Statistics and Probability Letters 94 (2014) 181–191

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Consistency of a numerical approximation to the first principal component projection pursuit estimator

Juan Lucas Bali, Graciela Boente *

Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
CONICET, Argentina

ARTICLE INFO

Article history:
Received 25 January 2014
Received in revised form 17 July 2014
Accepted 17 July 2014
Available online 23 July 2014

MSC:
62F35
62H25

Keywords:
Fisher-consistency
Functional principal components
Outliers
Projection-pursuit
Robust estimation

ABSTRACT

We generalize to functional data, the approach given by Croux and Ruiz-Gazen (1996) to compute robust projection-pursuit principal direction estimators, allowing also for smoothness in the estimators. Consistency of the approximated first principal direction estimator is derived.

1. Introduction

Principal component analysis is a standard multivariate tool used for dimension reduction. The idea is to search for directions that maximize the dispersion of the projected data. The classical estimators normally use the standard deviation as a measure of dispersion leading to estimators very sensitive to atypical observations. To overcome this problem, in the multivariate setting, Li and Chen (1985) introduced a robust projection-pursuit approach to principal component analysis that looks for projections of the data which maximize a robust scale estimator of the projected data. The projection-pursuit idea can be applied almost with no change to the functional data setting (for example, in $L^2(\mathcal{I})$ with $\mathcal{I}$ a bounded interval of $\mathbb{R}$). More generally, in this paper we will consider $\mathcal{H}$ a separable Hilbert space.

One problem of the projection-pursuit estimators is that of actually computing them. Li and Chen (1985) proposed an algorithm and later, Croux and Ruiz-Gazen (1996) contributed their own method of estimation, which is especially interesting because it can be easily adapted to the functional setting. The main problem regarding the projection-pursuit estimators is the difficulty of covering the unit sphere so as to obtain the direction which maximizes the scale of the projected data. Croux and Ruiz-Gazen (1996) proposed instead to use the available data as possible candidates. Empirical results showed that, when considering only few directions in the finite-dimensional setting, the algorithm gives good approximations. However, there is still a lack of theoretical results allowing us to show that the resulting estimators are in fact consistent. In this paper, we face the problem of obtaining consistency results for the computed first principal direction

* Correspondence to: Departamento de Matemáticas, FCEyN, UBA, Ciudad Universitaria, Pabellón 1, Buenos Aires, C1428EHA, Argentina. Tel.: +54 11 45763335; fax: +54 11 45763375.
E-mail addresses: lucasbali@dm.uba.ar (J.L. Bali), gboente@dm.uba.ar, gboente@fibertel.com.ar (G. Boente).

http://dx.doi.org/10.1016/j.spl.2014.07.019
0167-7152/

estimator in the case of an elliptical distribution. Moreover, we generalize the Croux and Ruiz-Gazen algorithm (*the CR algorithm*) to the functional case to include a penalization either in the norm or in the scale, and we prove its consistency under mild conditions. In particular, consistency of the penalizing approach is obtained under the assumption that the trajectories themselves are smooth. If this is not the case, new candidates are needed. For that reason, in Section 5, we discuss a modification of the *CR algorithm*, so as to increase the search space by considering random directions.

The paper is organized as follows. In Section 2, we introduce the notation to be used later, we review some previous results and we define a general algorithm which includes the functional version of the *CR algorithm* as well as related penalized proposals. In Section 3, we derive the consistency of the first principal direction estimator obtained through the proposed procedure when the underlying process is elliptically distributed. The performance of the proposed procedure is evaluated through a numerical study for different sample sizes, in Section 4. Finally, Section 5 contains some concluding remarks. Proofs are relegated to Appendix.

2. The computable projection-pursuit estimator

2.1. Preliminaries

From now on, $\mathcal{H}$ will stand for a separable Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and related norm $\|\alpha\|^2 = \langle\alpha,\alpha\rangle$. Besides, $X \in \mathcal{H}$ will stand for a random element. If X has a finite second moment then its covariance operator $\mathbf{\Gamma}_X$ is linear, self-adjoint and continuous. $\mathbf{\Gamma}_X$ is also Hilbert–Schmidt, so it has a countable number of eigenvalues, all of which are real and nonnegative since $\mathbf{\Gamma}_X$ is positive semi-definite. Denote as $\{\lambda_i\}_{i\in\mathbb{N}}$ the eigenvalues in decreasing order $(\lambda_1 \geq \lambda_2 \geq \cdots)$ and let $\{\phi_i\}_{i\in\mathbb{N}}$ be the associated eigenfunctions, which form an orthonormal basis of $\mathcal{H}$. Using the spectral decomposition theorem, we have that $\mathbf{\Gamma}_X = \sum_{i\geq 1}\lambda_i\phi_i\otimes\phi_i$ where $(u\otimes v)w = \langle v, w\rangle u$. Thus, $\phi_i\otimes\phi_i$ is the orthogonal projection operator over the linear space spanned by ϕ_i, which corresponds to the ith principal direction.

Dauxois et al. (1982) studied the asymptotic properties of the principal direction estimators obtained as the eigenfunctions of the sample covariance operator $\widehat{\mathbf{\Gamma}}_X = (1/n)\sum_{i=1}^n X_i\otimes X_i$. Rice and Silverman (1991) proposed to smooth the principal components by adding a roughness penalty to the sample variance. On the other hand, Silverman (1996) and Ramsay and Silverman (2005) introduced smooth principal components for functional data, also based on roughness penalty methods, but penalizing the norm rather than the sample variance, while Boente and Fraiman (2000) considered a kernel-based approach. More recent work on the estimation of the principal components and the covariance function includes Gervini (2006), Hall and Hosseini-Nasab (2006), Hall et al. (2006) and Yao and Lee (2006). All these methods are sensitive to atypical observations since they are based on classical scale estimators. Locantore et al. (1999) provide resistant estimators of the principal components in a functional setting, although their approach is multivariate in nature. Their proposal was then studied in a fully functional approach by Gervini (2008). Hyndman and Ullah (2007) give an application of a robust projection-pursuit approach, while Bali et al. (2011) introduce robust estimators of the principal directions based on robust projection-pursuit combined with different smoothing methods and establish their strong consistency. For the sake of completeness, we summarize the proposal given in Bali et al. (2011). The basic idea is to view principal components as any direction α_1 such that

$$\operatorname{var}(\langle\alpha_1, X\rangle) = \sup_{\{\alpha:\|\alpha\|=1\}} \operatorname{var}\left(\langle\alpha, X\rangle\right), \tag{1}$$

but replacing the variance with a robust scale functional. For that purpose let σ_{R} be a univariate scale functional, i.e., σ_{R} is location invariant and scale equivariant. Taking into account (1) and given the robust scale functional $\sigma_{\mathrm{R}}(F)$, the robust functional principal component directions are defined as

$$\begin{cases} \phi_{\mathrm{R},1}(P) = \operatorname*{argmax}_{\|\alpha\|=1} \sigma^2(\alpha) \\ \phi_{\mathrm{R},m}(P) = \operatorname*{argmax}_{\|\alpha\|=1,\alpha\in\mathcal{B}_m} \sigma^2(\alpha), \quad 2\leq m, \end{cases} \tag{2}$$

where $\sigma(\alpha) = \sigma_{\mathrm{R}}(P[\alpha])$ and $P[\alpha]$ stands for the distribution of $\langle\alpha, X\rangle$ when $X\sim P$, and $\mathcal{B}_m = \{\alpha\in\mathcal{H} : \langle\alpha,\phi_{\mathrm{R},j}(P)\rangle = 0,\ 1\leq j\leq m-1\}$, for $m\geq 2$.

In this paper, our goal will be the estimation of $\phi_{\mathrm{R},1}(P)$, given a random sample X_i, $1\leq i\leq n$. To define the estimators, consider the empirical version of σ^2, $s_n^2 : \mathcal{H}\to\mathbb{R}$ as $s_n^2(\alpha) = \sigma_{\mathrm{R}}^2(P_n[\alpha])$, where P_n is the empirical measure of P, that is, $P_n(A) = (1/n)\sum_{i=1}^n \delta_A(X_i)$.

We will state all the definitions in a separable Hilbert space $\mathcal{H}$ keeping in mind that the main application will be $\mathcal{H} = L^2(\mathcal{I})$ with $\mathcal{I}\subset\mathbb{R}$ a bounded interval. To define penalized estimators, we assume that $\mathcal{H}_S\subset\mathcal{H}$ stands for a linear subspace of "smooth elements" of $\mathcal{H}$, not necessarily equal to $\mathcal{H}$. Furthermore, $\mathcal{H}_S$ corresponds to the domain of a linear operator $D : \mathcal{H}_S\to\mathcal{H}$ related to the smoothness. This operator is usually identified as the "differentiator", since in $L^2(\mathcal{I})$ it is defined as $D\alpha = \alpha''$ when $\mathcal{H}_S$ is the linear space of twice continuously differentiable functions α such that $\int_{\mathcal{I}}\left[\alpha''(s)\right]^2 ds < \infty$. Using D, we define the symmetric positive semi-definite bilinear form $\lceil\cdot,\cdot\rceil : \mathcal{H}_S\times\mathcal{H}_S\to\mathbb{R}$ as $\lceil\alpha,\beta\rceil = \langle D\alpha, D\beta\rangle$. The "penalization operator" $\Psi : \mathcal{H}_S\to\mathbb{R}$ is assumed to be $\Psi(\alpha) = \lceil\alpha,\alpha\rceil$, while the penalized inner

product equals $\langle \alpha, \beta \rangle_v = \langle \alpha, \beta \rangle + v\lceil \alpha, \beta \rceil$, leading to the penalized norm $\|\alpha\|_v^2 = \|\alpha\|^2 + v\Psi(\alpha)$. Hence, for any $\alpha \in \mathcal{H}_S$, $\Psi(\alpha) = \|D\alpha\|^2 < \infty$.

