left(\sqrt{\beta_n} \|\boldsymbol{\Gamma}_{nh} - \boldsymbol{\Gamma}_h\|_{\mathscr{F}} > \varepsilon \right) &= P\left(\left\| \sum_{i=1}^n (\boldsymbol{V}_{ih} - E(\boldsymbol{V}_{ih})) \right\|_{\mathscr{F}} > xB_n \right) \\ &\leqslant 2\exp\left(-\varepsilon^2 n \beta_n^{-1} \left[2 + 3.24 \frac{\varepsilon A_1}{\sqrt{\beta_n} b_1} \right]^{-1} \right) \end{aligned}$$

which entails (16), since $\beta_n = \mathrm{o}(n/\log n)$. □

Note that no condition on h is needed.

Proposition 3. *Assume that* $E(|X_1|^2) < \infty$.
(a) *If* $h \to 0$, *we have that* $\|\boldsymbol{\Gamma}_h - \boldsymbol{\Gamma}\|_{\mathscr{F}} \to 0$.
(b) *If* $\int |t| K(t)\, \mathrm{d}t < \infty$ *and*

$$|\gamma(t,u) - \gamma(t,t)| \leqslant \boldsymbol{C}|t-u|, \tag{18}$$

then $\beta_n h \to 0$ *implies that* $\sqrt{\beta_n}\|\boldsymbol{\Gamma}_h - \boldsymbol{\Gamma}\|_{\mathscr{F}} \to 0$.
(c) *If* $\int t^2 K(t)\, \mathrm{d}t < \infty, \int tK(t)\, \mathrm{d}t = 0$ *and the covariance function* $\gamma(t,u)$ *is continuously differentiable and*

$$g(t,u_0) = \left| \frac{\partial}{\partial u}\gamma(t,u) \right|_{u=t} - \frac{\partial}{\partial u}\gamma(t,u)|_{u=u_0} | \leqslant \boldsymbol{C}|t-u_0|, \tag{19}$$

then $\beta_n h^2 \to 0$ *entails that* $\sqrt{\beta_n}\|\boldsymbol{\Gamma}_h - \boldsymbol{\Gamma}\|_{\mathscr{F}} \to 0$.

Proof. (a) Since $\boldsymbol{\Gamma}_h = E(\boldsymbol{V}_{1h})$ and $\boldsymbol{\Gamma} = E(\boldsymbol{V}_1)$, we have that

$$\|\boldsymbol{\Gamma}_h - \boldsymbol{\Gamma}\|_{\mathscr{F}} = \|E(\boldsymbol{V}_{1h} - \boldsymbol{V}_1)\|_{\mathscr{F}} \leqslant E(\|\boldsymbol{V}_{1h} - \boldsymbol{V}_1\|_{\mathscr{F}}). \tag{20}$$

On the other hand, since $\|\boldsymbol{V}_{1h} - \boldsymbol{V}_1\|_{\mathscr{F}} \leqslant 2|X_i|^2$, the Dominated Convergence Theorem will entail the desired result if we show that

$$\|\boldsymbol{V}_{1h} - \boldsymbol{V}_1\|_{\mathscr{F}} \to 0 \quad \text{a.s.} \tag{21}$$

For each $\omega \in \Omega$ from the fact that $X_{1h}(t) = (K_h * X_1)(t)$ and $X_1(t) = X_1(t,\omega)$ is a continuous function of t, we obtain that

$$\int_0^1 (X_{1h}(t,\omega) - X_1(t,\omega))^2\, \mathrm{d}t = |X_{1h} - X_1|^2 \to 0 \quad \text{as } h \to 0.$$

Therefore, since $\|\boldsymbol{V}_{1h} - \boldsymbol{V}_1\|_{\mathscr{F}}^2 = |X_{1h}|^4 + |X_1|^4 - 2(X_{1h}, X_1)^2$, the continuity of the inner product with respect to the norm entails that (21) holds for each $\omega \in \Omega$.

