Robust estimators under a functional common principal components model *

Lucas Bali
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
Graciela Boente
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina

Abstract

When dealing with several populations of functional data, equality of the covariance operators is often assumed even when seeking for a lower–dimensional approximation to the data. Usually, if this assumption does not hold, one estimates the covariance operator of each group separately, which leads to a large number of parameters. As in the multivariate setting, this is not satisfactory since the covariance operators may exhibit some common structure. In this paper, we discuss the extension to the functional setting of projection–pursuit estimators for the common directions under a common principal component model that has been widely studied when dealing with multivariate observations. We present estimators of the unknown parameters combining robust projection–pursuit with different smoothing methods. We obtain consistency results under mild assumptions.

Key Words: Fisher–consistency; Functional common principal component model; Outliers; Projection–pursuit; Robust estimation

AMS Subject Classification: MSC 62G35, 62H25 MSC 62G20

*This research was partially supported by Grants X018 and 20020100100276 from Universidad of Buenos Aires, PIP 112-200801-00216 from CONICET and PICT 2011-0397 from ANPCYT, Argentina.

1 Introduction

The common principal components, CPC, model introduced by Flury (1984) for $p-$th dimensional data, generalizes proportionality of the covariance matrices by allowing the matrices to have different eigenvalues but identical eigenvectors, that is, $\boldsymbol{\Sigma}_i = \boldsymbol{\beta}\boldsymbol{\Lambda}_i\boldsymbol{\beta}^{\mathrm{T}}$, $1 \le i \le k$, where $\boldsymbol{\Lambda}_i$ are diagonal matrices and $\boldsymbol{\beta}$ is the orthogonal matrix of the common eigenvectors. This model can be viewed as a generalization of principal components to k groups, since the principal transformation is identical in all populations considered while the variances associated with them vary among groups. In biometric applications, principal components are frequently interpreted as independent factors determining the growth, size or shape of an organism. It seems therefore reasonable to consider a model in which the same factors arise in different, but related species. The common principal components model clearly serves this purpose.

In this paper, we go further and we will consider several populations of functional data instead of finite–dimensional ones. To be more precise, the observations to be considered are elements of a separable Hilbert space $\mathcal{H}$ with the inner product $\langle \cdot, \cdot \rangle$ and related norm $\|\alpha\|^2 = \langle \alpha, \alpha \rangle$. If $X \in \mathcal{H}$ is a random element with finite second moment, i.e., $\mathbb{E}(\|X\|^2 < \infty$, the bilinear operator $a_X : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ defined as $a_X(\alpha, \beta) = \mathrm{cov}(\langle \alpha, X \rangle, \langle \beta, X \rangle)$ leads to a continuous operator. Hence, Riesz representation theorem implies that there exists a bounded operator, $\boldsymbol{\Gamma}_X : \mathcal{H} \to \mathcal{H}$, such that $a_X(\alpha, \beta) = \langle \alpha, \boldsymbol{\Gamma}_X \beta \rangle$. The operator $\boldsymbol{\Gamma}_X$ is called the covariance operator of X and is linear, self–adjoint and continuous. Moreover, $\boldsymbol{\Gamma}_X$ is a Hilbert-Schmidt operator so, it has a countable number of eigenvalues, all of which are real. Furthermore, since the covariance operator $\boldsymbol{\Gamma}_X$ is also positive semi-definite, its eigenvalues are non-negative. As with symmetric matrices in finite dimensional Euclidean spaces, one can choose the eigenfunctions of a Hilbert-Schmidt operator so that they form an orthonormal basis for $\mathcal{H}$. Let $\{\phi_j : j \ge 1\}$ and $\{\lambda_j : j \ge 1\}$ be respectively an orthonormal basis of eigenfunctions and their corresponding eigenvalues for the covariance operator $\boldsymbol{\Gamma}_X$, with $\lambda_j \ge \lambda_{j+1}$. Let $\otimes$ stand for the tensor product on $\mathcal{H}$, e.g., for $u, v \in \mathcal{H}$, the operator $u \otimes v : \mathcal{H} \to \mathcal{H}$ is defined as $(u \otimes v)w = \langle v, w \rangle u$. With this notation, the spectral value decomposition for $\boldsymbol{\Gamma}_X$ can then be expressed as $\boldsymbol{\Gamma}_X = \sum_{j=1}^{\infty} \lambda_j \phi_j \otimes \phi_j$. Besides, the covariance operator $\boldsymbol{\Gamma}_X$ can also be written as $\boldsymbol{\Gamma}_X = \mathbb{E}\{(X - \mu) \otimes (X - \mu)\}$. In particular, principal components analysis has been successfully extended from the multivariate setting to accommodate functional data. The j-th principal component variable is defined as $Z_j = \langle \phi_j, X - \mu \rangle$, leading to the Karhunen–Loève expansion $X = \mu + \sum_{j=1}^{\infty} Z_j \phi_j$, with $\mu = \mathbb{E}(X)$ and the Z_j's being uncorrelated and having variances λ_j in descending order.

As in the $p-$dimensional case, in many situations, one collects functional data $X_{i,1}, \cdots, X_{i,n_i}$ from k independent samples with mean μ_i and different covariance operators $\boldsymbol{\Gamma}_i$ which may exhibit some common structure to be taken into account in the estimation procedure. The simplest generalization of equal covariance operators consists of assuming their proportionality, i.e., $\boldsymbol{\Gamma}_i = \rho_i \boldsymbol{\Gamma}_1$, for $1 \le i \le k$ and $\rho_1 = 1$. On the other hand, a natural extension of functional principal components to several populations, which also corresponds to a generalization to the functional setting of the CPC model introduced by Flury (1984), is to assume that the covariance operators $\boldsymbol{\Gamma}_i$ have common eigenfunctions ϕ_j but different eigenvalues $\lambda_{i,j}$, i.e.,

$$\boldsymbol{\Gamma}_i = \sum_{j=1}^{\infty} \lambda_{i,j} \phi_j \otimes \phi_j \ . \tag{1}$$

As in Boente *et al.* (2010), we will assume that the eigenvalues preserve the order among populations, i.e., $\lambda_{i,1} \ge \lambda_{i,2} \ge \cdots \ge \lambda_{i,j} \ge \lambda_{i,j+1} \cdots$, for $1 \le i \le k$. In this sense, the processes $X_{i,1}$, $1 \le i \le k$, can be written as $X_{i,1} = \mu_i + \sum_{j=1}^{\infty} \lambda_{i,j}^{\frac{1}{2}} \xi_{ij} \phi_j$, with $\lambda_{i1} \ge \lambda_{i2} \ge \ldots \ge 0$ and ξ_{ij}

zero mean random variables such that $\mathbb{E}(\xi_{ij}^2) = 1$, $\mathbb{E}(\xi_{ij}\ \xi_{is}) = 0$ for $j \neq s$ and so, the common eigenfunctions, as in the one–population setting, exhibit the same major modes of variation. This model is usually denoted the functional common principal component (FCPC) model. As in principal component analysis, the FCPC model could be used to reduce the dimensionality of the data, retaining as much as possible of the variability present in each of the populations. Besides, this model provides a framework for analysing different population data that share their main modes of variation $\phi_1, \phi_2, \ldots$. It is worth noticing that when considering a functional proportional model, $X_{i,1}$, $1 \leq i \leq k$, can be written as $X_{i,1} = \mu_i + \rho_i^{\frac{1}{2}} \sum_{j=1}^{\infty} \lambda_j^{\frac{1}{2}} \xi_{ij}\, \phi_j$, with $\rho_1 = 1$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ and ξ_{ij} random variables as described above. A similar problem was studied by Benko *et al.* (2009) who considered the case of $k = 2$ populations and provide tests for equality of means and equality of a fixed number of eigenfunctions and by Boente *et al.* (2010) who considered estimators under a general FCPC model.

The estimators defined in Boente *et al.* (2010) are based on the sample covariance operators of each population being, therefore, sensitive to atypical trajectories. Clearly, when $X_{i,1} \in L^2(\mathcal{I})$, with $\mathcal{I} \subset \mathbb{R}$ a finite interval, it is always possible to reduce the functional problem to a multivariate one, by evaluating the observations on a common output grid or by using the coefficients of a basis expansion, as done in Locantore *et al.* (1999), for the one–population case. However, even when $k = 1$, this approach has several drawbacks which are discussed, for instance, in Gervini (2008), who also study a fully functional approach to robust estimation of the principal components. Up to our knowledge, robust proposals for functional principal components consider only the one–population case. For instance, Gervini (2009) develop robust functional principal component estimators for sparsely and irregularly observed functional data and use it for outlier detection while, Sawant *et al.* (2011) consider a robust approach of principal components based on a robust eigen–analysis of the coefficients of the observed data on some known basis. On the other hand, Hyndman and Ullah (2007) give an application of a robust projection–pursuit approach, applied to smoothed trajectories, but do not study the properties of their method in detail. More recently, Bali *et al.* (2011) introduce robust estimators of the principal directions based on robust projection–pursuit combined with different smoothing methods through a penalization in the scale or in the norm and establish their strong consistency. On the other hand, when dealing with several populations of multivariate observations, robust estimators under a CPC model are considered in Boente and Orellana (2001). Further developments are given by Boente *et al.* (2006) who define a general class of projection–pursuit estimators in order to improve the efficiency of the robust estimators for a given scale and also, to recover the maximum likelihood estimators when the scale is the standard deviation.

The aim of this paper is to extend some of the previous proposals to a functional setting with observations from several populations, in order to provide robust estimators of the common directions under a FCPC model. In Section 2, we introduce the notation to be used and we generalize the FCPC model described above to avoid second moment conditions. In Section 3, robust estimators for the common directions are considered through a projection–pursuit approach combined with different smoothing procedures. The strong consistency of the given proposals is stated in Section 4 and a robust cross–validation procedure to select the penalization parameter is described in Section 5. Section 6 summarizes the results of a Monte Carlo study conducted to compare the performance of the robust proposals between them and also with that of the classical estimators based on the sample variance while a real data set is studied in Section 7. Some preliminary results are given in Appendix A while proofs are relegated to Appendix B.

2 Preliminaries

2.1 The FCPC model: notation and definitions

We recall some definitions given in Bali *et al.* (2011) which will help to generalize the FCPC model to the situation in which second moments do not exist.

If $X \sim P$ and $\alpha \in \mathcal{H}$, $P[\alpha]$ will denote the measure of the real random variable $\langle \alpha, X \rangle$. In the context of several independent populations with probability measures $P_1, \dots, P_k$, that is, when $X_{i,1}, \cdots, X_{i,n_i}$ are independent and such that $X_{i,j} \sim X_{i,1} \sim P_i$, we will denote by P_{i,n_i} the empirical measure under P_i, i.e.,

$$P_{i,n_i}(A) = \frac{1}{n_i} \sum_{j=1}^{n_i} I_A(X_{i,j}).$$

while $P_{i,n_i}[\alpha]$ will be the empirical measure of the real random variables $\{\langle X_{i,1}, \alpha \rangle, \dots, \langle X_{i,n_i}, \alpha \rangle\}$.

Denote by $\mathcal{G}$ the set of all univariate distributions and $\sigma_{\mathrm{R}} : \mathcal{G} \to [0, +\infty)$ a scale functional, that is, a functional over the set of univariate distributions which is location invariant and scale equivariant, i.e., if $G_{a,b}$ stands for the distribution of $aY + b$ when $Y \sim G$, then, $\sigma_{\mathrm{R}}(G_{a,b}) = |a| \sigma_{\mathrm{R}}(G)$, for all real numbers a and b. We refer to Bali *et al.* (2011) for a discussion on scale functionals in the context of functional principal components.