In Bali et al. (2011), three families of estimators are considered. The first one leads to raw estimators while the two other families lead to smoothed estimators. To be more precise, the raw projection-pursuit estimator of the first principal direction is defined as $\widehat{\phi}_{\mathrm{RAW},1} = \mathrm{argmax}_{\|\alpha\|=1} s_n^2(\alpha)$. The first direction estimator obtained by penalizing the scale is the direction $\widehat{\phi}_{\mathrm{PS},1}$ such that $\widehat{\phi}_{\mathrm{PS},1} = \mathrm{argmax}_{\|\alpha\|=1} \left\{ s_n^2(\alpha) - \rho\Psi(\alpha) \right\}$. Finally, the robust estimator obtained penalizing the norm is defined as $\widehat{\phi}_{\mathrm{PN},1} = \mathrm{argmax}_{\|\alpha\|_v=1} s_n^2(\alpha)$. It is worth noting that the approach penalizing the scale maximizes $\left\{ s_n^2(\alpha) - \rho\Psi(\alpha) \right\} / \|\alpha\|^2$ over $\alpha \neq 0$, while that based on penalizing the norm maximizes $s_n^2(\alpha) / \left\{ \|\alpha\|^2 + v\Psi(\alpha) \right\}$. In both cases roughness will be penalized, but in a somewhat different way. Indeed, using that $\|\alpha\|_v^2 = \|\alpha\|^2 + v\Psi(\alpha)$, we get that

$$\max_{\|\alpha\|_v=1} s_n^2(\alpha) = \max_{\|\alpha\|=1} \frac{s_n^2(\alpha)}{1 + v\Psi(\alpha)} = \max_{\|\alpha\|=1} \left\{ 1 - \frac{v\Psi(\alpha)}{1 + v\Psi(\alpha)} \right\} s_n^2(\alpha).$$

Hence, recalling that $\|\widehat{\phi}_{\mathrm{PN},1}\|_v = 1$, we have that

$$\frac{\widehat{\phi}_{\mathrm{PN},1}}{\|\widehat{\phi}_{\mathrm{PN},1}\|} = \mathrm{argmax}_{\|\alpha\|=1} \frac{s_n^2(\alpha)}{1 + v\Psi(\alpha)} = \mathrm{argmax}_{\|\alpha\|=1} \left\{ 1 - \frac{v\Psi(\alpha)}{1 + v\Psi(\alpha)} \right\} s_n^2(\alpha). \tag{3}$$

Thus, the approach based on penalizing the norm can be viewed as a penalization in the scale where the penalization takes into account the dispersion in the direction.

Under general conditions, the consistency of all these proposals was studied in Bali et al. (2011) where also the subsequent principal direction estimators are defined imposing an orthogonality condition with respect to the previous ones. The inner product used to impose the orthogonality restriction is the usual one for the raw estimators and for those penalizing the scale, while it corresponds to the penalized inner product $\langle \cdot, \cdot \rangle_v$, when considering the estimators obtained penalizing the norm.

2.2. The proposed estimators of the principal directions

It is clear that the computation of $\widehat{\phi}_{\mathrm{RAW},1}$, $\widehat{\phi}_{\mathrm{PS},1}$ or $\widehat{\phi}_{\mathrm{PN},1}$ involves a non-trivial maximization problem, which has been considered as a major disadvantage of the projection-pursuit approach. In particular, the computation of $\widehat{\phi}_{\mathrm{RAW},1}$ and $\widehat{\phi}_{\mathrm{PS},1}$ is harder since the unitary ball $\mathcal{S}_1 = \{\alpha : \|\alpha\| = 1\}$ in an infinite dimensional space is not compact. On the other hand, in many situations, for instance, when considering $\mathcal{H} = L^2([0, 1])$ and $D\alpha = \alpha''$, the set $\{\alpha : \|\alpha\|_v = 1\}$ is a compact subset of $\mathcal{H}$. However, due to the high dimension it is rather sparse so, the computation of the maximum is still a challenging problem. To simplify the presentation, note that the three estimators defined in Section 2.1 can be written using a unified approach as

$$\widehat{\phi}_1 = \mathrm{argmax}_{\|\alpha\|_v=1} \left\{ s_n^2(\alpha) - \rho\Psi(\alpha) \right\},$$

where the products $v\Psi(\alpha)$ or $\rho\Psi(\alpha)$ are defined as 0 when $v = 0$ or $\rho = 0$ respectively, even when $\alpha \notin \mathcal{H}_S$ for which case $\Psi(\alpha) = \infty$. When $v = \rho = 0$, $\widehat{\phi}_1 = \widehat{\phi}_{\mathrm{RAW},1}$, while $\widehat{\phi}_1 = \widehat{\phi}_{\mathrm{PS},1}$ and $\widehat{\phi}_1 = \widehat{\phi}_{\mathrm{PN},1}$ when $v = 0$ and $\rho = 0$, respectively. Alternatively, one may use (3) to define $\widehat{\phi}_1$ as

$$\widehat{\phi}_1 = \mathrm{argmax}_{\|\alpha\|=1} \frac{s_n^2(\alpha) - \rho\Psi(\alpha)}{1 + v\Psi(\alpha)}, \tag{4}$$

in which case, $\widehat{\phi}_{\mathrm{PN},1} = \widehat{\phi}_1 / \|\widehat{\phi}_1\|_v$ when $\rho = 0$ and $v > 0$.

To approximate the estimators $\widehat{\phi}_{\mathrm{RAW},1}$, $\widehat{\phi}_{\mathrm{PS},1}$ and $\widehat{\phi}_{\mathrm{PN},1}$ or more generally, $\widehat{\phi}_1$, we propose the following algorithm which is a generalization of the Croux and Ruiz-Gazen (1996) algorithm. In a first step, we compute a robust location estimator $\widehat{\mu}$, such as the functional spatial median defined in Gervini (2008) and we center the observations $\widetilde{X}_i = X_i - \widehat{\mu}$.

The general projection-pursuit algorithm can be described as follows.

(i) Normalize the observations $\alpha_i = \widetilde{X}_i / \|\widetilde{X}_i\|$ and consider the set of possible directions $\mathcal{A}_n = \{\alpha_i = \widetilde{X}_i / \|\widetilde{X}_i\|, 1 \leq i \leq n, \} \subset \mathcal{S}_1$.

(ii) Estimate $\phi_{\mathrm{R},1}(P)$ using the approximation

$$\widehat{\widehat{\phi}}_1 = \mathrm{argmax}_{\alpha \in \mathcal{A}_n} \frac{s_n^2(\alpha) - \rho\Psi(\alpha)}{1 + v\Psi(\alpha)} = \mathrm{argmax}_{1 \leq i \leq n} \frac{s_n^2(\alpha_i) - \rho\Psi(\alpha_i)}{1 + v\Psi(\alpha_i)}.$$

Then, the *raw approximation*, denoted by $\widehat{\widehat{\phi}}_{\mathrm{RAW},1}$ corresponds to $v = \rho = 0$, while the scale-penalized projection-pursuit algorithm leads to the estimator $\widehat{\widehat{\phi}}_{\mathrm{PS},1}$ taking $v = 0$ and $\rho > 0$. Finally, when $\rho = 0$ and $v > 0$, to obtain an approximation to the true estimator $\widehat{\phi}_{\mathrm{PN},1}$, we proceed as above and define $\widehat{\widehat{\phi}}_{\mathrm{PN},1} = \widehat{\widehat{\phi}}_1 / \|\widehat{\widehat{\phi}}_1\|_v$.

It is worth noting that, when $v > 0$ or $\rho > 0$ the general algorithm assumes that $\mathbb{P}(\widetilde{X}_i \in \mathcal{H}_S) = 1$; otherwise $\Psi(\widetilde{X}_i) = \infty$ and the above definition is meaningless.

2.3. Elliptical families

In the finite-dimensional setting, elliptical families play an important role when deriving Fisher-consistency and infinitesimal robustness of the principal component robust estimators. This notion has been extended to the functional setting by Bali and Boente (2009) and used in Bali et al. (2011) to obtain Fisher-consistency of the projection-pursuit estimators. For the sake of completeness, we recall here their definition.

Let $\mathbf{Z} \in \mathbb{R}^d$ be a random vector. We say that $\mathbf{Z}$ has an elliptical distribution, and we denote it as $\mathbf{Z} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, if there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^d$, a positive semi-definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ and a function $\phi : \mathbb{R}_+ \to \mathbb{R}$ such that the characteristic function of $\mathbf{Z} - \boldsymbol{\mu}$ is given by $\varphi_{\mathbf{Z}-\boldsymbol{\mu}}(\mathbf{t}) = \phi(\mathbf{t}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{t})$, for all $\mathbf{t} \in \mathbb{R}^d$. In particular, if $\boldsymbol{\mu} = \mathbf{0}_d$ and $\boldsymbol{\Sigma} = \mathbf{I}_d$, $\mathbf{Z}$ has a spherical distribution, that is, a distribution which is invariant by orthogonal transformations. To extend the definition to the functional setting, let X be a random element in a separable Hilbert space $\mathcal{H}$, $\mu \in \mathcal{H}$ and let $\boldsymbol{\Gamma} : \mathcal{H} \to \mathcal{H}$ be a self-adjoint, positive semi-definite and compact operator. The random element X has an elliptical distribution with parameters $(\mu, \boldsymbol{\Gamma})$, denoted as $X \sim \mathcal{E}(\mu, \boldsymbol{\Gamma}, \phi)$, if for any linear and bounded operator $A : \mathcal{H} \to \mathbb{R}^d$, AX has a multivariate elliptical distribution with parameters $A\mu$ and $A\boldsymbol{\Gamma}A^*$, i.e., $AX \sim \mathcal{E}_d(A\mu, A\boldsymbol{\Gamma}A^*, \phi)$, where $A^* : \mathbb{R}^p \to \mathcal{H}$ stands for the adjoint operator of A. For the sake of simplicity, we will omit the symbol ϕ and denote $X \sim \mathcal{E}(\mu, \boldsymbol{\Gamma})$. As in the finite-dimensional setting, if the covariance operator $\boldsymbol{\Gamma}_X$ of X exists, then $\boldsymbol{\Gamma}_X = a\,\boldsymbol{\Gamma}$, for some $a \in \mathbb{R}$. Elliptical distributions in $\mathcal{H}$ include, among others, the Gaussian distributions and scale mixtures of Gaussian. Both in the univariate and multivariate settings, a scale mixture of a Gaussian distributions has a longer tail than a Gaussian distribution. For this reason, they are often viewed as attractive alternatives to the Gaussian model, in particular, when dealing with robust procedures. In an infinite-dimensional setting, the rank of the scatter operator $\boldsymbol{\Gamma}$ can either be finite or infinite. When the rank of $\boldsymbol{\Gamma}$ equals q, $X \sim \mathcal{E}(\mu, \boldsymbol{\Gamma})$ admits a finite Karhunen–Loève expansion, $X = \mu + \sum_{j=1}^{q} \lambda_j^{1/2} Z_j \phi_j$ with $\mathbf{Z} = (Z_1, \ldots, Z_q)^{\mathrm{T}}$ having a spherical distribution and where $\lambda_1 \geq \cdots \geq \lambda_q > 0$ and $\{\phi_j, j = 1, \ldots q\}$ are respectively the non-null eigenvalues and the associated orthonormal set of eigenfunctions of $\boldsymbol{\Gamma}$. On the other hand, when the rank of $\boldsymbol{\Gamma}$ is not finite, Proposition 2.1 in Boente et al. (2014) shows that if $X \sim \mathcal{E}(\mu, \boldsymbol{\Gamma})$ then X is a scale mixture of Gaussian.