(b) Using that $|X_{1h}|\leqslant|X_1|$, it follows easily that

$$\|\boldsymbol{V}_{1h}-\boldsymbol{V}_1\|_{\mathscr{F}}^2=|X_{1h}|^4+|X_1|^4-2(X_{1h},X_1)^2\leqslant 3|X_1|^2|X_{1h}-X_1|^2. \tag{22}$$

Therefore, using (20), (22) and Cauchy–Schwartz inequality, we obtain that

$$\sqrt{\beta_n}\|\boldsymbol{\Gamma}_h-\boldsymbol{\Gamma}\|_{\mathscr{F}}\leqslant\sqrt{3}[E(|X_1|^2)]^{1/2}\sqrt{\beta_n}[E(|X_{1h}-X_1|^2)]^{1/2},$$

which entails that it will be enough to show that

$$\beta_n E(|X_{1h}-X_1|^2)\to 0. \tag{23}$$

Denote by $\tilde{\gamma}_h(t,t)=\int K_h(t-u)\gamma(t,u)\,\mathrm{d}u$. Then, from (18) we get $|\tilde{\gamma}_h(t,t)-\gamma(t,t)|\leqslant \boldsymbol{C}h\int|z|K(z)\,\mathrm{d}z=\boldsymbol{C}_1h$ which entails

$$\int_0^1|\tilde{\gamma}_h(t,t)-\gamma(t,t)|\,\mathrm{d}t\leqslant\boldsymbol{C}_1h. \tag{24}$$

On the other hand, using again (18) we obtain

$$\begin{aligned}\int_0^1|\gamma_h(t,t)-\gamma(t,t)|\,\mathrm{d}t &\leqslant \iint K(u)K(z)|u|h\,\mathrm{d}u\,\mathrm{d}z+\iint K(u)K(z)|z|h\,\mathrm{d}u\,\mathrm{d}z.\\ &\leqslant \boldsymbol{C}_2h.\end{aligned} \tag{25}$$

Therefore, since $E(X_1(t)X_{1h}(t))=\tilde{\gamma}_h(t,t)$, we have

$$E(|X_{1h}-X_1|^2)=\int_0^1[\gamma_h(t,t)-\gamma(t,t)]\,\mathrm{d}t-2\int_0^1[\tilde{\gamma}_h(t,t)-\gamma(t,t)]\,\mathrm{d}t, \tag{26}$$

which together with (24) and (25) entails that $E(|X_{1h}-X_1|^2)\leqslant\boldsymbol{C}_3h$, which concludes the proof since $\beta_nh\to 0$.

(c) As in (b) it will be enough to show (23).

From (19) and using $\int uK(u)\,\mathrm{d}u=0$, similar arguments as those used in (b) led to $E(|X_{1h}-X_1|^2)\leqslant\boldsymbol{C}_3h^2$, which concludes the proof since $\beta_nh^2\to 0$. □

Remark 2. The previous proposition provides convergence rates for the bias term. Note that (b) and (c) show that now, the trade-off is between the regularity of the covariance function and the speed at which the smoothing parameter converges to 0. Propositions 2 and 3(a) entail that no compromise between n and h is needed in order to get just consistency of the smoothed covariance operator $\boldsymbol{\Gamma}_{nh}$.

Order the eigenvalues of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_{nh}$ as decreasing sequences, i.e., $\lambda_j\geqslant\lambda_{j+1}$ and $\hat{\lambda}_{jh}\geqslant\hat{\lambda}_{(j+1),h}$ and denote by $\mathscr{M}_j=\{k\geqslant 1\colon \lambda_k=\lambda_j\}$ and $m_j=\#\mathscr{M}_j$, where $\#\mathscr{M}$ stands for the number of elements of the finite set $\mathscr{M}$. Define $\boldsymbol{P}_j$ and $\hat{\boldsymbol{P}}_{jh}$ as the projection operators associated with λ_j and $\hat{\lambda}_{jh}$, respectively, more precisely,

$$(\boldsymbol{P}_ju)(t)=\lambda_j\sum_{k\in\mathscr{M}_j}\phi_k(t)\,(u,\phi_k)\quad\text{and}\quad(\hat{\boldsymbol{P}}_{jh}u)(t)=\hat{\lambda}_{jh}\sum_{k\in\mathscr{M}_j}\hat{\phi}_{kh}(t)(u,\hat{\phi}_{kh}).$$

The following theorem summarizes our consistency results.