Given a probability measure P and a scale functional σ_{R}, robust functional principal components operators were defined in Bali *et al.* (2011) as

$$\begin{cases} \phi_{\mathrm{R},1}(P) & = \underset{\|\alpha\|=1}{\operatorname{argmax}} \, \sigma_{\mathrm{R}} \left(P[\alpha] \right) \\ \phi_{\mathrm{R},m}(P) & = \underset{\|\alpha\|=1, \alpha \in \mathcal{B}_m}{\operatorname{argmax}} \, \sigma_{\mathrm{R}} \left(P[\alpha] \right) \quad 2 \le m \,, \end{cases} \tag{2}$$

where $\mathcal{B}_m = \{\alpha \in \mathcal{H} : \langle \alpha, \phi_{\mathrm{R},j}(P) \rangle = 0, \ 1 \le j \le m-1\}$. These authors also define the robust eigenvalue operators as $\lambda_{\mathrm{R},i}(P) = \sigma_{\mathrm{R}}^2(P[\phi_{\mathrm{R},i}(P)])$. If σ_{R} is the standard deviation, the usual definition of principal components is obtained. As mentioned in Bali *et al.* (2011), if the scale functional σ_{R} is (weakly) continuous, the maximum above is attained.

Assume now that we are dealing with several populations with finite second moment, i.e., $\mathbb{E}\|X_{i,1}\|^2 < \infty$, and that the scale functional is the standard deviation. Then, $\sigma_{\mathrm{R}}^2 \left(P_i[\alpha] \right) = \langle \alpha, \mathbf{\Gamma}_i \alpha \rangle$ with $\mathbf{\Gamma}_i$ the covariance operator of the $i-$th population. Then, under a FCPC model, (1) holds, so $\phi_{\mathrm{R},j}(P_i) = \phi_j$ for all $j \ge 1$ and $1 \le i \le k$. This property allows to extend the definition of a FCPC model to the situation in which the covariance operator does not exist, as follows.

Definition 2.1.

- *We will say that $P_1, \dots, P_k$ are weakly–FCPC for the scale functional σ_{R} if $\phi_{\mathrm{R},j}(P_i) = \phi_{\mathrm{R},j}(P_m)$ (except for a sign change) for all $j \ge 1$ and $1 \le i, m \le k$.*
- *We will say that $P_1, \dots, P_k$ are strongly–FCPC if there exist constants $c_i > 0$ and self–adjoint, positive semidefinite and compact operators $\mathbf{\Gamma}_i$ such that for any $\alpha \in \mathcal{H}$, $\sigma_{\mathrm{R}}^2(P_i[\alpha]) = c_i \langle \alpha, \mathbf{\Gamma}_i \alpha \rangle$, where $\mathbf{\Gamma}_i$ satisfies (1), with $\lambda_{i,1} \ge \lambda_{i,2} \ge \cdots \ge \lambda_{i,j} \ge \lambda_{i,j+1} \cdots$, for $1 \le i \le k$. Hence, $\mathbf{\Gamma}_i$ have the same eigenfunctions and the order among eigenvalues is preserved along populations.*

Clearly, strong–FCPC implies weak–FCPC. Moreover, if second moment exists and σ_{R}^2 is the variance, $P_1, \dots, P_k$ are strongly–FCPC when the covariance operators satisfy (1).

Elliptical distributions were defined in defined in Bali and Boente (2009). For the sake of completeness, we recall their definition. Let X be a random element in a separable Hilbert space $\mathcal{H}$ and $\mu \in \mathcal{H}$. Let $\mathbf{\Gamma} : \mathcal{H} \to \mathcal{H}$ be a self–adjoint, positive semidefinite and compact operator. We will say that X has an elliptical distribution with parameters $(\mu, \mathbf{\Gamma})$, denoted as $X \sim \mathcal{E}(\mu, \mathbf{\Gamma})$, 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 $A\mathbf{\Gamma}A^*$, i.e., $AX \sim \mathcal{E}_d(A\mu, A\mathbf{\Gamma}A^*)$, where $A^* : \mathbb{R}^p \to \mathcal{H}$ stands for the adjoint operator of A. As in the finite–dimensional setting, if the covariance operator, $\mathbf{\Gamma}_X$, of X exists then, $\mathbf{\Gamma}_X = a\, \mathbf{\Gamma}$, for some $a \in \mathbb{R}$. Elliptical distributions in $\mathcal{H}$ include the Gaussian distributions, while other elliptical distributions can be obtained as mixtures of Gaussian processes.

Recall that if $X \sim P = \mathcal{E}(\mu, \mathbf{\Gamma})$, then $\sigma^2_{\mathrm{R}}(P[\alpha]) = c\langle \alpha, \mathbf{\Gamma}\alpha\rangle$ for any scale functional. Thus, when the different samples $X_{i,1}$ are elliptically distributed, i.e., $X_{i,1} \sim \mathcal{E}_d(\mu_i, \mathbf{\Gamma}_i)$ with dispersion operators $\mathbf{\Gamma}_i$ satisfying (1), we have that $P_1, \dots, P_k$ are strongly–FCPC.

The principal directions can be estimated applying a sample version of (2) to each population. However, in most cases, the estimators obtained in such a way will not be equal over populations, even if a FCPC model holds. Hence, a unified approach is needed. Let us define $\sigma_i : \mathcal{H} \to [0, +\infty)$ as $\sigma_i(\alpha) = \sigma_{\mathrm{R}}(P_i[\alpha])$. Note that if $c \in \mathbb{R}$, then using that σ_{R} is a scale functional we get that $\sigma_i^2(c\alpha) = c^2\sigma_i^2(\alpha)$. As in Boente *et al.* (2006), let $f : \mathbb{R}_{>0} \to \mathbb{R}_{\geq 0}$ be a general increasing score function and define $\varsigma_f(\alpha) = \sum_{i=1}^k \tau_i f(\sigma_i^2(\alpha))$. Moreover, let P stand for the product measure $P = P_1 \times \ldots \times P_k$.

It is clear that when $P_1, \dots, P_k$ are weakly–FCPC under σ_{R}, then, for any $1 \leq i \leq k$, $\phi_{\mathrm{R},1}(P_i)$ will maximize $\sum_{i=1}^k \tau_i \sigma^2_{\mathrm{R}}(P_i[\alpha])$ over $\mathcal{S}_1 = \{\alpha \in \mathcal{H} : \|\alpha\| = 1\}$. More generally, it will maximize $\varsigma_f(\alpha)$ over $\mathcal{S}_1$. This motivates to define the common directions projection–pursuit functional as

$$\begin{cases} \phi_{f,1}(P) &= \underset{\|\alpha\|=1}{\operatorname{argmax}}\, \varsigma_f(\alpha) = \underset{\|\alpha\|=1}{\operatorname{argmax}} \sum_{i=1}^k \tau_i f(\sigma^2_{\mathrm{R}}(P_i[\alpha])) \\ \phi_{f,m}(P) &= \underset{\|\alpha\|=1, \alpha \in \mathcal{B}_{f,m}}{\operatorname{argmax}} \varsigma_f(\alpha) = \underset{\|\alpha\|=1, \alpha \in \mathcal{B}_{f,m}}{\operatorname{argmax}} \sum_{i=1}^k \tau_i f(\sigma^2_{\mathrm{R}}(P_i[\alpha])) \quad 2 \leq m\,, \end{cases} \tag{3}$$

where $\mathcal{B}_{f,m} = \{\alpha \in \mathcal{H} : \langle \alpha, \phi_{f,j}(P)\rangle = 0,\ 1 \leq j \leq m-1\}$. We also define the robust principal values functionals as

$$\lambda_{f,i,m}(P) = \sigma^2_{\mathrm{R}}(P_i[\phi_{f,m}(P)]) = \sigma_i^2(\phi_{f,m}(P))\,. \tag{4}$$

Among others, the identity function id or the log can be chosen as score functions f. When $f = \mathrm{id}$ and σ^2_{R} is the variance, the functionals defined in (3) correspond to the eigenfunctions of the pooled covariance operator whose sample version was studied in Boente *et al.* (2010). On the other hand, the function $f = \log$ leads, in the multivariate setting, to the maximum likelihood estimators when considering the sample variance. When considering robust scale estimators, the choice $f = \log$ was recommended in Boente *et al.* (2006) for multivariate observations, based on their simulation results and on the fact that, under a proportional model, the related estimators maximize the asymptotic variance of the common principal directions over the class of strictly increasing twice continuously differentiable score functions f, for a given choice of σ_{R}.

The following two lemmas justify the definition given in (3) since they show that these functionals are properly defined.

Lemma 2.1. *If $f : [0, \infty] \to \mathbb{R}$ is a continuous function and σ_i is weakly continuous, then $\sup_{\|\alpha\|=1} \varsigma_f(\alpha)$ is reached for some $\alpha \in \mathcal{S}_1$ and so the functional will be well defined.*

Similar arguments to those considered in the proof of Lemma 2.1 allow to show that the conclusion of Lemma 2.1 still holds when considering $\sup_{\|\alpha\|=1,\alpha\in\mathcal{B}_{f,m}} \varsigma_f(\alpha)$.

The following Lemma ensures the existence of $\phi_{f,1}(P)$ when $f = \log$.

Lemma 2.2. *If $f = \log$, σ_i is a weakly-continuous function and there exists α_0 such that $\sigma_i(\alpha_0) > 0$ for all i, then $\sup_{\|\alpha\|=1} \varsigma_f(\alpha)$ will be reached for some $\alpha \in \mathcal{S}_1$ and so the functional is well defined.*

As above, the same ideas used in the proof can be considered to obtain that $\sup_{\|\alpha\|=1,\alpha\in\mathcal{B}_{f,m}} \varsigma_f(\alpha)$ is attained if there exists $\alpha \in \mathcal{B}_{f,m}$ such that $\sigma_i(\alpha) > 0$ for all i.

2.2 Fisher-Consistency

The following Lemma show that, if $P_1, \dots, P_k$ are weakly–FCPC, the weights τ_i and the score function f do not play a major role when defining the functional $\phi_{f,j}(P)$.

Lemma 2.3. *Assume that $\tau_i \geq 0$, $\sum_{i=1}^k \tau_i = 1$, $f : \mathbb{R} \to \mathbb{R}$ is an strictly increasing function and that $P_1, \dots, P_k$ are weakly–FCPC under σ_{R}. Then, $\phi_{f,j}(P) = \phi_{\mathrm{R},j}(P_1)$.*

It is worth noting that the above result does not ensure uniqueness of the solution of (3) which will be a condition needed to ensure consistency of the estimators to be defined below. As in the one–population setting, one important issue is what the functions $\phi_{f,m}$ represent, at least in some particular situations. Lemma 2.4 below shows that, for functional elliptical families, the functionals $\phi_{f,m}(P)$ and $\lambda_{f,i,m}(P)$ are well defined, that is, the solution of (3) is unique, and have a simple interpretation. In particular, our result holds if all the populations have an elliptical distribution, but is not restricted to them. Fisher-consistency of the functionals defined through (3) will be obtained under the following assumption

A1. There exists a constant $c_i > 0$ and a self–adjoint, positive semidefinite and compact operator $\mathbf{\Gamma}_{i,0}$, such that for any α, we have $\sigma_i^2(\alpha) = c_i \langle \alpha, \mathbf{\Gamma}_{i,0} \alpha \rangle$.

Note that **A1** entails that the function $\sigma_i : \mathcal{H} \to \mathbb{R}$ defined as $\sigma_i(\alpha) = \sigma_{\mathrm{R}}(P_i[\alpha])$ is weakly continuous.