On the other hand, as mentioned above, elliptical families are helpful in robustness since they allow us to obtain Fisher-consistency. In particular, Bali et al. (2011) showed that, when $X \sim \mathcal{E}(\mu, \boldsymbol{\Gamma})$ where $\boldsymbol{\Gamma} = \sum_{j\geq 1} \lambda_j \phi_j \otimes \phi_j$ is a compact operator with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$ and orthonormal related eigenfunctions ϕ_j, $\sigma^2(\alpha) = c\langle \alpha, \boldsymbol{\Gamma}\alpha\rangle$ for some constant $c > 0$. Hence, $\phi_{\mathrm{R},m}(P) = \phi_m$, that is, we are estimating the target directions.

3. Consistency of the first principal direction approximated estimators

Throughout this section, we will assume, for simplicity, that μ is known. Clearly, when μ is known, one may assume, without lack of generality, that $\mu = 0$, so that $\widetilde{X}_i$ defined in Section 2.2 equals $\widetilde{X}_i = X_i$.

To derive consistency of the approximations, we will consider assumption **A1** below, which was also considered in Bali et al. (2011). Recall that $\sigma(\alpha) = \sigma_{\mathrm{R}}(P[\alpha])$, $s_n(\alpha) = \sigma_{\mathrm{R}}(P_n[\alpha])$ and denote $\lambda_{\mathrm{R},m}(P) = \sigma^2(\phi_{\mathrm{R},m})$.

A1. $\sup_{\|\alpha\|=1} \left|s_n^2(\alpha) - \sigma^2(\alpha)\right| \xrightarrow{a.s.} 0$.

Conditions under which **A1** holds are discussed in Bali et al. (2011). In particular, if σ_{R} is a continuous functional then **A1** holds. Moreover, we also have that σ is a weakly continuous function, i.e., continuous with respect to the weak topology in $\mathcal{H}$.

It is worth noting that $\lambda_{\mathrm{R},1}(P) > \lambda_{\mathrm{R},2}(P)$ if and only if $\phi_{\mathrm{R},1}(P)$ is unique up to a sign change. Moreover, as mentioned above, if $X \sim \mathcal{E}(\mu, \boldsymbol{\Gamma})$ and $\lambda_1 \geq \lambda_2 \geq \ldots$ stand for the ordered eigenvalues of the compact operator $\boldsymbol{\Gamma}$, then $\sigma^2(\alpha) = c\langle \alpha, \boldsymbol{\Gamma}\alpha\rangle$ for some constant $c > 0$ which entails that $\lambda_{\mathrm{R},j}(P) = c\lambda_j$. Without loss of generality, we assume that $c = 1$, so that $\lambda_{\mathrm{R},j}(P) = \lambda_j$ which means that the functional related to the eigenvalues is also Fisher-consistent.

Theorem 3.1 provides the consistency of the "raw", the norm-penalization and the scale-penalization projection-pursuit estimators computed through the approximation algorithm described in Section 2.2.

Theorem 3.1. *Let $\mathcal{H}$ be a separable Hilbert space and $X_i \in \mathcal{H}$, $1 \leq i \leq n$ a random sample with common distribution P. Assume that $P \sim \mathcal{E}(0, \boldsymbol{\Gamma})$ where $\boldsymbol{\Gamma}$ is a compact operator with eigenvalues λ_ℓ ordered in decreasing order, $\lambda_1 > \lambda_2 \geq \ldots$, and related eigenfunctions ϕ_ℓ. Furthermore, assume that $\mathbb{P}(X_1 = 0) = 0$ and that **A1** holds. Let $\{\nu_n\}_{n\in\mathbb{N}}$ and $\{\rho_n\}_{n\in\mathbb{N}}$ be sequences of non-negative real numbers such that $\nu = \nu_n \to 0$ and $\rho = \rho_n \to 0$. Then, if $\mathbb{P}(X_1 \in \mathcal{H}_S) = 1$, where $\mathcal{H}_S = \mathcal{H}$ whenever $\nu = \rho = 0$, we have that*

(a) $\langle \widehat{\widehat{\phi}}_1, \phi_1\rangle^2 \xrightarrow{a.s.} 1$, *that is,* $\widehat{\widehat{\phi}}_1 \xrightarrow{a.s.} \phi_1$ *except for a sign change.*

(b) $\nu\Psi(\widehat{\widehat{\phi}}_1) \to 0$ *and* $\rho\Psi(\widehat{\widehat{\phi}}_1) \to 0$.

(c) $\widehat{\widehat{\lambda}}_1 = s_n^2(\widehat{\widehat{\phi}}_1) \xrightarrow{a.s.} \lambda_1$.

In particular, Theorem 3.1 implies that the approximate estimators $\widehat{\widehat{\phi}}_{\mathrm{RAW},1}$, $\widehat{\widehat{\phi}}_{\mathrm{PN},1}$ and $\widehat{\widehat{\phi}}_{\mathrm{PS},1}$ provide consistent estimators of ϕ_1 for elliptical distributions. Furthermore, it allows us to derive the consistency of the first principal component direction estimator computed using the algorithm introduced in Croux and Ruiz-Gazen (1996, 2005) as given in Corollary 3.1.

Corollary 3.1. *Let $X_i \in \mathbb{R}^p$, $1 \le i \le n$ be a random sample with common distribution P such that $\mathbb{P}(X_1 = 0) = 0$. Assume that $P \sim \mathcal{E}_p(0, \mathbf{\Sigma})$ where $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ is a symmetric positive definite matrix with eigenvalues λ_ℓ ordered in decreasing order and such that $\lambda_1 > \lambda_2 \ge \cdots \ge \lambda_p$. Denote as $\boldsymbol{\beta}_1$ the eigenvector related to λ_1. Then, if* **A1** *holds, we have that the algorithm described in Croux and Ruiz-Gazen (1996) provides consistent estimators of the first principal direction and its size. To be more precise, if $\widehat{\widehat{\boldsymbol{\beta}}}_1 = \operatorname{argmin}_{1\le i\le n} s_n^2(X_i/\|X_i\|)$ and $\widehat{\widehat{\lambda}}_1 = s_n^2(\widehat{\widehat{\boldsymbol{\beta}}}_1)$, then $\widehat{\widehat{\boldsymbol{\beta}}}_1 \xrightarrow{a.s.} \boldsymbol{\beta}_1$ up to a sign change and $\widehat{\widehat{\lambda}}_1 \xrightarrow{a.s.} \lambda_1$.*

4. Monte Carlo study

In this section, we report the results of a numerical study conducted to assess the performance of the algorithm in practice. We considered as sample sizes $n = 50$, 100, 250, 500, 1000 and 5000. For each replication, we generate i.i.d. observations $X_1, \ldots, X_n$, $X_i \in L^2(0, 1)$ such that $X_i \sim P$ where P is defined below. We choose $D\alpha = \alpha''$ and $\mathcal{H}_S$ the linear space of twice continuously differentiable functions α such that $\int_{\mathcal{I}} \left[\alpha''(s)\right]^2 ds < \infty$. It is clear that, if the trajectories are not smooth the approximations $\widehat{\widehat{\phi}}_{\mathrm{PS},1}$ and $\widehat{\widehat{\phi}}_{\mathrm{PN},1}$ of the penalized estimators $\widehat{\phi}_{\mathrm{PS},1}$ and $\widehat{\phi}_{\mathrm{PN},1}$ are not properly defined and clearly they will not be smooth, so that Theorem 3.1 does not guarantee consistency. However, in this numerical study, we want to cover situations with smooth and non-smooth trajectories to see the impact on the numerical approximations. For that purpose, we consider Gaussian processes with two different covariance kernels and we also include a contamination that produces irregular trajectories by introducing a peak contamination. In all cases, we performed 100 replications. To implement the estimators the observations are computed over a grid of M points, where $M = 50$, 100 and 250.

4.1. The models

In all tables, the distribution P of the observations will be labeled C_0 and C_1 according to the following situations.