Theorem 1. *Let $K_h(.)=h^{-1}K(./h)$ be a non-negative kernel function with smoothing factor $h\to 0$, such that $\int K(u)\,\mathrm{d}u=1$ and $\int K^2(u)\,\mathrm{d}u<\infty$. Assume that (1) and (3) hold and that for some positive constant A*

$$E(|X_1|^{2m})\leqslant\frac{m!}{2}E(|X_1|^2)\,A^{m-2}\quad\forall m\geqslant 2.$$

Then, we have that, for each $j\geqslant 1$, if λ_j has multiplicity m_j,

(a) *there exists* m_j *sequences* $\{\hat{\lambda}_{kh}, k \in \mathscr{M}_j\}$ *converging to* λ_j, *a.s.*
(b) $\hat{\boldsymbol{P}}_{jh}$ *converges to* $\boldsymbol{P}_j$ in $\mathscr{F}$, *a.s.*
(c) *In particular, when* $m_j = 1$, $\hat{\phi}_{jh}$ *converges to* ϕ_j *in* $L^2[0,1]$, *a.s.*

Remark 3. It is worthwhile noting that there is no trade-off between bias and variance. It is just needed that $h \to 0$. In spite of the non-parametric regression setting, for functional principal components, the non-smooth solution is also consistent.

Theorem 2 (Strong convergence rates)**.** *Let* $\{\beta_n\}_{n\geqslant 1}$ *be a sequence, such that* $\beta_n = \mathrm{o}(n/\log n)$. *Under the assumptions given in Theorem* 1 *and* (b) *or* (c) *of Proposition* 3, *we have that, for each* $j \geqslant 1$, *if* λ_j *has multiplicity* m_j,
(a) *with probability one, there exists* m_j *sequences*, $\{\hat{\lambda}_{kh}, k \in \mathscr{M}_j\}$, *such that* $\sqrt{\beta_n}(\hat{\lambda}_{kh} - \lambda_j)$ *converges to* 0, *for any* $k \in \mathscr{M}_j$.
(b) $\sqrt{\beta_n}\|\hat{\boldsymbol{P}}_{jh} - \boldsymbol{P}_j\|_{\mathscr{F}} \to 0$ *a.s.*
(c) *In particular, when* $m_j = 1$, $\sqrt{\beta_n}|\hat{\phi}_{jh} - \phi_j| \to 0$ (*in* $L^2[0,1]$), *a.s.*

The proof of Theorem 2 follows easily from Proposition 2 and 3(b) or (c) together with the inequalities relating the distance between the corresponding eigenvalues (or projection operators) of two operators and the norm between the operators themselves, used in Proposition 2 (Proposition 3) of Dauxois et al. (1982).

4.2. Asymptotic distribution

As mentioned above, using the Central Limit Theorem in Hilbert spaces, Dauxois et al. (1982) have shown that $\sqrt{n}(\boldsymbol{\Gamma}_n - \boldsymbol{\Gamma})$ converges in distribution to a zero mean Gaussian random element, $\boldsymbol{U}$, of $\mathscr{F}$ with covariance operator $\boldsymbol{\Sigma}$ and derived from it the asymptotic distribution of the non-smooth estimate of the eigenvalues and of the associated manifolds of the linear operator $\boldsymbol{\Gamma}$. Therefore, in order to obtain the same result for the smooth estimate, it will suffice to show that

$$\|\sqrt{n}(\boldsymbol{\Gamma}_{nh} - \boldsymbol{\Gamma}_n)\|_{\mathscr{F}} \to 0 \quad \text{in probability.}$$