Lemma 2.4. *Let $\phi_{f,m}$ and $\lambda_{f,i,m}$ be the functionals defined in (2) and (4), respectively. Let $X_{i,1} \sim P_i$ be random elements such that **A1** holds. Assume that $\mathbf{\Gamma}_{i,0}$ have the same eigenfunctions for any i. Moreover, assume that $\mathbf{\Gamma}_{i,0}$ satisfies (1). Denote by $\lambda_{i,j}$ the eigenvalue of $\mathbf{\Gamma}_{i,0}$ related to the eigenfunction ϕ_j, such that $\lambda_{i,1} \geq \lambda_{i,2} \geq \dots$. Assume that for some $1 \leq i_0 \leq k$ there exists $q \geq 2$ such that for all $1 \leq j \leq q$, $\lambda_{i_0,1} > \lambda_{i_0,2} > \dots > \lambda_{i_0,q} > \lambda_{i,q+1}$. Then, if f is an strictly increasing function and $\tau_{i_0} > 0$, we have that, for all $1 \leq j \leq q$, $\phi_{f,j}(P) = \phi_j$ and $\lambda_{f,j}(P) = c_i \lambda_{i,j}$.*

If $\mathbf{\Gamma}_{i,0}$ in **A1** is the covariance operator of P_i, then the eigenfunctions functionals $\phi_{f,m}$ are the common principal components. Besides, we also have that $\lambda_{f,j} = \sigma_i^2(\phi_{f,j}) = c_i \lambda_{i,j}$ where $\lambda_{i,j}$ is the $j-$eigenvalue of the covariance operator of $i-$th population, that is, the traditional principal value in the classical approach. Therefore, the robust eigenvalue functional will be Fisher-consistent except by multiplying factor c_i that can be chosen to be equal to 1 for all populations under a common central Gaussian model to ensure Fisher-consistency of the robust eigenvalue functionals.

3 The estimators

Let $X_{i,1}, \cdots, X_{i,n_i}$ in $\mathcal{H}$ be independent observations from k independent populations, that is, $X_{i,j}$ are independent and such that $X_{i,j} \sim P_i$. Denote $N = \sum_{i=1}^k n_i$ and $\widehat{\tau}_i = n_i/N$. We will assume that $\widehat{\tau}_i \to \tau_i$ with $0 < \tau_i < 1$, for $1 \le i \le k$, and $\sum_{i=1}^k \tau_i = 1$.

To define the estimators of the common functional principal directions, define the empirical version of σ_i^2, $s_{i,n_i}^2 : \mathcal{H} \to \mathbb{R}$ as $s_{i,n_i}^2(\alpha) = \sigma_{\mathrm{R}}^2(P_{i,n_i}[\alpha])$ and the estimators of $\varsigma_f(\alpha)$, $\widehat{\varsigma} : \mathcal{H} \to \mathbb{R}$, as $\widehat{\varsigma}(\alpha) = \sum_{i=1}^k \widehat{\tau}_i f(s_{i,n_i}^2(\alpha))$ with $\widehat{\tau}_i = n_i/N$.

3.1 General robust raw projection-pursuit estimators

As in Boente *et al.* (2006), the general raw projection–pursuit functional common direction estimators are defined as

$$\begin{cases} \widehat{\phi}_1 & = \underset{\|\alpha\|=1}{\operatorname{argmax}} \displaystyle\sum_{i=1}^k \widehat{\tau}_i f(s_{i,n_i}^2(\alpha)) = \underset{\|\alpha\|=1}{\operatorname{argmax}}\, \widehat{\varsigma}(\alpha) \\ \widehat{\phi}_m & = \underset{\|\alpha\|=1, \alpha \in \widehat{\mathcal{B}}_{f,m}}{\operatorname{argmax}} \displaystyle\sum_{i=1}^k \widehat{\tau}_i f(s_{i,n_i}^2(\alpha)) = \underset{\|\alpha\|=1, \alpha \in \widehat{\mathcal{B}}_{f,m}}{\operatorname{argmax}} \widehat{\varsigma}(\alpha) \quad 2 \le m \,, \end{cases} \tag{5}$$

where $\widehat{\mathcal{B}}_m = \{\alpha \in \mathcal{H} : \langle \alpha, \widehat{\phi}_j \rangle = 0, \ 1 \le j \le m-1\}$ while the estimators of their size in the $i-$th population are defined as $\widehat{\lambda}_{i,m} = s_{i,n_i}^2(\widehat{\phi}_m)$.

3.2 Smoothed robust common principal direction estimators

As discussed in Bali *et al.* (2011), sometimes the practiser is interested in smoothed common principal directions. When considering just one–population, the advantages of smoothed functional PCA are well documented, see for instance, Rice and Silverman (1991) and Ramsay and Silverman (2005). The same arguments apply to the case of several populations. For the one–population setting, Rice and Silverman (1991) and Silverman (1996) proposed two smoothing approaches by penalizing the variance and the norm, respectively. These approaches will be extended here to several populations. In this sense, our results will cover not only a robust approach to estimate the common principal directions, but they also provide results for the classical common principal component estimators obtained through a penalization approach, which have not been considered up to now.

We will state the definition in a separable Hilbert space $\mathcal{H}$ keeping in mind that the main applications correspond to $\mathcal{H} = L^2(\mathcal{I})$ with $\mathcal{I} \subset \mathbb{R}$ a finite interval. Let us consider $\mathcal{H}_{\mathrm{S}} \subset \mathcal{H}$ the subset of "smooth elements"of $\mathcal{H}$. In order to obtain consistency results, we will need that $\phi_{\mathrm{R},j}(P) \in \mathcal{H}_{\mathrm{S}}$. Let $D : \mathcal{H}_{\mathrm{S}} \to \mathcal{H}$, a linear operator that we will call the "differentiator". Using D, we will define the symmetric positive semidefinite bilinear form $\lceil \cdot, \cdot \rceil : \mathcal{H}_{\mathrm{S}} \times \mathcal{H}_{\mathrm{S}} \to \mathbb{R}$, $\lceil \alpha, \beta \rceil = \langle D\alpha, D\beta \rangle$. The "penalization operator"is then defined as $\Psi : \mathcal{H}_{\mathrm{S}} \to \mathbb{R}$, $\Psi(\alpha) = \lceil \alpha, \alpha \rceil$. Moreover, define as above, the penalized inner product $\langle \alpha, \beta \rangle_\nu = \langle \alpha, \beta \rangle + \nu \lceil \alpha, \beta \rceil$. Then, $\|\alpha\|_\nu^2 = \|\alpha\|^2 + \nu \Psi(\alpha)$. As in Pezzulli and Silverman (1993), we will assume that the bilinear form is closable.

The smoothed general robust functional common principal components estimators are defined either

a) by penalizing the pooled transformed scales as

$$\begin{cases} \widehat{\phi}_{\mathrm{S},1} & = \operatorname*{argmax}_{\|\alpha\|=1} \left\{ \sum_{i=1}^{k} \widehat{\tau}_i f(s^2_{i,n_i}(\alpha)) - \rho\lceil \alpha, \alpha \rceil \right\} = \operatorname*{argmax}_{\|\alpha\|=1} \widehat{\varsigma}(\alpha) - \rho\Psi(\alpha) \\ \widehat{\phi}_{\mathrm{S},m} & = \operatorname*{argmax}_{\alpha \in \widehat{\mathcal{B}}_{\mathrm{S},m}} \left\{ \sum_{i=1}^{k} \widehat{\tau}_i f(s^2_{i,n_i}(\alpha)) - \rho\lceil \alpha, \alpha \rceil \right\} = \operatorname*{argmax}_{\alpha \in \widehat{\mathcal{B}}_{\mathrm{S},m}} \widehat{\varsigma}(\alpha) - \rho\Psi(\alpha) \quad 2 \le m \end{cases} \tag{6}$$

where $\widehat{\mathcal{B}}_{\mathrm{S},m} = \{\alpha \in \mathcal{H} : \ \|\alpha\| = 1, \langle \alpha, \widehat{\phi}_{\mathrm{S},j} \rangle = 0 \ , \ \forall\, 1 \le j \le m-1\}$.

b) or by penalizing the norm as

$$\begin{cases} \widehat{\phi}_{\mathrm{PN},1} & = \operatorname*{argmax}_{\|\alpha\|_\nu=1} \sum_{i=1}^{k} \widehat{\tau}_i f(s^2_{i,n_i}(\alpha)) = \operatorname*{argmax}_{\|\alpha\|_\nu=1} \widehat{\varsigma}(\alpha) \\ \widehat{\phi}_{\mathrm{PN},m} & = \operatorname*{argmax}_{\alpha \in \widehat{\mathcal{B}}_{\mathrm{PN},m,\nu}} \sum_{i=1}^{k} \widehat{\tau}_i f(s^2_{i,n_i}(\alpha)) = \operatorname*{argmax}_{\alpha \in \widehat{\mathcal{B}}_{\mathrm{PN},m,\nu}} \widehat{\varsigma}(\alpha) \quad 2 \le m \end{cases} \tag{7}$$

where $\widehat{\mathcal{B}}_{\mathrm{PN},m,\nu} = \{\alpha \in \mathcal{H} : \ \|\alpha\|_\nu = 1, \langle \alpha, \widehat{\phi}_{\mathrm{PN},j} \rangle_\nu = 0 \ , \ \forall\, 1 \le j \le m-1\}$.

The eigenvalue estimators are thus defined as

$$\widehat{\lambda}_{\mathrm{S},i,m} = s^2_{i,n_i}(\widehat{\phi}_{\mathrm{S},m})\,, \tag{8}$$

$$\widehat{\lambda}_{\mathrm{PN},i,m} = s^2_{i,n_i}(\widehat{\phi}_{\mathrm{PN},m})\,. \tag{9}$$

To help formulate a unified approach to the different estimators considered in sections 3.1 and 3.2, let the products $\rho\Psi(\alpha)$ or $\nu\Psi(\alpha)$ be defined as 0 when $\rho = 0$ or $\nu = 0$ respectively, even when $\alpha \notin \mathcal{H}_{\mathrm{S}}$ for which case $\Psi(\alpha) = \infty$. All the projection pursuit estimators considered can be viewed as special cases of the following general class of estimators.

$$\begin{cases} \widehat{\phi}_{f,1} & = \operatorname*{argmax}_{\|\alpha\|_\nu=1} \{\widehat{\varsigma}(\alpha) - \rho\Psi(\alpha)\} \\ \widehat{\phi}_{f,m} & = \operatorname*{argmax}_{\alpha \in \widehat{\mathcal{B}}_{f,m,\nu}} \{\widehat{\varsigma}(\alpha) - \rho\Psi(\alpha)\} \quad 2 \le m, \end{cases} \tag{10}$$

where $\widehat{\mathcal{B}}_{f,m,\nu} = \{\alpha \in \mathcal{H} : \ \|\alpha\|_\nu = 1, \langle \alpha, \widehat{\phi}_{f,j} \rangle_\nu = 0 \ , \ \forall\, 1 \le j \le m-1\}$.

With this definition the raw estimators are obtained when $\rho = \nu = 0$, while $\widehat{\phi}_{\mathrm{PN},m}$ and $\widehat{\phi}_{\mathrm{PS},m}$ correspond to $\rho = 0$ and $\nu = 0$, respectively.