- C_0: In this case $X_i(t)$ is a Gaussian process in $L^2(0, 1)$. We have considered two possible covariance kernels
  - Under **Model 1**: $\gamma_X(s, t) = (1/2)(1/2)^{0.9|s-t|}$, which corresponds to the Ornstein–Uhlenbeck process.
  - Under **Model 2**: $\gamma_X(s, t) = (1/2)(1/2)^{0.9(s-t)^2}$ which generates smooth trajectories.
- C_1: In this situation, we have contaminated the data using a peak contamination considered by Sawant et al. (2012). The contaminated observations $X_i^{(c)}$ are defined as $X_i^{(c)}(s) = X_i(s) + V_i D_i M \mathbb{I}_{\{T_i<s<T_i+\ell\}}$, where $V_i \sim Bi(1, p)$, D_i is such that $\mathbb{P}(D_i = 1) = \mathbb{P}(D_i = -1) = 1/2$, $T_i \sim \mathcal{U}(0, 1-\ell)$, $\ell < 1/2$ and V_i, X_i, D_i and T_i are independent. We choose $\ell = 1/15$, $M = 30$ and $p = 0.2$.

When considering the standard deviation, the first target principal direction corresponds to the eigenfunction of the covariance operator related to the largest eigenvalue. Denote it as ϕ_1 for the uncontaminated data and ϕ_{1,C_1} for the contaminated ones. We have numerically computed the target directions when the scale is the standard deviation and we obtained that

- Under **Model 1**: $|\cos(\phi_{1,C_1}, \phi_1)| \simeq 0.9871$, so that $\|\phi_{1,C_1} - \phi_1\|^2 = 0.0258$.
- Under **Model 2**: $|\cos(\phi_{1,C_1}, \phi_1)| \simeq 0.9883$, so that $\|\phi_{1,C_1} - \phi_1\|^2 = 0.0235$.

Hence, the contamination does not produce a huge change on the first principal direction when considering the classical approach but rather generates spurious data which loose the smoothness present in Model 2. The purpose of this contamination is to see how it affects the computation of the approximated estimators.

Besides considering the classical estimators based on the standard deviation, we also compute the approximations for robust projection-pursuit estimators based on an M-scale estimator. It is worth noting that, under C_0, the functional $\phi_{\mathrm{R},1}(P)$ defined in (2) also equals ϕ_1 when considering the robust scale estimator, since the process is Gaussian. For the M-scale estimator, we use the biweight Tukey function $\chi_c(y) = \min\left(3\,(y/c)^2 - 3\,(y/c)^4 + (y/c)^6\,,\, 1\right)$, with tuning constant $c = 1.56$ and breakdown point 1/2, which ensures Fisher-consistency of the principal values at Gaussian processes. To compute the M-scale, we choose the MAD as the initial scale estimator. It is clear that, when considering the classical estimators, i.e., those based on the sample variance, the raw estimators of the principal directions can be obtained as the eigenfunctions of the sample covariance operator. However, in order to get the classical penalized estimators or the robust projection-pursuit estimators, we need to compute an approximation over the unit ball. For that reason, to make fair comparisons, we have also computed the raw estimators using the algorithm described in Section 2.2. For the classical procedures, i.e., those based on the standard deviation, we use the sample mean as the centering point $\widehat{\mu}$ for the trajectories to define the centered observations $\widetilde{X}_i = X_i - \widehat{\mu}$. For the robust procedures, i.e., those based on the M-SCALE, we center the data by using the sample spatial median, defined as the value

$$\widehat{\mu} = \operatorname*{argmin}_{\theta\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \|X_i - \theta\|.$$

In all tables, the classical and robust estimators are denoted as SD and M-SCALE, respectively.

Table 1
Mean values of $\|\widehat{\phi}_1 - \phi_1\|^2$ where $\widehat{\phi}_1$ is the first eigenfunction of the sample covariance operator.

n	$M = 50$	$M = 100$	$M = 250$	$M = 50$	$M = 100$	$M = 250$
	$\gamma_X(s,t) = (1/2)(1/2)^{0.9\|s-t\|}$			$\gamma_X(s,t) = (1/2)(1/2)^{0.9(s-t)^2}$		
	C_0					
50	0.0054	0.0061	0.0056	0.0027	0.0031	0.0027
100	0.0026	0.0026	0.0028	0.0013	0.0012	0.0014
250	0.0012	0.0011	0.0011	0.0007	0.0006	0.0005
500	0.0005	0.0005	0.0004	0.0003	0.0003	0.0002
1000	0.0003	0.0003	0.0002	0.0002	0.0001	0.0001
5000	0.00004	0.00004	0.00005	0.00002	0.00002	0.00002
	C_1					
50	1.1683	1.1879	1.1315	1.1501	1.1524	1.0982
100	0.9697	0.9757	0.9450	0.9303	0.9266	0.8983
250	0.6006	0.6049	0.6126	0.5478	0.5456	0.5563
500	0.3763	0.3417	0.3790	0.3301	0.2985	0.3313
1000	0.2194	0.2102	0.2184	0.1878	0.1778	0.1856
5000	0.0660	0.0672	0.0648	0.0576	0.0580	0.0560

We report the results for the raw estimators $\widehat{\widehat{\phi}}_{\mathrm{RAW},1}$ and also, for the estimators obtained penalizing the scale $\widehat{\widehat{\phi}}_{\mathrm{PS},1}$ and the norm $\widehat{\widehat{\phi}}_{\mathrm{PN},1}$. For the latter ones, the penalizing parameter ρ was taken fixed and equal to $\rho = 10^{-5}$ to avoid the high computation time required by a robust cross-validation procedure. To compare the performance of the estimators, we computed the square distance between the true and the estimated direction (normalized to have L^2 norm 1), i.e., $\widehat{\Delta} = \|\widehat{\widehat{\phi}}_1/\|\widehat{\widehat{\phi}}_1\| - \phi_1\|^2$, where $\widehat{\widehat{\phi}}_1 = \widehat{\widehat{\phi}}_{\mathrm{RAW},1}$, $\widehat{\widehat{\phi}}_{\mathrm{PS},1}$ or $\widehat{\widehat{\phi}}_{\mathrm{PN},1}$. It is worth noting that all the estimators except those penalizing the norm are such that $\|\widehat{\widehat{\phi}}_1\| = 1$.

4.2. Performance of the estimators

Tables 2 and 3 report the mean value over replications of $\widehat{\Delta}$ under C_0 and C_1, for Models 1 and 2, respectively. When considering the classical raw estimators and in order to compare the results obtained using the algorithm described in Section 2.2 with the exact procedure, Table 1 reports the results obtained using the eigenfunctions of the sample covariance operator, $\widehat{\mathbf{\Gamma}}$, over grids of M points. When $n = 5000$, we report 5 digits after the comma when needed to avoid reporting only 0 and help in the comparisons.

The results show that the convergence is faster if the trajectories are smooth, even when the estimators are computed as the first eigenfunction of the sample covariance estimator. In particular, under C_1, when considering the raw estimators based on the standard deviation, poor results are obtained for sample sizes smaller than 500, which may be confused with a lack of robustness. This bad performance is not due to a lack of efficiency in the algorithm to approximate the real maximum over the unit ball since when computing the eigenfunctions of the sample covariance operator similar results are obtained (see Table 1). In fact, when the sample size equals 1000 and even for 5000, the algorithm described in this paper gives more accurate results than computing the first eigenfunction. Indeed, under C_1, when the raw estimators are computed using the approximation given by the introduced algorithm, that is, when considering $\widehat{\widehat{\phi}}_{\mathrm{RAW},1}$, the mean values of $\widehat{\Delta}$ are closer to the true values reported above (0.0258 and 0.0235, under Models 1 and 2, respectively) than when computing the first eigenfunction of $\widehat{\mathbf{\Gamma}}$. This behavior may be explained by the fact that the peak contamination considered produces an unsmooth sample covariance operator which makes more difficult to estimate the first direction. In general, the size of the grid M does not lead to an improvement over the obtained results.

With respect to the behavior of $\widehat{\widehat{\phi}}_{\mathrm{RAW},1}$ and the penalized estimators, under C_0, the latter ones behaves as the first one under Model 2, since the trajectories are very smooth and give a good set of candidates $\mathcal{A}_n = \{\alpha_i = X_i/\|X_i\|,\ 1 \le i \le n\} \subset \mathcal{S}_1$ to approximate the true first principal direction which is smooth under both models. On the other hand, under Model 1, penalizing gives worse results than the raw estimators under C_0 due to the lack of smoothness of the trajectories. Besides, as expected, under C_0 and Model 1, computing the first eigenfunction of $\widehat{\mathbf{\Gamma}}$ produces smaller mean square errors than those obtained with $\widehat{\widehat{\phi}}_{\mathrm{RAW},1}$. Again, this can be explained by the fact that the trajectories are not smooth while the true principal component is smooth, so that, in this case, ϕ_1 is better estimated using the sample covariance operator.

As mentioned above, the estimators computed using the algorithm described in Section 2.2 are not properly defined if the trajectories are not smooth when considering penalized estimators, which explains the bad performance obtained in this situation under Model 1 and the much better one observed under Model 2. A possible method to overcome this problem is discussed in Section 5.