Proposition 4. *If* $E(|X_1|^4) < \infty$ *and* $h = h_n \to 0$, *then*,

$$\|\boldsymbol{Z}_n\|_{\mathscr{F}} = \sqrt{n}\|(\boldsymbol{\Gamma}_{nh} - \boldsymbol{\Gamma}_h) - (\boldsymbol{\Gamma}_n - \boldsymbol{\Gamma})\|_{\mathscr{F}} \to 0 \quad \textit{in probability.}$$

Proof. Using Tchébishev's inequality, we obtain that

$$P(\|\boldsymbol{Z}_n\|_{\mathscr{F}} > \varepsilon) \leqslant \frac{1}{\varepsilon^2 n} E\left(\left\|\sum_{i=1}^{n} \xi_{ih}\right\|_{\mathscr{F}}^2\right), \tag{27}$$

where $\xi_{ih} = (\boldsymbol{V}_{ih} - E(\boldsymbol{V}_{ih})) - (\boldsymbol{V}_i - E(\boldsymbol{V}_i))$. Since, $E(\xi_{ih}) = 0$ and $\{\boldsymbol{V}_{ih} - \boldsymbol{V}_i\colon 1 \leqslant i \leqslant n\}$ are independent random operators, we have that

$$E\left(\left\|\sum_{i=1}^{n} \xi_{ih}\right\|_{\mathscr{F}}^2\right) = E\left(\sum_{i,j} (\xi_{ih}, \xi_{jh})_{\mathscr{F}}\right) = nE(\|\xi_{1h}\|_{\mathscr{F}}^2). \tag{28}$$

Thus, (27) and (28) entail that $P(\|\boldsymbol{Z}_n\|_{\mathscr{F}} \geqslant \varepsilon) \leqslant (1/\varepsilon^2)E(\|\xi_{1h}\|_{\mathscr{F}}^2)$. Therefore, the result will follow if we show that

$$E(\|\xi_{1h}\|_{\mathscr{F}}^2) \to 0. \tag{29}$$

Since $\|\xi_{1h}\|_{\mathscr{F}} \leqslant \|V_{1h}-V_1\|_{\mathscr{F}} + E\|V_{1h}-V_1\|_{\mathscr{F}}$, we have that $E(\|\xi_{1h}\|^2_{\mathscr{F}}) \leqslant 2[E(\|V_{1h}-V_1\|^2_{\mathscr{F}}) + (E\|V_{1h}-V_1\|_{\mathscr{F}})^2]$. On the other hand, since $\|V_{1h}-V_1\|_{\mathscr{F}} \leqslant 2|X_i|^2$, the Dominated Convergence Theorem entails (29) and thus, from (21), we get $\|V_{1h}-V_1\|_{\mathscr{F}} \to 0$ a.s. $\square$

The following theorem gives the asymptotic marginal distribution of the smoothed eigenvalues and eigenmanifolds. The proof is a consequence of Proposition 3(b) or (c) (with $\beta_n = n$) which deals with the bias term, Proposition 4 and the results in Section 2.1 of Dauxois et al. (1982).

Theorem 3. *Let $K_h(.) = h^{-1}K(./h)$ be a non-negative kernel function with smoothing factor $h = h_n$, such that $\int K(u)\,\mathrm{d}u = 1$. Assume that* (1) *and* (3) *hold and that $E(|X_1|^4) < \infty$. Let $\boldsymbol{U}$ be a zero mean Gaussian random element of $\mathscr{F}$ with covariance operator $\boldsymbol{\Sigma}$. Then, if assumptions given in* (b) *or* (c) *of Proposition* 3 *hold for $\beta_n = n$, we have that, for each $j \geqslant 1$*

(a) *$\sqrt{n}(\hat{\boldsymbol{P}}_{jh} - \boldsymbol{P}_j)$ converges in distribution in $\mathscr{F}$ to the zero mean Gaussian random element $\boldsymbol{W}_j\boldsymbol{U}\boldsymbol{P}_j + \boldsymbol{P}_j\boldsymbol{U}\boldsymbol{W}_j$, where $\boldsymbol{W}_j$ stands for the linear operator*

$$(\boldsymbol{W}_j u)(t) = \sum_{k \notin \mathscr{M}_j} \frac{1}{\lambda_k - \lambda_j}\phi_k(t)(u, \phi_k).$$