4 Consistency

As mentioned above, in the finite–dimensional case, if dispersion operators are proportional, that is, under the second level of hierarchy defined by Flury (1984), the score function $f = \log$ minimizes the asymptotic variance over a family of functions, we refer to Boente *et al.* (2006) for details. The main disadvantage of log is that $\varsigma_f(\alpha)$ and $\widehat{\varsigma}(\alpha)$ are not defined when $\alpha = 0$. Moreover, they will not be weakly continuous in $\mathcal{H}$ due to the singularity at $\alpha = 0$. For that reason, most of the statements and proofs are given separately considering on one side, the case of a continuous function $f : [0, \infty) \to \mathbb{R}$ and on the other one, the logarithm.

To derive consistency results for the estimators defined in Section 3, we will consider the following set of assumptions.

C0. For some $q \geq 2$ and $1 \leq j \leq q$, $\phi_{f,j}(P)$ are unique up to a sign change where $P = P_1 \times \ldots \times P_k$.

C1. $\sigma_i : \mathcal{H} \to \mathbb{R}$ is a weakly continuous function, i.e., continuous with respect to the weak topology in $\mathcal{H}$.

C2. $f : [0, +\infty) \to \mathbb{R}$ is an strictly increasing and continuous function.

C3. $\sup_{\|\alpha\|=1} \left| s^2_{i,n_i}(\alpha) - \sigma^2_i(\alpha) \right| \xrightarrow{a.s.} 0$, for any $1 \leq i \leq k$.

C4. $\widehat{\tau}_i \longrightarrow \tau_i$.

C5. $\sup_{\|\alpha\|=1} |\widehat{\varsigma}(\alpha) - \varsigma_f(\alpha)| \xrightarrow{a.s.} 0$.

It is clear that **C0** holds if for some $1 \leq i \leq k$, $\lambda_{f,i,1} > \lambda_{f,i,2} > \ldots > \lambda_{f,i,q} > \lambda_{f,i,q+1}$. On the other hand, if **C0** holds then, for any $1 \leq \ell \leq q$, there exists $1 \leq i = i_\ell \leq k$ such that $\lambda_{f,i,\ell} > \lambda_{f,i,\ell+1}$.

Remark 4.1 It is worth noticing that **C1** and **C2** imply that $\varsigma_f : \mathcal{H} \to \mathbb{R}$ is a weakly continuous function. Note also that **C1** hold if the univariate scale functional σ_{R} is qualitatively robust, that is, continuous with respect to the weak topology on the space of probability measures, which is induced by the Prohorov distance. Nevertheless, this is not strictly necessary. For instance, if the scale functional satisfies **A1**, as is the case when σ_{R} is the standard deviation, we also obtain weak continuity of σ_i. Assumption **C1** also imply that the functional σ^2_i is weakly uniformly continuous in the unit sphere $\mathcal{S}_1$. Besides, assumption **C3** follows from the consistency of the sample covariance operators (see Dauxois *et al.*, 1982), if σ_{R} equals the standard deviation, while for any scale functional σ_{R} continuous with respect to the weak topology, **C3** follows from Theorem 6.2 in Bali *et al.* (2011). Finally, **C2** to **C4** imply **C5**.

For the sake of simplicity denote by $\widehat{\varsigma}_{i,n_i}(\alpha) = f(s^2_{i,n_i}(\alpha))$ and $\varsigma_{f,i}(\alpha) = f(\sigma^2_i(\alpha))$. Moreover, from now on let $o_{\mathrm{a.s.}}(1)$ stand for a term converging to 0 almost surely.

The following lemma will be useful for deriving consistency of the general eigenfunction estimators.

Lemma 4.1. *Let* $P = P_1 \times \ldots \times P_k$, $\phi_{f,m} = \phi_{f,m}(P)$ *and* $\lambda_{f,i,m} = \lambda_{f,i,m}(P)$ *be defined as in (3) and (4) and let* $\widehat{\phi}_m \in \mathcal{V}_1$ *be such that* $\widehat{\phi}_m \neq 0$, $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$ *and* $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle \xrightarrow{a.s.} 0$. *Assume that **C0** to **C2** hold. We have that*

a) If $\varsigma_f(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$, *then,* $\langle \widehat{\phi}_1, \phi_{f,1} \rangle^2 \xrightarrow{a.s.} 1$.

b) Given $2 \leq m \leq q$, *if* $\varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$, $\widehat{\phi}_s \xrightarrow{a.s.} \phi_{f,s}$ *for* $1 \leq s \leq m-1$, *then* $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$ *and so we can choose the sign of* $\widehat{\phi}_m$ *so that* $\widehat{\phi}_m \xrightarrow{a.s.} \phi_{f,m}$.

Similarly, the following lemma will be useful for deriving consistency of the general eigenfunction estimators when using the logarithm as score function. An extra condition on the principal values $\lambda_{f,i,j}$ is needed to avoid singularities.

Lemma 4.2. *Let* $P = P_1 \times \ldots \times P_k$, $\phi_{f,m} = \phi_{f,m}(P)$ *and* $\lambda_{f,i,m} = \lambda_{f,i,m}(P)$ *be defined as in (3) and (4), respectively, with* $f = \log$ *and* $\widehat{\phi}_m \in \mathcal{V}_1$ *be such that* $\widehat{\phi}_m \neq 0$, $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$ *and* $\langle \widehat{\phi}_m, \widehat{\phi}_j \rangle \xrightarrow{a.s.} 0$. *Assume that **C0** and **C1** hold and that, for any* $1 \leq i \leq k$, $\lambda_{f,i,1} > \lambda_{f,i,2} > \ldots > \lambda_{f,i,q} > \lambda_{f,i,q+1}$. *We have that*

a) *If* $\varsigma_f(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$*, then,* $\langle \widehat{\phi}_1, \phi_{f,1}\rangle^2 \xrightarrow{a.s.} 1$.

b) *Given* $2 \le m \le q$*, if* $\varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$, $\widehat{\phi}_s \xrightarrow{a.s.} \phi_{f,s}$ *for* $1 \le s \le m-1$*, then* $\langle \widehat{\phi}_m, \phi_{f,m}\rangle^2 \xrightarrow{a.s.} 1$ *and so we can choose the sign of* $\widehat{\phi}_m$ *so that* $\widehat{\phi}_m \xrightarrow{a.s.} \phi_{f,m}$.

Theorems 4.1 and 4.2 below establish the continuity of the functionals defined in (3) and (4), for general continuous score functions defined at 0 and for $f = \log$, respectively, and hence the asymptotic robustness of the estimators derived from them, as defined in Hampel (1971). This can be seen just by replacing almost sure convergence by convergence in its statement and by taking P_{i,n_i}, $1 \le i \le k$, fixed sequences of probability measures instead of random ones. From Theorem 6.2 in Bali *et al.* (2011), the uniform convergence required in assumption ii) in Theorem 4.2 below holds if σ_{R} is a continuous scale functional when $P_{i,n_i} \xrightarrow{\omega} P_i$.

Theorem 4.1. *Let* P_{i,n_i}, $1 \le i \le k$*, be random sequences of probability measures,* $N = \sum_{i=1}^k n_i$, $\widehat{\tau}_{i,n_i}$ *be random variables such that* $\widehat{\tau}_{i,n_i} \xrightarrow{a.s.} \tau_i$ *with* $0 < \tau_i < 1$, $\sum_{i=1}^k \tau_i = 1$*. Let* $\nu = \nu_N \ge 0$, $\rho = \rho_N \ge 0$ *be random smoothing parameters. Denote by* $\sigma^2_{i,n_i}(\alpha) = \sigma^2_{\mathrm{R}}(P_{i,n_i}[\alpha])$ *and* $\varsigma_N(\alpha) = \sum_{i=1}^k \widehat{\tau}_i f\left(\sigma^2_{i,n_i}(\alpha)\right)$ *with* $f : [0, +\infty) \to \mathbb{R}$*. Define* $\widehat{\lambda}_{i,m} = \sigma^2_{i,n_i}(\widehat{\phi}_m)$ *with*

$$\begin{cases} \widehat{\phi}_1 &= \underset{\|\alpha\|_\nu = 1}{\operatorname{argmax}} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \\ \widehat{\phi}_m &= \underset{\alpha \in \widehat{\mathcal{B}}_{m,\nu}}{\operatorname{argmax}} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \quad 2 \le m, \end{cases}$$

where $\widehat{\mathcal{B}}_{m,\nu} = \{\alpha \in \mathcal{H} : \|\alpha\|_\nu = 1, \langle \alpha, \widehat{\phi}_j\rangle_\nu = 0 \ , \ \forall\ 1 \le j \le m-1\}$*. Let* $P = P_1 \times \ldots \times P_k$ *be a probability measure satisfying* ***C0*** *and* $\phi_{f,m} = \phi_{f,m}(P)$ *and* $\lambda_{f,i,m} = \lambda_{f,i,m}(P)$ *be defined as in (3) and (4), respectively. Assume that*

i) ***C1*** *and* ***C2*** *hold.*

ii) $\sup_{\|\alpha\|=1} \left|\sigma^2_{i,n_i}(\alpha) - \sigma^2_i(\alpha)\right| \xrightarrow{a.s.} 0$.

iii) $\nu_N \xrightarrow{a.s.} 0$ *and* $\rho_N \xrightarrow{a.s.} 0$.

iv) *Moreover, if* $\nu_N > 0$ *or* $\rho_N > 0$*, for all* $N \ge N_0$*, assume that* $\phi_{f,j} \in \mathcal{H}_{\mathrm{S}}$*, i.e.,* $\Psi(\phi_{f,j}) < \infty$*, for all* $1 \le j \le q$.

Then,

a) $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$ *and* $\varsigma_f(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$*. Moreover,* $\rho\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$ *and* $\nu\lceil \widehat{\phi}_1, \widehat{\phi}_1\rceil \xrightarrow{a.s.} 0$*, and so,* $\|\widehat{\phi}_1\| \xrightarrow{a.s.} 1$.

b) $\langle \widehat{\phi}_1, \phi_{f,1}\rangle^2 \xrightarrow{a.s.} 1$ *and* $\widehat{\lambda}_{i,1} \xrightarrow{a.s.} \lambda_{f,i,1}$.

c) *For any* $2 \le m \le q$*, if* $\widehat{\phi}_\ell \xrightarrow{a.s.} \phi_{f,\ell}$, $\nu\Psi(\widehat{\phi}_\ell) \xrightarrow{a.s.} 0$ *and* $\rho\Psi(\widehat{\phi}_\ell) \xrightarrow{a.s.} 0$ *for* $1 \le \ell \le m-1$*, then,* $\varsigma_N(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$ *and* $\varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$*. Moreover,* $\rho\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, $\nu\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ *and so,* $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$.

d) *For* $1 \le m \le q$, $\langle \widehat{\phi}_m, \phi_{f,m}\rangle^2 \xrightarrow{a.s.} 1$ *and* $\widehat{\lambda}_{i,m} \xrightarrow{a.s.} \sigma^2_i(\phi_{f,m})$.

Note that assumption ii) corresponds to **C3** when P_{i,n_i} is the empirical probability measure of the $i-$th population. On the other hand, when $\sigma_{\mathrm{R}}(\cdot)$ is a continuous scale functional, Theorem 6.2 in Bali *et al.* (2011) implies that ii) holds whenever $d_{\mathrm{PR}}(P_{i,n_i}, P_i) \xrightarrow{a.s.} 0$. Moreover, if $\sigma_{\mathrm{R}}(\cdot)$ is a continuous scale functional and P_i satisfy **C0**, Theorem 4.1 entails the continuity of the functionals $\phi_{f,j}(\cdot)$ and $\lambda_{f,i,j}(\cdot)$ at P_i, for $1 \le j \le q$, and so the proposed estimators are qualitatively robust and consistent. In particular, the estimators are robust if the populations are independent each with an elliptical distribution $\mathcal{E}(\mu_1, \mathbf{\Gamma}_1) \times \ldots \times \mathcal{E}(\mu_k, \mathbf{\Gamma}_k)$, as defined in Section 2.2, such that, for some $1 \le i \le k$, the q largest eigenvalues of the operators $\mathbf{\Gamma}_i$ are all distinct.