As noted in Silverman (1996) and pointed out in Bali et al. (2011) some degree of smoothing in the procedure based on penalizing the norm will give a better estimation of the principal directions in the L^2 sense under mild conditions when

Table 2

Mean values of $\|\widehat{\phi}_1/\|\widehat{\phi}_1\| - \phi_1\|^2$ when $\gamma_X(s,t) = (1/2)(1/2)^{0.9|s-t|}$ and $\widehat{\phi}_1 = \widehat{\widehat{\phi}}_{\mathrm{RAW},1}$, $\widehat{\widehat{\phi}}_{\mathrm{PS},1}$ and $\widehat{\widehat{\phi}}_{\mathrm{PN},1}$.

n	Scale	$M=50$	$M=100$	$M=250$	$M=50$	$M=100$	$M=250$	$M=50$	$M=100$	$M=250$
		$\widehat{\widehat{\phi}}_{\mathrm{RAW},1}$			$\widehat{\widehat{\phi}}_{\mathrm{PS},1}\,(\rho=10^{-5})$			$\widehat{\widehat{\phi}}_{\mathrm{PN},1}\,(\rho=10^{-5})$		
		C_0								
50	SD	0.0226	0.0234	0.0219	0.0386	0.0445	0.0411	0.0344	0.0375	0.0354
	M-SCALE	0.0342	0.0333	0.0326	0.0410	0.0421	0.0418	0.0375	0.0391	0.0377
100	SD	0.0147	0.0149	0.0155	0.0290	0.0349	0.0369	0.0273	0.0333	0.0348
	M-SCALE	0.0183	0.0197	0.0212	0.0294	0.0351	0.0374	0.0269	0.0333	0.0339
250	SD	0.0109	0.0115	0.0102	0.0269	0.0336	0.0253	0.0222	0.0302	0.0244
	M-SCALE	0.0122	0.0144	0.0127	0.0265	0.0347	0.0266	0.0219	0.0290	0.0245
500	SD	0.0084	0.0084	0.0079	0.0260	0.0274	0.0247	0.0195	0.0251	0.0233
	M-SCALE	0.0090	0.0100	0.0088	0.0245	0.0281	0.0247	0.0200	0.0242	0.0231
1000	SD	0.0068	0.0067	0.0064	0.0210	0.0231	0.0208	0.0185	0.0210	0.0190
	M-SCALE	0.0075	0.0076	0.0071	0.0213	0.0238	0.0204	0.0185	0.0210	0.0184
5000	SD	0.0043	0.0044	0.0043	0.0186	0.0170	0.0166	0.0158	0.0158	0.0155
	M-SCALE	0.0044	0.0046	0.0044	0.0185	0.0170	0.0166	0.0165	0.0158	0.0157
		C_1								
50	SD	1.4037	1.4231	1.3850	0.1252	0.1287	0.1267	0.1318	0.1370	0.1377
	M-SCALE	0.0447	0.0375	0.0324	0.0494	0.0444	0.0462	0.0421	0.0400	0.0374
100	SD	1.3779	1.3308	1.3617	0.0790	0.0724	0.0704	0.0822	0.0784	0.0741
	M-SCALE	0.0248	0.0233	0.0226	0.0338	0.0386	0.0355	0.0291	0.0335	0.0313
250	SD	0.8164	0.8397	0.8195	0.0423	0.0433	0.0372	0.0433	0.0392	0.0372
	M-SCALE	0.0158	0.0141	0.0148	0.0257	0.0297	0.0259	0.0222	0.0248	0.0239
500	SD	0.3027	0.1935	0.2244	0.0327	0.0350	0.0325	0.0336	0.0327	0.0309
	M-SCALE	0.0113	0.0113	0.0096	0.0212	0.0311	0.0258	0.0195	0.0240	0.0234
1000	SD	0.0901	0.0831	0.0791	0.0250	0.0277	0.0246	0.0254	0.0254	0.0250
	M-SCALE	0.0086	0.0087	0.0082	0.0213	0.0253	0.0211	0.0185	0.0230	0.0195
5000	SD	0.0491	0.0432	0.0459	0.0178	0.0181	0.0171	0.0172	0.0178	0.0170
	M-SCALE	0.0052	0.0053	0.0049	0.0155	0.0177	0.0164	0.0134	0.0162	0.0157

Table 3

Mean values of $\|\widehat{\phi}_1/\|\widehat{\phi}_1\| - \phi_1\|^2$ when $\gamma_X(s,t) = (1/2)(1/2)^{0.9(s-t)^2}$ and $\widehat{\phi}_1 = \widehat{\widehat{\phi}}_{\mathrm{RAW},1}$, $\widehat{\widehat{\phi}}_{\mathrm{PS},1}$ and $\widehat{\widehat{\phi}}_{\mathrm{PN},1}$.

n	Scale	$M=50$	$M=100$	$M=250$	$M=50$	$M=100$	$M=250$	$M=50$	$M=100$	$M=250$
		$\widehat{\widehat{\phi}}_{\mathrm{RAW},1}$			$\widehat{\widehat{\phi}}_{\mathrm{PS},1}\,(\rho=10^{-5})$			$\widehat{\widehat{\phi}}_{\mathrm{PN},1}\,(\rho=10^{-5})$		
		C_0								
50	SD	0.0034	0.0042	0.0039	0.0036	0.0042	0.0040	0.0036	0.0042	0.0039
	M-SCALE	0.0108	0.0099	0.0108	0.0108	0.0099	0.0107	0.0108	0.0099	0.0107
100	SD	0.0020	0.0020	0.0020	0.0020	0.0020	0.0021	0.0020	0.0020	0.0020
	M-SCALE	0.0049	0.0044	0.0059	0.0049	0.0044	0.0060	0.0049	0.0044	0.0059
250	SD	0.0010	0.0008	0.0007	0.0010	0.0008	0.0007	0.0010	0.0008	0.0007
	M-SCALE	0.0019	0.0025	0.0022	0.0019	0.0025	0.0023	0.0019	0.0025	0.0022
500	SD	0.0004	0.0003	0.0003	0.0005	0.0003	0.0003	0.0004	0.0003	0.0002
	M-SCALE	0.0008	0.0012	0.0008	0.0008	0.0012	0.0008	0.0008	0.0012	0.0008
1000	SD	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0001
	M-SCALE	0.0006	0.0007	0.0004	0.0006	0.0006	0.0004	0.0005	0.0006	0.0003
5000	SD	0.00004	0.00003	0.00005	0.00004	0.00004	0.00005	0.00004	0.00004	0.00005
	M-SCALE	0.0001	0.0001	0.00009	0.0001	0.0001	0.00009	0.0001	0.0001	0.0001
		C_1								
50	SD	1.3846	1.4136	1.3686	0.0993	0.1004	0.0989	0.1016	0.1010	0.1016
	M-SCALE	0.0183	0.0153	0.0108	0.0138	0.0133	0.0180	0.0128	0.0123	0.0147
100	SD	1.3699	1.3265	1.3460	0.0542	0.0513	0.0507	0.0534	0.0513	0.0536
	M-SCALE	0.0068	0.0076	0.0070	0.0058	0.0104	0.0165	0.0058	0.0098	0.0143
250	SD	0.7453	0.7611	0.7945	0.0257	0.0279	0.0223	0.0248	0.0271	0.0230
	M-SCALE	0.0036	0.0033	0.0030	0.0034	0.0043	0.0096	0.0034	0.0038	0.0083
500	SD	0.2592	0.1353	0.1535	0.0163	0.0181	0.0172	0.0165	0.0175	0.0167
	M-SCALE	0.0017	0.0017	0.0018	0.0015	0.0024	0.0075	0.0015	0.0022	0.0057
1000	SD	0.0496	0.0352	0.0340	0.0125	0.0164	0.0106	0.0125	0.0156	0.0105
	M-SCALE	0.0011	0.0001	0.0001	0.0010	0.0012	0.0037	0.0010	0.0012	0.0032
5000	SD	0.0317	0.0278	0.0274	0.0036	0.0082	0.0084	0.0037	0.0080	0.0084
	M-SCALE	0.0003	0.0002	0.0002	0.0002	0.0003	0.0012	0.0002	0.0002	0.0009

the true estimators are considered. The advantages of smoothing are in general visible when estimating directions related to higher eigenvalues. In our situation, the classical estimators computed using a penalized procedure are less sensitive to

the contamination C_1 since the rougher candidates generated by the contamination are highly penalized. This effect is more striking under Model 2 than under Model 1. Under C_0 and Model 2, all the estimators including the robust ones perform similarly due to the smoothness of the trajectories. As expected, under Model 1, the penalized estimators give larger values of the mean square error. In particular, those penalizing the scale perform worse than those penalizing the norm. We recall again that the estimators obtained through the algorithm described in Section 2.2 are not properly defined in this case, even if they may be computed due to the numerical approximation of the penalization $\Psi(\alpha)$ over the grid of points. This fact explains the large values obtained for the mean square errors under C_0 for Model 1. For instance, when $n = 5000$, the classical or robust estimators which penalize the norm have mean square errors which are four times larger than those obtained with their raw counterpart estimators. Note that the mean values of $\widehat{\Delta}$ are even larger for $\widehat{\widehat{\phi}}_{\mathrm{PS},1}$. With respect to the robust estimators, under Model 2 and C_1, some advantage is observed when penalizing in particular, for small or moderate sample sizes and $M = 50$.

5. Concluding remarks

In this paper, we derive consistency results for a computable estimator of the first principal, when the underlying distribution is elliptical. The estimators considered are a generalization to the functional setting of those proposed by Croux and Ruiz-Gazen (1996, 2005) in the finite-dimensional context. In the multivariate setting, there are other proposals to approximate the robust projection-pursuit principal component estimators, besides the *CR algorithm*. These methods show their advantage when computing the kth principal direction estimator for $k > 1$. Hubert et al. (2002) construct a fast two-step algorithm (the *reflection-based algorithm for principal components analysis*) that is expected to be more numerically stable than the *CR algorithm*. Croux et al. (2006) proposed an algorithm (the GRID algorithm) based on the idea that the optimization problem is easy when dimension p is equal to 2, since it can be solved with a grid search. Hence, when $p > 2$, the GRID algorithm performs iterative optimizations in two-dimensional planes. The generalization of the *reflection-based* or the GRID algorithm to the functional setting is an interesting topic that we leave for further research.

Another important issue is how to provide an algorithm leading to consistent penalized estimators even when the trajectories are not smooth, since Theorem 3.1 requires smoothness of the trajectories to obtain consistency of the penalized estimators defined in Section 2.2. To overcome this problem, an algorithm based on random directions as in Cuesta-Albertos et al. (2014) may be implemented as follows. Denote as $Be(p_1, p_2)$ the beta distribution of parameters p_1, p_2 and consider any basis δ_i of $\mathcal{H}$. Generate $b_0 \in [0, 1]$ such that $b_0 \sim Be(p_1, p_2)$ and given $m \geq 1$; generate $b_m \in [0, 1 - \sum_{i=0}^{m-1} b_i]$ such that $b_m \sim (1 - \sum_{i=0}^{m-1} b_i)Be(p_1, p_2)$. Define the random direction $\alpha = \sum_{i\geq 1} b_{i-1}\delta_i$. Then, in Cuesta-Albertos et al. (2014) it is shown that $\|\alpha\| = 1$ almost surely. When the trajectories are not smooth, one may compute the penalized projection-pursuit estimators maximizing over a large number of random directions generated as described (or adding these random directions to the set of candidates $\mathcal{A}_n$ if the user has no previous information on the smoothness of the trajectories). The consistency of this procedure is beyond the scope of the paper; however, we expect that the random directions provide better approximations when the trajectories are not smooth if the basis is chosen appropriately.