(b) *$(\sqrt{n}(\hat{\lambda}_{kh} - \lambda_k))_{k \in \mathscr{M}_j}$ converges in distribution to the distribution of the decreasing ordered eigenvalues of $\boldsymbol{P}_j\boldsymbol{U}\boldsymbol{P}_j$. Moreover, its joint asymptotic density is given by*

$$f_{m_j}(t_1,\ldots,t_{m_j}) = C\exp\left[-\sum_{l=1}^{m_j}\frac{t_l^2}{4\lambda_j^2}\right]\prod_{k<l}(t_k - t_l),$$

where

$$C^{-1} = 2^{m_j(m_j+3)/4}\prod_{l=1}^{m_j}\left[\Gamma\left(m_j + \frac{1-l}{2}\right)\lambda_j^{m_j(m_j+1)/2}\right],$$

and $\Gamma(p) = \int_0^{+\infty} x^p\,\mathrm{e}^{-x}\,\mathrm{d}x$ denotes the real function Gamma.

In particular, if $m_j = 1$,

(c) *$\sqrt{n}(\hat{\lambda}_{jh} - \lambda_j)$ converges in distribution to a normal distribution with zero mean and variance $2\lambda_j^2$ as in the finite-dimensional case.*

(d) *$\sqrt{n}(\hat{\phi}_{jh} - \phi_j)$ converges in distribution to a zero mean Gaussian random function in $L^2[0,1]$, with covariance function $\rho(t,s)$ given by*

$$\rho(t,s) = \lambda_j\sum_{k \neq j}\frac{\lambda_k}{(\lambda_j - \lambda_k)^2}\phi_k(t)\phi_j(s).$$

Remark 4. Theorem 3 shows that in this problem a root-n rate of convergence is attained, although some smoothing is done. Note that also, when estimating the cumulative distribution function through as smoothed empirical distribution a root-n speed of convergence is obtained. As in that case, non-smooth estimates are consistent unlike the case of the classical density estimation. Nevertheless, the problem of bandwidth selection, which is not addressed in this paper, is crucial. In Pezzulli and Silverman (1993) there is some discussion of whether (using roughness penalty methods) smoothing will actually help, from a second order study of the mean square error.

Acknowledgements

The authors would like to thank an anonymous referee by his/her valuable suggestions which improved the presentation of the paper.

References

Besse, P., Ramsay, J.O., 1986. Principal component analysis of sampled functions. Psychometrika 51, 285–311.

Dauxois, J., Pousse, A., Romain, Y., 1982. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. J. Multivariate Anal. 12, 136–154.

Fraiman, R., Pérez Iribarren, G., 1991. Nonparametric regression estimation in models with weak error's structure. J. Multivariate Anal. 21, 180–196.

Hart, J.D., Wehrly, T.E., 1986. Kernel regression estimation using repeated measurements data. J. Amer. Statist. Assoc. 81, 1080–1088.

Pezzulli, S.D., Silverman, B.W., 1993. Some properties of smoothed principal components analysis for functional data. Comput. Statist. Data Anal. 8, 1–16.

Ramsay, J.O., 1982. When the data are functions. Psychometrika 47, 379–396.

Ramsay, J.O., Dalzell, C.J., 1991. Some tools for functional data analysis (with discussion). J. Roy. Statist. Soc. Ser. B 53, 539–572.

Ramsay, J.O., Silverman, B.W., 1997. Functional Data Analysis. Springer Series in Statistics. Springer, New York.

Rice, J., Silverman, B.W., 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. J. Roy. Statist. Soc. Ser. B 53, 233–243.

Riesz, F., Nagy, B., 1965. Lecons d'analyse functionelle. Gauthiers-Villars, Paris.

Silverman, B.W., 1996. Smoothed functional principal components analysis by choice of norm. Ann. Statist. 24, 1–24.

Yurinskii, V.V., 1976. Exponential inequalities for sums of random vectors. J. Multivariate Anal. 6, 473–499.