From Theorem 4.1, we get that the raw estimators of the principal components are consistent, under **C0** to **C5** by taking $\rho = \nu = 0$. Moreover, the smooth estimators (6) and (7) are also consistent if $\phi_{f,j} \in \mathcal{H}_{\mathrm{S}}$, $1 \le j \le q$.

The following Theorem states an analogous result when using the logarithm as score function.

Theorem 4.2. *Let P_{i,n_i}, $1 \le i \le k$, be sequences of probability measures, $N = \sum_{i=1}^k n_i$, $\widehat{\tau}_{i,n_i}$ be random variables such that $\widehat{\tau}_{i,n_i} \xrightarrow{a.s.} \tau_i$ with $0 < \tau_i < 1$, $\sum_{i=1}^k \tau_i = 1$. Let $\nu = \nu_N \ge 0$, $\rho = \rho_N \ge 0$ be random smoothing parameters. Denote by $\sigma^2_{i,n_i}(\alpha) = \sigma^2_{\mathrm{R}}(P_{i,n_i}[\alpha])$ and $\varsigma_N(\alpha) = \sum_{i=1}^k \widehat{\tau}_i f\left(\sigma^2_{i,n_i}(\alpha)\right)$ with $f = \log$ and define $\widehat{\lambda}_{i,m} = \sigma^2_{i,n_i}(\widehat{\phi}_m)$ with*

$$\begin{cases} \widehat{\phi}_1 & = \underset{\|\alpha\|_\nu = 1}{\operatorname{argmax}} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \\ \widehat{\phi}_m & = \underset{\alpha \in \widehat{\mathcal{B}}_{m,\nu}}{\operatorname{argmax}} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \quad 2 \le m, \end{cases}$$

*where $\widehat{\mathcal{B}}_{m,\nu} = \{\alpha \in \mathcal{H} : \ \|\alpha\|_\nu = 1, \langle \alpha, \widehat{\phi}_j \rangle_\nu = 0 \ , \ \forall \ 1 \le j \le m-1\}$. Let $P = P_1 \times \ldots \times P_k$ be a probability measure satisfying **C0** and $\phi_{f,m} = \phi_{f,m}(P)$ and $\lambda_{f,i,m} = \lambda_{f,i,m}(P)$ be defined as in (3) and (4), respectively, with $f = \log$ and assume that, for any $1 \le i \le k$, $\lambda_{f,i,1} > \lambda_{f,i,2} > \ldots > \lambda_{f,i,q} > \lambda_{f,i,q+1}$. Moreover, assume that*

*i) **C1** holds.*

ii) $\sup_{\|\alpha\|=1} \left|\sigma^2_{i,n_i}(\alpha) - \sigma^2_i(\alpha)\right| \xrightarrow{a.s.} 0$.

iii) $\nu_N \xrightarrow{a.s.} 0$ and $\rho_N \xrightarrow{a.s.} 0$.

iv) Moreover, if $\nu_N > 0$ or $\rho_N > 0$, for all $N \ge N_0$, assume that $\phi_{f,j} \in \mathcal{H}_{\mathrm{S}}$, i.e., $\Psi(\phi_{f,j}) < \infty$, for all $1 \le j \le q$.

Then,

a) $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$ and $\varsigma_f(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$. Moreover, $\rho\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$ and $\nu\lceil\widehat{\phi}_1, \widehat{\phi}_1\rceil \xrightarrow{a.s.} 0$, and so, $\|\widehat{\phi}_1\| \xrightarrow{a.s.} 1$.

b) $\langle \widehat{\phi}_1, \phi_{f,1} \rangle^2 \xrightarrow{a.s.} 1$ and $\widehat{\lambda}_{i,1} \xrightarrow{a.s.} \lambda_{f,i,1}$.

c) For any $2 \le m \le q$, if $\widehat{\phi}_\ell \xrightarrow{a.s.} \phi_{f,\ell}$, $\nu\Psi(\widehat{\phi}_\ell) \xrightarrow{a.s.} 0$ and $\rho\Psi(\widehat{\phi}_\ell) \xrightarrow{a.s.} 0$ for $1 \le \ell \le m-1$, then, $\varsigma_N(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$ and $\varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$. Moreover, $\rho\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, $\nu\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ and so, $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$.

d) For $1 \le m \le q$, $\langle \widehat{\phi}_m, \phi_{f,m} \rangle^2 \xrightarrow{a.s.} 1$ and $\widehat{\lambda}_{i,m} \xrightarrow{a.s.} \sigma^2_i(\phi_{f,m})$.

5 A robust cross–validation procedure

As it is well known, the selection of the smoothing parameters is an important practical issue. Usually, L^2 cross–validation is considered to address this problem. However, it has been extensively pointed out that procedures based on L^2 cross–validation methods are sensitive to outliers even if they are combined with a robust estimation method. When considering just one population, Bali *et al.* (2011) describe a robust cross–validation procedure to choose the penalization parameters when estimating the principal directions using a penalized projection–pursuit method.

In this section, we consider a modified procedure which allows for different degrees of penalization as the common principal directions are obtained. For the sake of simplicity, we describe the procedure only for the robust estimators obtained through (6), that is, those penalizing the pooled transformed scales. In our procedure,once the parameters $\widehat{\rho}_1, \ldots, \widehat{\rho}_{m-1}$ are selected, the estimators of the first $m-1$ common principal directions are kept fixed, that is, we do not recompute the estimators of those directions.

To compute $\widehat{\rho}_1$, the value of the penalizing parameter ρ used to estimate the first common principal direction, we consider the following procedure:

RCV0. Take $Y_{i,j} = X_{i,j}$ and set $m = 1$.

RCV1. Center the data $Y_{i,j}$, that is, define $\widetilde{Y}_{i,j} = Y_{i,j} - \widehat{\mu}_i$, where $\widehat{\mu}_i$ is a robust location estimator of the observations $\{Y_{i,j}, 1 \leq j \leq n_i\}$ from the $i-$th population.

RCV2. Randomly partition the centered data $\{\widetilde{Y}_{i,j}\}$ of each population in K subsets of approximately the same size, with the ℓ-th subset of the population i with size $n_{i,\ell} \geq 2$, $\sum_{\ell=1}^{K} n_{i,\ell} = n_i$. Let $\{\widetilde{Y}_{i,j}^{(\ell)}\}_{1\leq j\leq n_{i,\ell}}$ the elements of the $\ell-$th subset, while $\{\widetilde{Y}_{i,j}^{(-\ell)}\}_{1\leq j\leq n-n_{i,\ell}}$ correspond to the elements in the complement of the ℓ-th subset. The set $\{\widetilde{Y}_{i,j}^{(-\ell)}\}_{1\leq j\leq n-n_{i,\ell}}$ is the training set and $\{\widetilde{Y}_{i,j}^{(\ell)}\}_{1\leq j\leq n_{i,\ell}}$ is the validation one.

RCV3. For each value of ρ in the range of consideration, compute the estimator of the $m-$th direction using only the elements of the training set $\{\widetilde{Y}_{i,j}^{(-\ell)}\}_{1\leq j\leq n-n_{i,\ell}}$. Denote $\widehat{\phi}_{\rho,m}^{(-\ell)}$ this estimator.

RCV4. Using the validation set, define $Y_{i,j}^{(\ell)\perp}(\rho) = \widetilde{Y}_{i,j}^{(\ell)} - \pi_{\widehat{\mathcal{L}}_m^{(-\ell)}}(\widetilde{Y}_{i,j}^{(\ell)})$, $1 \leq j \leq n_{i,\ell}$, where $\widehat{\mathcal{L}}_m^{(-\ell)}$ is the subspace spanned by $\{\widehat{\phi}_{\widehat{\rho}_1,1}, \ldots, \widehat{\phi}_{\widehat{\rho}_{m-1},m-1}, \widehat{\phi}_{\rho,m}^{(-\ell)}\}$ and $\pi_{\mathcal{L}} : \mathcal{H} \to \mathcal{L}$ stands for the orthogonal projection onto the closed linear space $\mathcal{L}$.

RCV5. Given a robust scale estimator around zero σ_n, the robust $K-$th fold cross–validation procedure selects the smoothing parameter as the value $\widehat{\rho}_m$ that minimizes

$$RCV_{\ell,\text{KCV}}(\rho) = \sum_{\ell=1}^{K} \sum_{i=1}^{k} \sigma_n^2(\|X_{i,1}^{(\ell)\perp}(\rho)\|, \ldots, \|X_{i,n_{i,\ell}}^{(\ell)\perp}(\rho)\|) \,.$$

RCV6. Using the value $\widehat{\rho}_m$, the estimator $\widehat{\phi}_{\widehat{\rho}_m,m}$ is obtained as

$$\widehat{\phi}_{\widehat{\rho}_m,m} = \underset{\|\alpha\|=1, \widehat{\mathcal{B}}_{\text{S},m}}{\text{argmax}} \;\widehat{\varsigma}(\alpha) - \widehat{\rho}_m \Psi(\alpha) \,.$$

where $\widehat{\mathcal{B}}_{\text{S},m} = \{\alpha \in \mathcal{H} : \ \|\alpha\| = 1, \langle \alpha, \widehat{\phi}_{\widehat{\rho}_j,j}\rangle = 0 \;, \ \forall \, 1 \leq j \leq m-1\}$.

Once, the parameters $\widehat{\rho}_1, \dots, \widehat{\rho}_{m-1}$ are selected, to select the smoothing parameter for the estimator of the $m-$th common principal direction, as mentioned above, we fix the first $m-1$ estimators of the common principal directions $\widehat{\phi}_{\widehat{\rho}_1,1}, \dots, \widehat{\phi}_{\widehat{\rho}_{m-1},m-1}$ and we project the data over the orthogonal linear space $\widehat{\mathcal{L}}_{m-1}$ spanned by $\widehat{\phi}_{\widehat{\rho}_1,1}, \dots, \widehat{\phi}_{\widehat{\rho}_{m-1},m-1}$. That is, define $Y_{i,j} = \pi_{\widehat{\mathcal{L}}_{m-1}}(X_{i,j})$ and repeat RCV1 to RCV6. Note that since $Y_{i,j} = \pi_{\widehat{\mathcal{L}}_{m-1}}(X_{i,j})$, in RCV4 we only have to project over $\widehat{\phi}^{(-\ell)}_{\rho,m}$.

By a robust measure of scale about zero, we mean that no location estimator is applied to center the data. For instance, in the classical setting, one takes $\sigma_n^2(z_1, \dots, z_n) = (1/n)\sum_{i=1}^n z_i^2$, while in the robust situation, one might choose $\sigma_n(z_1, \dots, z_n) = \text{median}(|z_1|, \dots, |z_n|)$ (see Bali *et al.*, 2011, for details).

6 Monte Carlo study

The algorithm to be used to compute the estimators considered in this paper is a modification of the algorithm proposed by Croux and Ruiz–Gazen (1996) for the computation of principal components using projection-pursuit adapted to the situation of several populations as in Boente *et al.* (2006). We refer to Bali *el al.* (2011) for details on the the algorithm for the one–population case. In our case, the so–called index ξ_n corresponds to $\widehat{\varsigma}$ when considering the raw estimators or those obtained penalizing the norm. On the other hand, when penalizing the objective function the index is $\widehat{\varsigma} - \rho\Psi$. As in Bali *et al.* (2011), in the simulation study, to apply the algorithm to functional data, we discretize the domain of the observed function $X \in L^2(\mathcal{I})$, over $m = 50$ equally spaced points in $\mathcal{I}$. As in Bali *et al.* (2011), the algorithm is adapted to allow for smoothed principal components.