The consistency results stated in this paper consider only the first principal direction estimators and assume that the location parameter is known. To obtain approximations for the subsequent directions when for $2 \leq k \leq q$, one can proceed as follows. Let $\widehat{\phi}_1, \ldots, \widehat{\phi}_{k-1}$ be the previously obtained direction estimators, where we have omitted the subscripts RAW, PN or PS, for the sake of simplicity. Denote $Y_i^{(1)} = \widetilde{X}_i$ and define recursively $Y_i^{(k)} = Y_i^{(1)} - \pi_{\widehat{\mathcal{V}}_{k-1}}(Y_i^{(1)}) = Y_i^{(k-1)} - \widehat{\pi}_{k-1}(Y_i^{(k-1)})$, where $\pi_{\widehat{\mathcal{V}}_{k-1}}(Y)$ and $\widehat{\pi}_{k-1}(Y)$ stand for the orthogonal projection of Y over the linear space $\widehat{\mathcal{V}}_{k-1}$ spanned by $\widehat{\phi}_1, \ldots, \widehat{\phi}_{k-1}$ and that spanned only by $\widehat{\phi}_{k-1}$, respectively. It should be taken into account that for the penalized norm approach the orthogonal projections $\pi_{\widehat{\mathcal{V}}_{k-1}}(Y)$ and $\widehat{\pi}_{k-1}(Y)$ are projections with respect to the inner product $\langle\cdot,\cdot\rangle_\nu$. Moreover, define $\xi_n(\alpha) = s_n^2(\alpha) - \rho\Psi(\alpha)$ where $\rho = 0$ when considering the raw or the norm-penalized projection-pursuit algorithm. Let the set of candidate directions for the kth principal direction be $\mathcal{A}_{n,k}(\mathbf{Y}) = \{Y_i^{(k)}/\|Y_i^{(k)}\|_\nu,\ 1 \leq i \leq n\}$, where $\nu = 0$ for the raw or the scale-penalized projection-pursuit algorithm. An approximation for the estimator of the kth principal direction may be defined as $\widehat{\phi}_k = \mathrm{argmax}_{\alpha\in\mathcal{A}_{n,k}(\mathbf{Y})}\xi_n(\alpha)$. The generalization of the consistency results to the kth principal direction estimator defined above requires $\lambda_1 > \lambda_2 > \cdots > \lambda_k > \lambda_{k+1}$ and involves the approximation of the projection operator $\pi_{\mathcal{V}_{k-1}}$ by $\pi_{\widehat{\mathcal{V}}_{k-1}}$, where $\mathcal{V}_{k-1}$ and $\widehat{\mathcal{V}}_{k-1}$ stand for the linear spaces spanned by the true principal directions $\phi_1, \ldots, \phi_{k-1}$ and by the estimated ones, respectively. We argue that in order to ensure consistency for the estimators of the kth principal direction computed using the *CR algorithm*, the distance between the projection operators, that is, the norm $\|\pi_{\mathcal{V}_{k-1}} - \pi_{\widehat{\mathcal{V}}_{k-1}}\|$ needs to converge to 0 with some given rate instead of just requiring its convergence to 0. We leave this challenging problem as well as the consistency of the method when the location parameter is estimated, however, as open topics for future research.

Acknowledgments

This research was partially supported by Grants W276 from Universidad of Buenos Aires, PIP 112-2011-01-00339 from CONICET and PICT 2011-0397 from ANPCYT, Argentina. The authors wish to thank the Associate Editor and two anonymous referees for valuable comments which led to an improved version of the original paper.

Appendix. Proofs

We first begin by deriving the following lemma that will be useful in the sequel. Recall that $\mathcal{H}_S \subset \mathcal{H}$ is the linear space of "smooth elements" of $\mathcal{H}$ and the domain of D. Thus, for any $\alpha \in \mathcal{H}_S$ we have $\Psi(\alpha) < \infty$. As mentioned above, when $\nu = \rho = 0$, we set $\mathcal{H}_S = \mathcal{H}$.

Lemma A.1. *Let $X \in \mathcal{H}$ where $\mathcal{H}$ is a separable Hilbert space be such that $\mathbb{P}(X = 0) = 0$. Assume that $X \sim \mathcal{E}(0, \mathbf{\Gamma})$ with $\mathbf{\Gamma}$ a compact operator such that $\mathbf{\Gamma} = \sum_{i\geq 1} \lambda_i \phi_i \otimes \phi_i$, with $\lambda_1 \geq \lambda_2 \geq \ldots$ its ordered eigenvalues and ϕ_i the related eigenfunctions. Let $\nu_n \geq 0$ be a numeric sequence such that $\nu = \nu_n \to 0$. Furthermore, when $\nu_n > 0$, assume that $\mathbb{P}(X \in \mathcal{H}_S) = 1$. Given $0 < \epsilon < 1$, let $p_n = \mathbb{P}\left(\langle X/\|X_1\|_{\nu_n}, \phi_1\rangle^2 < 1 - \epsilon\right)$. Then, there exist $n_0 \in \mathbb{N}$ and $0 < q < 1$ such that $p_n \leq q < 1$ for all $n \geq n_0$.*

Proof. Let $p(\epsilon) = \mathbb{P}\left(\langle X/\|X\|, \phi_1\rangle^2 < 1 - \epsilon\right)$. First, we begin by showing that

$$p(\epsilon) < 1 \quad \text{for all } 0 < \epsilon < 1. \tag{A.1}$$

We will consider two cases: (a) $\mathbf{\Gamma}$ has infinite positive eigenvalues and (b) $\mathbf{\Gamma}$ has finite rank.

(a) When the rank of $\mathbf{\Gamma}$ is not finite, it is clear that $\lambda_1 > 0$. Besides, Proposition 2.1 in Boente et al. (2014) entails that $X = VY$ where Y is a Gaussian process with zero mean and covariance operator $\mathbf{\Gamma}_Y$ proportional to $\mathbf{\Gamma}$, say $\mathbf{\Gamma}_Y = a\mathbf{\Gamma}$ for some $a \geq 0$, and V is a non-negative real random variable independent of Y. Using that $\mathbb{P}(X = 0) = 0$ and Y is Gaussian, we get that $a > 0$ and $\mathbb{P}(V = 0) = 0$. Without loss of generality, we will assume that $a = 1$.

Using that Y is a Gaussian process with covariance operator $\mathbf{\Gamma}_Y = \mathbf{\Gamma}$, from the Karhunen–Loève decomposition, we obtain that $Y = \sum_{j\geq 1} \lambda_j^{1/2} Z_j \phi_j$ in $\mathcal{H}$ where $Z_j = \langle Y, \phi_j\rangle \lambda_j^{-1/2}$ are independent and $Z_j \sim N(0, 1)$. Moreover, $Y - \lambda_1^{1/2} Z_1 \phi_1$ is independent of $Z_1\phi_1$. Therefore,

$$\left\langle \frac{X}{\|X\|}, \phi_1 \right\rangle^2 = \left\langle \frac{VY}{|V|\|Y\|}, \phi_1 \right\rangle^2 = \frac{V^2}{|V|^2} \left\langle \frac{Y}{\|Y\|}, \phi_1 \right\rangle^2 = \left\langle \frac{Y}{\|Y\|}, \phi_1 \right\rangle^2 = \frac{\lambda_1 Z_1^2}{\|Y\|^2}.$$

Then, using that $\mathbb{P}(\|Y\| > 0) = 1$ we get

$$p(\epsilon) = \mathbb{P}\left(\frac{\lambda_1 Z_1^2}{\|Y\|^2} < 1 - \epsilon\right) = \mathbb{P}\left(\lambda_1 Z_1^2 - (1 - \epsilon)\|Y\|^2 < 0\right) = \mathbb{P}\left(\epsilon\lambda_1 Z_1^2 - U_2 < 0\right),$$

where $U_2 = (1 - \epsilon)\|Y - \lambda_1^{1/2} Z_1 \phi_1\|^2$.

Observe that $Z_j^2 \sim \chi_1^2$, also Z_1^2 and U_2 are independent such that $\mathbb{P}(U_2 > 0) = 1$ since the rank $\mathbf{\Gamma}$ is not finite. Define the function $g : \mathbb{R}_{>0} \to [0, 1]$ as $g(u) = \mathbb{P}\left(Z_1^2 < u/(\epsilon\lambda_1)\right)$. Note that, $0 < g(u) < 1$ for all $u > 0$, since $\lambda_1 > 0$. Using that $p(\epsilon) = \mathbb{E}\left\{\mathbb{P}\left(\epsilon\lambda_1 Z_1^2 < U_2 | U_2\right)\right\}$, the independence between Z_1 and U_2 we obtain $p(\epsilon) = \mathbb{E}g(U_2)$ which entails that $p(\epsilon) < 1$.

Observe that if X is Gaussian, the result would be valid even when the rank of $\mathbf{\Gamma}$ is finite and equal to $r > 1$, that is, if $Z_j = \langle X, \phi_j\rangle/\lambda_j^{1/2} \sim N(0, 1)$ are independent, then,

$$\mathbb{P}\left(\frac{\lambda_1 Z_1^2}{\sum_{j=1}^{r} \lambda_j Z_j^2} < 1 - \epsilon\right) < 1. \tag{A.2}$$

(b) Let us now assume that $\mathbf{\Gamma}$ has finite rank. Note that since $\mathbb{P}(X = 0) = 0$, $\mathbf{\Gamma}$ has rank r greater than or equal to 1, that is, $\mathbf{\Gamma} = \sum_{i=1}^{r} \lambda_i \phi_i \otimes \phi_i$ with $\lambda_1 \geq \cdots \geq \lambda_r > 0$. In this case, the arguments to be used are analogous to those to be considered in the multivariate case, since $\mathbb{P}(X = \pi(X)) = 1$ where π is the orthogonal projection over the linear space spanned by $\phi_1, \ldots, \phi_r$. Effectively, define $\mathbf{Y} = (Y_1, \ldots, Y_r)$ with $Y_j = \langle X, \phi_j\rangle$, and then $\mathbf{Y} \sim \mathcal{E}_r(0, \mathbf{\Lambda})$ where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_r)$ and $\langle X/\|X\|, \phi_1\rangle^2 = Y_1^2/\sum_{j=1}^{r} Y_j^2$. Moreover, $\mathbb{P}(\mathbf{Y} = \mathbf{0}) = 0$, since $\lambda_1 \geq \cdots \geq \lambda_r > 0$. Let $\mathbf{Z} = \mathbf{\Lambda}^{-1/2}\mathbf{Y}$, and then $\mathbf{Z}$ has a spherical distribution, so that $\mathbf{U} = \mathbf{Z}/\|\mathbf{Z}\|$ is uniform over the unit sphere in $\mathbb{R}^r$. Moreover, the fact that $\mathbf{U} \sim \mathbf{V}/\|\mathbf{V}\|$ where $\mathbf{V} \sim N(\mathbf{0}, \mathbf{I})$ implies that $Y_1^2/\sum_{j=1}^{r} Y_j^2 = \lambda_1 Z_1^2/\sum_{j=1}^{r} \lambda_j Z_j^2 = \lambda_1 U_1^2/\sum_{j=1}^{r} \lambda_j U_j^2 \sim \lambda_1 V_1^2/\sum_{j=1}^{r} \lambda_j V_j^2$. Hence, we have that $p(\epsilon) = \mathbb{P}\left(\langle X/\|X\|, \phi_1\rangle^2 < 1 - \epsilon\right) = \mathbb{P}\left(\lambda_1 V_1^2/\sum_{j=1}^{r} \lambda_j V_j^2 < 1 - \epsilon\right)$ and (A.1) follows now easily using (A.2) since V_j are i.i.d. $V_j \sim N(0, 1)$.

So we have concluded the proof of (A.1). Define $q = (1 + p(\epsilon/2))/2 < 1$, and we will show that $p_n \leq q$ for large enough n. Using that $\langle X_1/\|X_1\|_{\nu_n}, \phi_1\rangle = (\|X_1\|/\|X_1\|_{\nu_n})\langle X_1/\|X_1\|, \phi_1\rangle$, we get that

$$p_n = \mathbb{P}\left(\left\langle \frac{X_1}{\|X_1\|}, \phi_1 \right\rangle^2 < (1 - \epsilon)\frac{\|X_1\|_{\nu_n}^2}{\|X_1\|^2}\right).$$

The convergence $\nu_n \to 0$ implies that $(1-\epsilon)\|X_1\|_{\nu_n}^2/\|X_1\|^2 \xrightarrow{a.s.}(1-\epsilon)$. Therefore, using that $\|X_1\|_{\nu_n} \geq \|X_1\|$, we get that $\lim_{n\to\infty}\mathbb{P}(\mathcal{B}_n)=1$, where $\mathcal{B}_n=\{\omega\in\Omega : 1-\epsilon < (1-\epsilon)\|X_1(\omega)\|_{\nu_n}^2/\|X_1(\omega)\|^2 < 1-\epsilon/2\}$. Hence, given $\delta=\eta/2$ where $\eta<1-p(\epsilon/2)$ there exists n_0 such that for all $n\geq n_0$, $\mathbb{P}(\mathcal{B}_n)\geq 1-\delta$. Thus,

$$
\begin{aligned}
p_n &= \mathbb{P}\left(\left\langle \frac{X_1}{\|X_1\|}, \phi_1\right\rangle^2 < (1-\epsilon)\frac{\|X_1\|_{\nu_n}^2}{\|X_1\|^2}\cap \mathcal{B}_n\right) + \mathbb{P}\left(\left\langle \frac{X_1}{\|X_1\|}, \phi_1\right\rangle^2 < (1-\epsilon)\frac{\|X_1\|_{\nu_n}^2}{\|X_1\|^2}\cap \mathcal{B}_n^c\right)\\
&\leq \mathbb{P}\left(\left\langle \frac{\|X_1\|}{\|X_1\|}, \phi_1\right\rangle^2 < 1-\epsilon/2\right) + \mathbb{P}\left(\mathcal{B}_n^c\right) = p(\epsilon/2)+\delta < \frac{1+p(\epsilon/2)}{2} = q < 1,
\end{aligned}
$$

concluding the proof. □

For simplicity, from now on we will denote by $o_{\text{a.s.}}(1)$ any term that converges to 0 almost everywhere.

Proof of Theorem 3.1. Recall that, since X_i is elliptically distributed, $\sigma^2(\alpha)=c\langle\alpha,\boldsymbol{\Gamma}\alpha\rangle$, where we may assume that $c=1$. Therefore, $\phi_{\mathrm{R},1}(P)=\phi_1$ and $\lambda_{\mathrm{R},1}(P)=\sigma^2(\phi_1)=\lambda_1>0$, since $\lambda_1>\lambda_2\geq 0$, which implies that $\phi_{\mathrm{R},1}(P)$ is uniquely defined except for a sign change. Moreover, since $\boldsymbol{\Gamma}$ is a compact operator, we also have that $\sigma:\mathcal{H}\to\mathbb{R}$ is a weakly continuous function.

Let $\tau=\max\{\nu,\rho\}$, define $\mathcal{A}_{n,\tau}=\{X_i/\|X_i\|_\tau,\ 1\leq i\leq n\}$, and $\widetilde{\phi}_1=\widetilde{\phi}_{1,n}=\operatorname{argmin}_{\alpha\in\mathcal{A}_{n,\tau}}\left\{1-\langle\alpha,\phi_1\rangle^2\right\}$, that is, $\langle\widetilde{\phi}_1,\phi_1\rangle^2=\max_{1\leq i\leq n}\langle X_i/\|X_i\|_\tau,\phi_1\rangle^2$. We will show that $\langle\widetilde{\phi}_1,\phi_1\rangle^2\xrightarrow{a.s.}1$, that is, $\widetilde{\phi}_1\xrightarrow{a.s.}\phi_1$ except for a possible sign change.

First, observe that $|\langle X_i/\|X_i\|_\tau,\phi_1\rangle|\leq\|X_i\|/\|X_i\|_\tau\leq 1$, implies that $\langle\widetilde{\phi}_1,\phi_1\rangle^2\leq 1$. Thus, we have to show that, for any $\epsilon>0$, $\lim_{M\to\infty}\mathbb{P}\left(\bigcup_{N\geq M}(1-\langle\widetilde{\phi}_{1,N},\phi_1\rangle^2)>\epsilon\right)=0$, where we have stressed the dependence on N for clarity. Hence, by Borel–Cantelli, it is enough to show that $\sum_{n\geq1}\mathbb{P}((1-\langle\widetilde{\phi}_{1,n},\phi_1\rangle^2)>\epsilon)<\infty$. Using that $\nu=\nu_n$ is a fixed numerical sequence, we have that

$$
\begin{aligned}
\mathbb{P}((1-\langle\widetilde{\phi}_{1,n},\phi_1\rangle^2)>\epsilon) &= \mathbb{P}\left(\max_{1\leq i\leq n}\left\langle\frac{X_i}{\|X_i\|_\tau},\phi_1\right\rangle^2<1-\epsilon\right)=\prod_{i=1}^n\mathbb{P}\left(\left\langle\frac{X_i}{\|X_i\|_\tau},\phi_1\right\rangle^2<1-\epsilon\right)\\
&= \left[\mathbb{P}\left(\left\langle\frac{X_1}{\|X_1\|_\tau},\phi_1\right\rangle^2<1-\epsilon\right)\right]^n = p_n^n,
\end{aligned}
$$

with $p_n=\mathbb{P}(\langle X_1/\|X_1\|_\nu,\phi_1\rangle^2<1-\epsilon)$. Lemma A.1 implies that there exist $n_0\in\mathbb{N}$ and $0<q<1$ such that $p_n\leq q<1$ for $n\geq n_0$, so that $\mathbb{P}(1-\langle\widetilde{\phi}_{1,n},\phi_1\rangle^2>\epsilon)\leq q^n$ for $n\geq n_0$. Thus, $\sum_{n=1}^\infty\mathbb{P}(1-\langle\widetilde{\phi}_{1,n},\phi_1\rangle^2>\epsilon)<\infty$ which entails that $\langle\widetilde{\phi}_1,\phi_1\rangle^2\xrightarrow{a.s.}1$. Without loss of generality, let us assume that $\widetilde{\phi}_1\xrightarrow{a.s.}\phi_1$. Using that $\|X_i\|/\|X_i\|_\tau\leq 1$, we get that $\|\widetilde{\phi}_1\|\leq 1$, which implies that $\langle\widetilde{\phi}_1,\phi_1\rangle^2\leq\|\widetilde{\phi}_1\|^2\|\phi_1\|^2=\|\widetilde{\phi}_1\|^2\leq 1$. Therefore, $\|\widetilde{\phi}_1\|^2\xrightarrow{a.s.}1$. Noting that $\|\widetilde{\phi}_1\|_\tau=1$, we obtain that $\tau\Psi(\widetilde{\phi}_1)\xrightarrow{a.s.}0$, which implies that $\nu\Psi(\widetilde{\phi}_1)\xrightarrow{a.s.}0$ and $\rho\Psi(\widetilde{\phi}_1)\xrightarrow{a.s.}0$.