Corresponding to non–resistant and robust estimators of the principal common directions, three scale functions are considered, the classical standard deviation (SD), the Median Absolute Deviation (MAD) and an $M-$estimator of scale ($M-$SCALE). The latter two are robust scale statistics. For the $M-$estimator, we used as score function $\chi_c(y) = \min\left(3\,(y/c)^2 - 3\,(y/c)^4 + (y/c)^6\,, 1\right)$, introduced by Beaton and Tukey (1974), with tuning constant $c = 1.56$ and breakdown point 1/2. To compute the $M-$scale, the initial estimator of scale was the MAD.

For the different procedures to be considered, smooth or raw, the maximization in (5) is performed over a set of candidates which correspond to the observations at hand, centered and normalized to the unit ball. To be more precise, when $j = 1$, the set of candidates correspond to $\mathcal{A} = \{(X_{i,j} - \widehat{\mu}_i)/\|X_{i,j} - \widehat{\mu}_i\|\}$ where $\widehat{\mu}_i$ is taken as the point–to–point mean of each population when the scale equals the SD, while for the robust procedures, $\widehat{\mu}_i$ equals the L^1 or spatial median of each population.

For both the classical and robust procedures a penalization depending on the L^2 norm of the second derivative is included, multiplied by a smoothing factor, that is, $\Psi(\alpha) = \int_{\mathcal{I}}(\alpha''(s))^2 ds$. Note that when ρ or $\nu = 0$, the raw estimators are obtained. We considered fixed penalization parameters and also, when penalizing the pooled transformed scales, data–driven parameters obtained through the robust $K-$fold cross–validation described in Section 5.

In all Figures and Tables, the estimators corresponding to each scale choice are labelled as SD, MAD, $M-$SCALE. For each scale, we considered three estimators, the raw estimators where no smoothing is used, the estimators obtained by penalizing the function $\widehat{\varsigma}$ and those obtained by penalizing the norm. In all Tables, as in Section 3.2, the jth principal direction estimators related to each method are labelled as $\widehat{\phi}_{\text{RAW},j}$, $\widehat{\phi}_{\text{PS},j}$ and $\widehat{\phi}_{\text{PN},j}$, respectively.

Several models were considered in our simulation study, including a finite range model, which is the several population counterpart of the model studied in Bali *et al.* (2011), and an infinite dimensional range model. Also, different relations among populations including a proportional model are studied in the infinite–dimensional case. In all cases, when the smoothing parameter is fixed we performed 1000 replications, while when the penalizing parameter is chosen using the cross–validation procedure described in Section 5, only 100 replications are performed due to the expensive computation time.

For each situation, we compute the estimators of the first three principal directions and the square distance between the true and the estimated direction (normalized to have L^2 norm 1), that is,

$$D_j = \left\| \frac{\widehat{\phi}_j}{\|\widehat{\phi}_j\|} - \phi_j \right\|^2 .$$

Note that all the estimators except those penalizing the norm, are such that $\|\widehat{\phi}_j\| = 1$. Mean values over replications, which hereafter is referred to as mean square error, are reported in the Tables summarizing the results.

6.1 Finite–range study

6.1.1 Description of the model and contaminations

As mentioned above, this model is the three population counterpart of the situation considered in Bali *et al.* (2011). We consider observations from $k = 3$ populations such that $X_{i,\ell} = Z_{i1,\ell}\phi_1 + Z_{i2,\ell}\phi_2 + Z_{i3,\ell}\phi_3$, $1 \le \ell \le n_i$, $1 \le i \le 3$, with $Z_{ij,\ell}$ independent of each other. The functions $\phi_i : [-1, 1] \to \mathbb{R}$ are given by $\phi_1(x) = \sin(4\pi x)$, $\phi_2(x) = \cos(7\pi x)$ and $\phi_3(x) = \cos(15\pi x)$. For all the populations, the sample sizes considered are $n_i = 100$, $1 \le i \le k = 3$, for a total sample size of $n = 300$.

Under the central model, labelled C_0, for $1 \le i \le 3$, $\mathbf{Z}_i = (Z_{i,1}, Z_{i,2}, Z_{i,3})^{\mathrm{T}} \sim N(\mathbf{0}, \mathbf{\Sigma}_i)$, where $\mathbf{\Sigma}_i = \mathrm{diag}(\lambda_{i,1}, \lambda_{i,2}, \lambda_{i,3})$ and $\boldsymbol{\lambda}_i = (\lambda_{i,1}, \lambda_{i,2}, \lambda_{i,3})^{\mathrm{T}}$ are given by $\boldsymbol{\lambda}_1 = (16, 4, 1)^{\mathrm{T}}$, $\boldsymbol{\lambda}_2 = (36, 9, 1)^{\mathrm{T}}$, $\boldsymbol{\lambda}_3 = (4, 1, 1/4)^{\mathrm{T}}$.

Several contaminations are considered. In all cases, the contaminated observations are defined as $X^{(c)}_{i,\ell} = Z^{(c)}_{i1,\ell}\phi_1 + Z^{(c)}_{i2,\ell}\phi_2 + Z^{(c)}_{i3,\ell}\phi_3$, $1 \le \ell \le$, $1 \le i \le 3$, where $Z^{(c)}_{ij,\ell}$ are independent of each other. The different contaminations correspond to different choices for the distribution of $Z^{(c)}_{ij,\ell}$.

- $C_{2,\epsilon}$ corresponds to contamination on the second component. In this case, $Z^{(c)}_{i1,\ell} \sim Z_{i1,\ell}$, $Z^{(c)}_{i3,\ell} \sim Z_{i3,\ell}$ and $Z^{(c)}_{i2,j} \sim (1-\epsilon)N(0, \sigma^2_{i2}) + \epsilon N(\mu_{i,2}, (\sigma^{(c)}_{i2})^2)$ where $\sigma^{(c)}_{i2} = 0.1$, $\mu_{1,2} = 10$, $\mu_{2,2} = 15$ and $\mu_{3,2} = 5$. We denote by $P_{C_{2,\epsilon}}$ the joint probability measure of $(X^{(c)}_{1,1}, X^{(c)}_{2,1}, X^{(c)}_{3,1})$ under $C_{2,\epsilon}$.

  The main effect of this contamination is that the directions ϕ_1 and ϕ_2 will be exchanged, when considering the standard deviation, either with $f = \mathrm{id}$ or $f = \log$.

- Two contaminations are considered in the third component.

  - ⋆ The first one, labelled $C_{3,a,\epsilon}$, is a strong contamination on the third component and corresponds to generating $Z^{(c)}_{i1,\ell} \sim Z_{i1,\ell}$, $Z^{(c)}_{i2,\ell} \sim Z_{i2,\ell}$ and $Z^{(c)}_{i3,j} \sim (1-\epsilon)N(0, \sigma^2_{i3}) +$

$\epsilon N(\mu_{i,3}, (\sigma_{i3}^{(c)})^2)$ with $\sigma_{i3}^{(c)} = 0.1$ and $\mu_{1,3} = 15$, $\mu_{2,3} = 20$ and $\mu_{3,3} = 7$. $P_{C_{3,a,\epsilon}}$ stands for the joint distribution of $(X_{1,1}^{(c)}, X_{2,1}^{(c)}, X_{3,1}^{(c)})$ under $C_{3,a,\epsilon}$.

⋆ Under $C_{3,b,\epsilon}$, the observations are such that $Z_{i1,\ell}^{(c)} \sim Z_{i1,\ell}$, $Z_{i2,\ell}^{(c)} \sim Z_{i2,\ell}$ and $Z_{i3,j}^{(c)} \sim (1-\epsilon)N(0, \sigma_{i3}^2) + \epsilon N(\mu_{i,3}, (\sigma_{i3}^{(c)})^2)$ where $\sigma_{i32}^{(c)} = 0.1$, $\mu_{1,3} = 6$, $\mu_{2,3} = 10$ and $\mu_{3,3} = 3$. As above, $P_{C_{3,b,\epsilon}}$ stands for the joint distribution of $(X_{1,1}^{(c)}, X_{2,1}^{(c)}, X_{3,1}^{(c)})$ under $C_{3,b,\epsilon}$.

- $C_{23,\epsilon}$ is a contamination both in the second and third component, according to the contamination level all the directions are modified when using the standard deviation. In this case, $Z_{i1,\ell}^{(c)} \sim Z_{i1,\ell}$, $Z_{i2,j}^{(c)} \sim (1-\epsilon)N(0, \sigma_{i2}^2) + \epsilon N(\mu_{i2}, (\sigma_{i2}^{(c)})^2)$ and $Z_{i3,j}^{(c)} \sim (1-\epsilon)N(0, \sigma_{i3}^2) + \epsilon N(\mu_{i3}, (\sigma_{i3}^{(c)})^2)$ where $\sigma_{i2}^{(c)} = 0.1$, $\sigma_{i3}^{(c)} = 0.1$ and the mean values are $\mu_{1,2} = 10$, $\mu_{2,2} = 15$ and $\mu_{3,2} = 5$ for the second component and $\mu_{1,3} = 15$, $\mu_{2,3} = 20$ and $\mu_{3,3} = 7$ for the third one. $P_{C_{23,\epsilon}}$ is the joint distribution of $(X_{1,1}^{(c)}, X_{2,1}^{(c)}, X_{3,1}^{(c)})$ under $C_{23,\epsilon}$.

Two contamination percentages are considered $\epsilon = 0.1$ and 0.2.

When $f = \text{id}$ and $f = \log$ and the standard deviation is taken as scale functional, the values of $\varsigma_f(\phi_j)$ obtained under C_0 and the different contaminations, are reported in Bali (2012). As described therein, under $C_{2,0.1}$, $\phi_{f,1}(P_{C_{2,0.1}}) = \phi_1$, $\phi_{f,21}(P_{C_{2,0.1}}) = \phi_2$ and $\phi_{f,3}(P_{C_{2,0.1}}) = \phi_3$. Hence, this amount of contamination may not affect the classical estimator, so the results obtained for $C_{2,0.1}$ are not reported. On the other hand, under $P_{C_{23,\epsilon}}$, when $f = \log$, the values of the objective function ς_f at ϕ_j, $1 \le j \le 3$, are very close to each other making difficult the estimation of the directions which may not distinguish between them. The same happens when considering $C_{3,b,0.1}$.

Besides, a penalized functional $\varsigma_{f,\rho}(\alpha) = \sum_{i=1}^{3}(1/3) f(\text{var}(P_i[\alpha])) - \rho\Psi(\alpha)$ is introduced to have an insight of the effect that penalizing the pooled transformed scales may have on the estimators, when contaminating the samples. It is clear that under C_0 for any value of ρ, $\varsigma_{f,\rho}((\phi_1) > \varsigma_{f,\rho}((\phi_2) > \varsigma_{f,\rho}((\phi_3)$, since $\Psi(\phi_1) < \Psi(\phi_2) < \Psi(\phi_3)$.

If we include now the penalization and consider contaminated samples, we need to ensure that the penalization amount ρ is such that the order observed when $\rho = 0$ among the functional common principal directions, $\phi_{f,j}$ is preserved, to observe an effect on the estimation procedure based on the SD, under contamination. Since a high value of ρ may produce that the dominating term is $\Psi(\alpha)$, maximum values for the penalizing amount were computed in Bali (2012) leading to the values reported in Table 1. Hence, values of ρ smaller than $\rho_{\max}$ should be considered.

$f = \text{id}$	$f = \log$	$f = \text{id}$	$f = \log$
$C_{2,0.1}$		$C_{23,0.1}$	
—	—	$4.54*10^{-7}$	$3.07*10^{-8}$
$C_{3,a,0.1}$		$C_{3,b,0.1}$	
$4.54*10^{-7}$	$3.07*10^{-8}$	$7.65*10^{-8}$	$1.17*10^{-8}$
$C_{2,0.2}$		$C_{23,0.2}$	
$1.79*10^{-5}$	$8.74*10^{-7}$	$3.01*10^{-6}$	$1.12*10^{-7}$
$C_{3,a,0.2}$		$C_{3,b,0.2}$	
$3.64*10^{-6}$	$1.44*10^{-7}$	$7.81*10^{-7}$	$1.15*10^{-7}$

Table 1: Values of $\rho_{\max}$ under $C_{2,\epsilon}$, $C_{3,a,\epsilon}$, $C_{3,b,\epsilon}$ and $C_{23,\epsilon}$ for $\epsilon = 0.1$ and 0.2.

6.1.2 Simulation results

When using the penalized estimators, several values for the penalizing parameters ρ and ν were chosen. Since large values of the smoothing parameters make the penalizing term to be the dominant component independently of the amount of contamination considered, we choose in this study ρ and ν equal to $aN^{-\alpha}$ for $\alpha = 2, 3$ and 4 and a equal to 0.1 to 5.5.

Tables 6 to 9, in the supplementary file, report the results obtained for the different estimators when $f = \text{id}$ while Table 10, in the supplementary file, report the results when $f = \log$. Due to the fact that the trajectories are smooth, smoothing does not improve the performance of the estimators. Since our conclusions are similar to those given in the one–population case, we refer to Bali *et al.* (2011) for a deeper discussion.

On the other hand, as reported in the one population case in Bali *et al.* (2011), the robust estimators have a behavior similar to that of the robust ones, while a larger efficiency loss is observed when using the MAD, as in the one–population case.

Tables 11 and 12, in the supplementary file, summarize the behavior of the estimators for the different contaminations considered when $f = \text{id}$ or $f = \log$, respectively. The results for the raw estimators, as well as that of the penalized estimators, when the penalization equals $3n^{-3}$ are reported. As expected, the robust estimators behave much better than the classical ones. Moreover, the simulation study confirms the expected inadequate behavior of the classical estimators in the presence of outliers since they do not estimate the target directions very accurately. The robust estimators of the first common principal direction are not heavily affected by the contaminations considered, except for contamination $C_{23,0.2}$ where the large amount of outliers seems to affect all the robust directions estimators, although much less than when using the classical methods. Probably, this amount of contamination is close to the breakdown point of the proposed estimators. Under $C_{3,a,\epsilon}$ or $C_{3,b,\epsilon}$, the projection–pursuit estimators based on the MAD seem to be more affected than those based on the $M-$scale by this type of contamination, in particular, when $\epsilon = 0.2$. Contamination $C_{2,0.2}$ affects more the first direction robust estimators than $C_{3,a,0.2}$. It is worth noting that, when using $f = \log$, an advantage is observed when some smoothing is introduced. Moreover, even if we are not considering a proportional model, $f = \log$ leads to smaller mean square errors under C_0 for all the scale estimators than $f = \text{id}$ except for the third direction when penalizing the norm. Tables 11 and 12 also reveal the high loss of efficiency for the MAD, in our functional setting even when using $f = \log$ the mean square errors are almost ten times larger than those of the classical procedure based on the SD.

For the procedure which penalizes the function $\widehat{\varsigma}$, the smoothing parameters $\widehat{\rho}_1, \widehat{\rho}_2, \widehat{\rho}_3$ were selected sequentially using the procedure described in Section 5 with $K = 5$ and $f = \text{id}$. As mentioned above, we have only performed 100 replications. The simulation results are reported in Table 2. For the contaminations schemes defined above, we only consider $\epsilon = 0.2$ which corresponds to the worst situation. As when fixing the penalization parameter, no major differences are observed between the penalized and the raw estimators. This fact can be explained by the fact that the trajectories are smooth.

6.2 Infinite–range case

In this case, two situations are considered, a model with three populations and a proportional model with two populations. The latter is considered to see if some advantage is observed when using $f = \log$.

Model	Scale Estimator	$\widehat{\phi}_{\mathrm{RAW},j}$			$\widehat{\phi}_{\mathrm{PS},j}$		
		$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
C_0	SD	0.0033	0.0039	0.0037	0.0031	0.0040	0.0038
	MAD	0.0436	0.0634	0.0368	0.0524	0.0667	0.0339
	M-scale	0.0108	0.0141	0.0077	0.0122	0.0152	0.0074
$C_{2,0.2}$	SD	1.4496	1.4496	0.0023	1.5410	1.5410	0.0020
	MAD	0.3780	0.3903	0.0271	0.3880	0.3972	0.0266
	M-scale	0.4347	0.4354	0.0058	0.4591	0.4604	0.0062
$C_{3,a,0.2}$	SD	1.8275	1.9238	1.9405	1.8253	1.9180	1.9367
	MAD	0.2598	0.7858	0.8071	0.2742	0.7945	0.8090
	M-scale	0.2997	1.0419	1.0977	0.2997	1.1054	1.1594
$C_{3,b,0.2}$	SD	0.0140	1.7596	1.7905	0.0145	1.7478	1.7745
	MAD	0.0874	0.5092	0.5013	0.0855	0.5329	0.5262
	M-scale	0.0427	0.4782	0.4966	0.0384	0.4378	0.4631
$C_{23,0.2}$	SD	1.8303	0.1930	1.8255	1.8196	0.1684	1.8113
	MAD	0.9350	1.0709	0.6005	0.9081	1.0608	0.5985
	M-scale	1.0599	1.2001	0.7009	1.0753	1.1905	0.6969

Table 2: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$ when $f = \mathrm{id}$ and the smoothing parameter is selected using a $K-$fold cross–validation procedure.

6.2.1 Three population model

As in the previous section, we considered $N = \sum_{i=1}^{k} n_i$ observations in $L^2([0,1])$ from $k = 3$ populations, with $n_i = 100$, $1 \le i \le 3$. Under the central model, labelled C_0, all the populations are Gaussian with distribution as follows

- For the first population, $X_{1,\ell} \sim P_1$ where P_1 corresponds to a Brownian motion in the interval $[0,1]$ with covariance kernel $\gamma_1(s,t) = 10\min(s,t)$. This choice of the covariance operator leads to principal directions $\phi_n(t) = \sqrt{2}\sin((2n-1)\pi t/2)$ with related principal values $\lambda_{1,n} = 10\,(2/\{(2n-1)\pi\})^2$.
- The second population is also a Gaussian process but with covariance kernel proportional to the previous one. To be more precise, we choose $\gamma_2(s,t) = 2\gamma_1(s,t)$.
- The third population is a finite–range one, generate as $X_{3,\ell} = Z_{1,\ell}\phi_1 + Z_{2,\ell}\phi_2 + Z_{3,\ell}\phi_3$, where $\phi_n(t) = \sqrt{2}\sin((2n-1)\pi t/2)$, $Z_{k,\ell} \sim N(0,\sigma_k^2)$, with $\sigma_1 = 3$, $\sigma_2 = 1$ and $\sigma_3 = 1/2$. Thus, $\lambda_{3,1} = \sigma_1^2 = 9$, $\lambda_{3,2} = \sigma_2^2 = 1$ and $\lambda_{3,3} = \sigma_3^2 = 1/4$ and $\lambda_{3,j} = 0$ for $j \ge 4$.

Note that the first two populations have continuous but rough trajectories while the third one has smooth trajectories. Hence, among the candidates to be considered in our maximization procedure, we have smooth candidates to approximate the true common principal direction estimators.

Each of the three populations is contaminated with a contaminating distribution highly concentrated on the fourth principal direction. Let us denote $C_{4,\epsilon}$ this contamination, where ϵ corresponds to the contamination level. The contaminated observations denoted $X_{i,j}^{(c)}$ are generated as $X_{i,j}^{(c)} = (1-V_{i,j})X_{i,j} + V_{i,j}W_{i,j}$, where $V_{i,j} \sim Bi(1,\epsilon)$ and $W_{i,j} \sim N(\mu_i,\sigma_c)\phi_4$ independent of $X_{i,j}$ with $\sigma_c = 0.1$, $\mu_1 = 10$, $\mu_2 = 15$ and $\mu_3 = 20$. Denote $P_\epsilon^{(c)}$ the joint distribution of $(X_{1,1}^{(c)}, X_{2,1}^{(c)}, X_{3,1}^{(c)})$. Two values for the proportion of atypical data are considered $\epsilon = 0.1$ and $\epsilon = 0.2$.

In Bali (2012), a detailed description on the effect that contamination has on the functional, when considering the standard deviation, is given to justify the selected values of μ_i. In particular, if σ_{R} is the standard deviation, for the contaminated data we will get that $\varsigma(\phi_4) > \varsigma(\phi_1)$ and so, the functional $\phi_{f,1}(P_\epsilon^{(c)})$ is no longer ϕ_1. Moreover, to evaluate the effect that the penalization may have on the estimators, when penalizing $\widehat{\varsigma}$, a penalized functional is considered and the effect of ρ is studied. The goal is to guarantee that the value of the penalizing parameter will not be chosen so large that its effect on the objective function will make the penalization to dominate over the pooled transformed scale, leading to a smaller value of the objective function at ϕ_4 than at ϕ_1 under $P_\epsilon^{(c)}$. This leads to maximum values for ρ, $\rho_{\max}$, equal to 0.0023 and 0.00013, when $f = \text{id}$ and $f = \log$, respectively. Taking values of ρ smaller than $\rho_{\max}$ ensure that the contamination will effectively have an effect when using the standard deviation.

Simulation results

Table 3 summarize the results of the simulation. The fact that the robust estimators, based on the MAD and $M-$scale, are more resistant under the presence of the contamination model than the classical estimator based on the standard deviation is again confirmed. It is worth noticing that the robust methods are sensitive to 20% of contamination, especially when considering the third direction. Evidently, for this contamination level, we are getting close to the breakdown point of the estimator due to the closeness of the eigenvalues. An approach to the computation of the breakdown point, in the finite–dimensional case, was given by Boente and Orellana (2001). Nevertheless, in the case of functional data the problem is more complex and is beyond the scope of the paper. However, in the one–population setting, it is well known that the sensitivity of the robust estimators is related to the relative size of the eigenvalues. On the other hand, when dealing with several populations and $f = \text{id}$, the sensitivity of the robust estimators to a given contamination is related to the relative size of the eigenvalues of the pooled covariance operator $\sum_{i=1}^{3} \tau_i \mathbf{\Gamma}_i$. When considering $f = \log$, in the finite–dimensional case, Boente *et al.* (2006) showed that the performance of the robust estimators is related to the closeness of the eigenvalues of the matrix $\mathbf{\Sigma}_{\ln} = \log \left(\prod_{i=1}^{k} \mathbf{\Sigma}_i^{\tau_i} \right)$, where $\mathbf{\Sigma}_i$ stands for the covariance matrix of the $i-$th population. These eigenvalues are related to the relative size of $\varsigma_f(\phi_j)$ and for that reason, using $f = \log$ leads to slightly better results. This closeness was also pointed out in Section 6.1, since in that case, the estimation procedure has more trouble to identify the correct directions.