As in the proof of Theorem 4.1 of Bali et al. (2011), **A1** implies

$$
\sup_{\|\alpha\|\leq1}\left|s_n^2(\alpha)-\sigma^2(\alpha)\right|\xrightarrow{a.s.}0. \tag{A.3}
$$

Therefore, using $\|\alpha\|_\tau=1$ implies that $\|\alpha\|\leq 1$, we get that $\sup_{\|\alpha\|_\tau=1}|s_n^2(\alpha)-\sigma^2(\alpha)|\xrightarrow{a.s.}0$. The weak continuity of σ entails that $\sigma^2(\widetilde{\phi}_1)-\sigma^2(\phi_1)\xrightarrow{a.s.}0$, while (A.3) implies that $s_n^2(\widetilde{\phi}_1)-\sigma^2(\widetilde{\phi}_1)\xrightarrow{a.s.}0$, which lead to $s_n^2(\widetilde{\phi}_1)\xrightarrow{a.s.}\sigma^2(\phi_1)=\lambda_{\mathrm{R},1}(P)>0$. Therefore, we have that $\widehat{b}_n=s_n^2(\widetilde{\phi}_1)-\sigma^2(\phi_1)\xrightarrow{a.s.}0$. Using $\widetilde{\phi}_1/\|\widetilde{\phi}_1\|\in\mathcal{A}_n$, we obtain the following bounds

$$
\begin{aligned}
s_n^2(\widehat{\widetilde{\phi}}_1) &\geq \frac{s_n^2(\widehat{\widetilde{\phi}}_1)}{1+\nu\Psi(\widehat{\widetilde{\phi}}_1)}\geq\frac{s_n^2(\widehat{\widetilde{\phi}}_1)-\rho\Psi(\widehat{\widetilde{\phi}}_1)}{1+\nu\Psi(\widehat{\widetilde{\phi}}_1)}=\max_{\alpha\in\mathcal{A}_n}\frac{s_n^2(\alpha)-\rho\Psi(\alpha)}{1+\nu\Psi(\alpha)}\\
&\geq \frac{s_n^2\left(\frac{\widetilde{\phi}_1}{\|\widetilde{\phi}_1\|}\right)-\rho\Psi\left(\frac{\widetilde{\phi}_1}{\|\widetilde{\phi}_1\|}\right)}{1+\nu\Psi\left(\frac{\widetilde{\phi}_1}{\|\widetilde{\phi}_1\|}\right)}=\frac{s_n^2\left(\widetilde{\phi}_1\right)-\rho\Psi\left(\widetilde{\phi}_1\right)}{\|\widetilde{\phi}_1\|^2+\nu\Psi\left(\widetilde{\phi}_1\right)},
\end{aligned} \tag{A.4}
$$

which together with the fact that $\|\widetilde{\phi}_1\|\leq 1$ entail

$$
s_n^2(\widehat{\widetilde{\phi}}_1)\geq\frac{s_n^2\left(\widetilde{\phi}_1\right)-\rho\Psi\left(\widetilde{\phi}_1\right)}{1+\nu\Psi\left(\widetilde{\phi}_1\right)}\geq s_n^2\left(\widetilde{\phi}_1\right)-\frac{\rho\Psi\left(\widetilde{\phi}_1\right)}{1+\nu\Psi\left(\widetilde{\phi}_1\right)}-\frac{\nu\Psi\left(\widetilde{\phi}_1\right)}{1+\nu\Psi\left(\widetilde{\phi}_1\right)}s_n^2\left(\widetilde{\phi}_1\right). \tag{A.5}
$$

Using that $\widehat{b}_n = s_n^2(\widetilde{\phi}_1) - \sigma^2(\phi_1) \xrightarrow{a.s.} 0$, $\nu\Psi(\widetilde{\phi}_1) \xrightarrow{a.s.} 0$ and $\rho\Psi(\widetilde{\phi}_1) \xrightarrow{a.s.} 0$, we get

$$s_n^2(\widehat{\widehat{\phi}}_1) \geq \sigma^2(\phi_1) + o_{\text{a.s.}}(1). \tag{A.6}$$

To conclude the proof it remains to show that

$$s_n^2(\widehat{\widehat{\phi}}_1) \leq \sigma^2(\phi_1) + o_{\text{a.s.}}(1). \tag{A.7}$$

Note that $\|\widehat{\widehat{\phi}}_1\| = 1$; hence $\sigma^2(\phi_1) \geq \sigma^2(\widehat{\widehat{\phi}}_1) = s_n^2(\widehat{\widehat{\phi}}_1) - d_n$, where $d_n = s_n^2(\widehat{\widehat{\phi}}_1) - \sigma^2(\widehat{\widehat{\phi}}_1) \xrightarrow{a.s.} 0$ from **A1**.

Using (A.6) and (A.7), we get that $s_n^2(\widehat{\widehat{\phi}}_1) \xrightarrow{a.s.} \sigma^2(\phi_1)$, and the proof follows now from Lemma 4.1 in Bali et al. (2011).

(b) Note that from (A.4)–(A.6) we have that

$$s_n^2(\widehat{\widehat{\phi}}_1) \geq \frac{s_n^2(\widehat{\widehat{\phi}}_1)}{1 + \nu\Psi(\widehat{\widehat{\phi}}_1)} \geq \sigma^2(\phi_1) + o_{\text{a.s.}}(1).$$

Hence, using that $s_n^2(\widehat{\widehat{\phi}}_1) \xrightarrow{a.s.} \sigma^2(\phi_1)$, we obtain that $\nu\Psi(\widehat{\widehat{\phi}}_1) \xrightarrow{a.s.} 0$. Finally, from (A.4) and (A.6), we get

$$s_n^2(\widehat{\widehat{\phi}}_1) \geq \frac{s_n^2(\widehat{\widehat{\phi}}_1) - \rho\Psi(\widehat{\widehat{\phi}}_1)}{1 + \nu\Psi(\widehat{\widehat{\phi}}_1)} \geq \sigma^2(\phi_1) + o_{\text{a.s.}}(1),$$

which together with the fact that $\nu\Psi(\widehat{\widehat{\phi}}_1) \xrightarrow{a.s.} 0$ entails that $\rho\Psi(\widehat{\widehat{\phi}}_1) \xrightarrow{a.s.} 0$.

(c) The proof of (c) follows immediately from the fact that $s_n^2(\widehat{\widehat{\phi}}_1) \xrightarrow{a.s.} \sigma^2(\phi_1)$. □

References

Bali, J., Boente, G., 2009. Principal points and elliptical distributions from the multivariate setting to the functional case. Statist. Probab. Lett. 79, 1858–1865.
Bali, J., Boente, G., Tyler, D.E., Wang, J.L., 2011. Robust functional principal components: a projection-pursuit approach. Ann. Statist. 39, 2852–2882.
Boente, G., Fraiman, R., 2000. Kernel-based functional principal components. Statist. Probab. Lett. 48, 335–345.
Boente, G., Salibian-Barrera, M., Tyler, D., 2014. A characterization of elliptical distributions and some optimality properties of principal components for functional data. J. Multivar. Anal. http://dx.doi.org/10.1016/j.jmva.2014.07.006 (in press).
Croux, C., Filzmoser, P., Oliveira, M.R., 2006. Algorithms for projection-pursuit robust principal component analysis. Chemometr. Intell. Lab. 87, 218–225.
Croux, C., Ruiz-Gazen, A., 1996. A fast algorithm for robust principal components based on projection pursuit. In: Prat, A. (Ed.), Compstat: Proceedings in Computational Statistics. Physica-Verlag, Heidelberg, pp. 211–217.
Croux, C., Ruiz-Gazen, A., 2005. High breakdown estimators for principal components: the projection-pursuit approach revisited. J. Multivar. Anal. 95, 206–226.
Cuesta-Albertos, J., Gamboa, F., Nieto-Reyes, A., 2014. A random–projection based procedure to test if a stationary process is Gaussian. Comput. Statist. Data Anal. 75, 124–141.
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. Multivar. Anal. 12, 136–154.
Gervini, D., 2006. Free–knot spline smoothing for functional data. J. Roy. Statist. Soc. B 68, 671–687.
Gervini, D., 2008. Robust functional estimation using the spatial median and spherical principal components. Biometrika 95, 587–600.
Hall, P., Hosseini-Nasab, M., 2006. On properties of functional principal components analysis. J. Roy. Statist. Soc. B 68, 109–126.
Hall, P., Müller, H.-G., Wang, J.-L., 2006. Properties of principal component methods for functional and longitudinal data analysis. Ann. Statist. 34, 1493–1517.
Hubert, M., Rousseeuw, P.J., Verboven, S., 2002. A fast method for robust principal components with applications to chemometrics. Chemometr. Intell. Lab. 60, 101–111.
Hyndman, R.J., Ullah, S., 2007. Robust forecasting of mortality and fertility rates: A functional data approach. Comput. Statist. Data Anal. 51, 4942–4956.
Li, G., Chen, Z., 1985. Projection–pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. J. Amer. Statist. Assoc. 80, 759–766.
Locantore, N., Marron, J.S., Simpson, D.G., Tripoli, N., Zhang, J.T., Cohen, K.L., 1999. Robust principal components for functional data (with Discussion). Test 8, 1–73.
Ramsay, J.O., Silverman, B.W., 2005. Functional Data Analysis, second ed. Springer-Verlag, New York.
Rice, J., Silverman, B.W., 1991. Estimating the mean and covariance structure nonparametrically when the data are curves. J. Roy. Statist. Soc. B 53, 233–243.
Sawant, P., Billor, N., Shin, H., 2012. Functional outlier detection with robust functional principal component analysis. Comput. Statist. 27, 83–102.
Silverman, B.W., 1996. Smoothed functional principal components analysis by choice of norm. Ann. Statist. 24, 1–24.
Yao, F., Lee, T.C.M., 2006. Penalized spline models for functional principal component analysis. J. R. Stat. Soc. Ser. B Stat. Methodol. 68, 3–25.