The results in Table 3 show that the function f plays a relevant role. It can be seen that, in general, the performance of the estimator is better when we use $f = \log$ than when $f = \text{id}$ is considered. Besides, a small improvement is observed when penalizing either the norm or the function $\widehat{\varsigma}$, in particular, when using the MAD as scale estimator. However, this improvement is smaller than that obtained if the logarithm function is considered in the estimation procedure. As mentioned above, the third population has smooth trajectories, so the inclusion of a penalization term will tend to improve the performance of the estimators.

6.3 Proportional model

To analyse the effect of having smooth or rough trajectories on the algorithm leading to the estimation procedure, we consider two different situations for a proportional model. For that purpose, the uncontaminated observations correspond to Gaussian processes being either a Wiener or a Ornstein–Uhlenbeck process. We generated $N = n_1 + n_2$ observations in $L^2([0, 1])$ from $k = 2$ populations, with $n_1 = n_2 = 100$.

For the uncontaminated observations, labelled C_0, the proportionality constant was equal to 10. To be more precise, we considered the models

Model	Scale Estimator	$\phi_{\mathrm{RAW},j}$		$\phi_{\mathrm{PS},j}$ ($\rho = 10^{-7}$)		$\phi_{\mathrm{PN},j}$ ($\nu = 10^{-7}$)	
		$j = 1$					
		$f = \mathrm{id}$	$f = \log$	$f = \mathrm{id}$	$f = \log$	$f = \mathrm{id}$	$f = \log$
C_0	SD	0.0037	0.0036	0.0037	0.0037	0.0037	0.0037
	MAD	0.0279	0.0279	0.0274	0.0268	0.0262	0.0268
	M-scale	0.0061	0.0065	0.0061	0.0065	0.0061	0.0065
$C_{4,0.1}$	SD	1.9274	1.9050	1.9274	1.9058	1.9265	1.9058
	MAD	0.0870	0.0845	0.0844	0.0726	0.0743	0.0727
	M-scale	0.0934	0.0897	0.0896	0.0778	0.0790	0.0781
$C_{4,0.2}$	SD	1.9475	1.9418	1.9475	1.9406	1.9469	1.9406
	MAD	0.1946	0.1864	0.1932	0.1780	0.1853	0.1782
	M-scale	0.2108	0.2003	0.2096	0.1945	0.2031	0.1946
		$j = 2$					
C_0	SD	0.0043	0.0043	0.0043	0.0045	0.0045	0.0045
	MAD	0.0585	0.0576	0.0551	0.0562	0.0550	0.0565
	M-scale	0.0130	0.0130	0.0130	0.0131	0.0132	0.0131
$C_{4,0.1}$	SD	1.9298	1.9291	1.9284	1.9271	1.9252	1.9271
	MAD	0.2588	0.2360	0.2105	0.1763	0.2317	0.2154
	M-scale	0.2635	0.2292	0.2121	0.1768	0.2394	0.2181
$C_{4,0.2}$	SD	1.9308	1.9295	1.9292	1.9264	1.9260	1.9264
	MAD	0.7966	0.7280	0.7551	0.6460	0.7761	0.7133
	M-scale	0.9394	0.8511	0.8974	0.7649	0.9012	0.8259
		$j = 3$					
C_0	SD	0.0036	0.0036	0.0035	0.0036	0.0036	0.0036
	MAD	0.0752	0.0672	0.0440	0.0460	0.0523	0.0520
	M-scale	0.0109	0.0101	0.0103	0.0100	0.0105	0.0100
$C_{4,0.1}$	SD	1.9209	1.9251	1.9206	1.9239	1.9229	1.9242
	MAD	0.8667	0.8521	0.7169	0.6329	0.7594	0.7443
	M-scale	1.1009	1.0773	0.9432	0.8297	0.9607	0.9474
$C_{4,0.2}$	SD	1.9174	1.9222	1.9164	1.9230	1.9181	1.9231
	MAD	1.5080	1.4755	1.4697	1.4103	1.4467	1.4362
	M-scale	1.6430	1.6300	1.6229	1.5674	1.6009	1.5873

Table 3: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$ for the three population model with infinite range.

- **Model 1**: Corresponds to a Brownian motion. The observations $X_{i,j}$, $1 \leq j \leq n_i$ correspond to Gaussian processes with covariance kernels $\gamma_1(s,t) = 10\min(s,t)$ and $\gamma_2(s,t) = 10\gamma_1(s,t)$. This model will be labelled BM in the Tables.

- **Model 2**: Corresponds to a Ornstein–Uhlenbeck process. In this case, the observations $X_{i,j}$, $1 \leq j \leq n_i$ are Gaussian with covariance kernels $\gamma_2(s,t) = (1/2)(1/2)^{0.9(s-t)^2}$ and $\gamma_2(s,t) = 10\gamma_1(s,t)$. This model will be labelled OU in the Tables.

For each model, a contamination in the fourth eigenfunction was considered, as follows. Let ϕ_j stand for the eigenfunctions of the covariance operator $\mathbf{\Gamma}_1$ related to the covariance kernel $\gamma_1(s,t)$. As above, denote $C_{4,\epsilon}$ this contamination, where ϵ corresponds to the contamination level. The contaminated observations denoted $X_{i,j}^{(c)}$ are generated as $X_{i,j}^{(c)} = (1 - V_{i,j})X_{i,j} + V_{i,j}W_{i,j}$, where $V_{i,j} \sim Bi(1,\epsilon)$ and $W_{i,j} \sim N(\mu_i, \sigma_c)\phi_4$ independent of $X_{i,j}$ with $\sigma_c = 0.1$, $\mu_1 = 10$, $\mu_2 = 30$. Two values for the proportion of atypical data are considered $\epsilon = 0.05$ and $\epsilon = 0.1$.

The results for the raw estimators of the first three common principal directions are reported in Table 4. For the uncontaminated samples, the advantage of using $f = \log$ can be appreciated. In the multivariate setting, the better performance of the estimators computed with the logarithm function can be explained since, as mentioned above, when σ_{R} is the standard deviation, they lead to the maximum likelihood estimators. Besides, for any fixed scale σ_{R}, estimators obtained using

$f = \log$ maximize the asymptotic variance of the common principal directions over the class of strictly increasing twice continuously differentiable score functions f (see Boente *et al.*, 2006). Our simulation results shows that the same improvement is obtained in the functional setting.

For contaminated samples, the procedures based on the standard deviation breakdown, especially when considering **Model 2** or the second and third components. The robust procedures are more stable, in particular, they lead to reliable results when estimating the first two common principal directions. This performance is much better for **Model 2** which has smooth trajectories, since the estimators are smoother in this case. On the other hand, when $\epsilon = 0.1$, the amount of outliers affect the robust estimators of the third common principal direction, although much less than when using the classical methods. It is worth noting that, when considering the robust scales, even if the contaminated samples do not follow a proportional model, choosing $f = \log$ leads to smaller mean square errors that those obtained with $f = \text{id}$ in most cases. In this sense, using a robust scale combined with the log seem to be the better choice.

Table 5 reports the results obtained when considering the penalized estimators $\widehat{\phi}_{\text{PS},j}$ obtained penalizing $\widehat{\varsigma}$. The smoothing parameters $\widehat{\rho}_1, \widehat{\rho}_2, \widehat{\rho}_3$ were selected using the $K-$th fold procedure described in Section 5 with $K = 5$. Again, under **Model 2**, penalizing the trajectories do not improve the behavior of the estimators under C_0 since the data are already smooth. On the other hand, a benefit is observed when using the robust scales, under **Model 1**. Note that the mean square errors obtained , in particular under $C_{4,0.1}$, suggest that some penalization combined with a robust scale is the recommended option even if the trajectories are smooth, since a reduction with respect to the mean square errors obtained for the raw estimators is obtained even when considering the Ornstein process.

	Scale	$\widehat{\phi}_{\text{RAW},1}$		$\widehat{\phi}_{\text{RAW},2}$		$\widehat{\phi}_{\text{RAW},3}$	
		$f = \text{id}$	$f = \log$	$f = \text{id}$	$f = \log$	$f = \text{id}$	$f = \log$
		BM					
C_0	SD	0.0143	0.0138	0.0956	0.0914	0.2483	0.2329
	MAD	0.0514	0.0407	0.2496	0.1929	0.5786	0.4619
	$M-$scale	0.0188	0.0167	0.1254	0.1095	0.3207	0.2834
$C_{4,0.05}$	SD	0.9135	0.9303	1.9097	1.9045	1.7811	1.8049
	MAD	0.0676	0.0588	0.3564	0.3011	0.8252	0.7757
	$M-$scale	0.0441	0.0372	0.2486	0.2170	0.7857	0.7130
$C_{4,0.1}$	SD	1.7338	1.8015	1.9111	1.9172	1.7865	1.8020
	MAD	0.1037	0.0941	0.5680	0.5092	1.1108	1.1202
	$M-$scale	0.0919	0.0895	0.5684	0.5499	1.1866	1.2269
		OU					
C_0	SD	0.0013	0.0009	0.0017	0.0011	0.0012	0.0009
	MAD	0.0281	0.0221	0.0511	0.0380	0.0528	0.0402
	$M-$scale	0.0045	0.0029	0.0060	0.0039	0.0040	0.0026
$C_{4,0.05}$	SD	1.9263	1.9249	1.9439	1.9541	1.9655	1.9723
	MAD	0.0421	0.0373	0.0974	0.0835	0.5355	0.5694
	$M-$scale	0.0321	0.0296	0.0584	0.0500	0.7801	0.8243
$C_{4,0.1}$	SD	1.9587	1.9666	1.9448	1.9545	1.9646	1.9719
	MAD	0.0615	0.0568	0.1624	0.1372	1.0835	1.1511
	$M-$scale	0.0540	0.0528	0.1212	0.1094	1.3264	1.3317

Table 4: Mean values of $\|\widehat{\phi}_{\text{RAW},j}/\|\widehat{\phi}_{\text{RAW},j}\| - \phi_{\text{RAW},j}\|^2$, under a proportional model, with trajectories generated from a Wiener process (BM) or from an Ornstein–Uhlenbeck process (OU).

	Scale	$\widehat{\phi}_{\text{PS},1}$		$\widehat{\phi}_{\text{PS},2}$		$\widehat{\phi}_{\text{PS},3}$	
		$f = \text{id}$	$f = \log$	$f = \text{id}$	$f = \log$	$f = \text{id}$	$f = \log$
		BM					
C_0	SD	0.0147	0.0143	0.0973	0.0931	0.2629	0.2321
	MAD	0.0354	0.0295	0.1805	0.1507	0.4131	0.3869
	$M-$scale	0.0173	0.0163	0.1158	0.1078	0.2917	0.2602
$C_{4,0.05}$	SD	0.9212	1.1136	1.9176	1.8961	1.7548	1.7496
	MAD	0.0429	0.0338	0.2525	0.1911	0.9367	0.8867
	$M-$scale	0.0250	0.0213	0.1677	0.1521	0.9885	0.8903
$C_{4,0.1}$	SD	1.7388	1.7780	1.9205	1.9048	1.7610	1.7524
	MAD	0.0529	0.0336	0.3549	0.2115	1.4118	1.3775
	$M-$scale	0.0311	0.0252	0.2259	0.1884	1.6285	1.6304
		OU					
C_0	SD	0.0013	0.0010	0.0016	0.0011	0.0013	0.0008
	MAD	0.0313	0.0203	0.0507	0.0378	0.0329	0.0392
	$M-$scale	0.0040	0.0027	0.0056	0.0037	0.0045	0.0036
$C_{4,0.05}$	SD	1.9311	1.9276	1.9413	1.9532	1.9642	1.9694
	MAD	0.0391	0.0377	0.0937	0.0875	0.4484	0.5684
	$M-$scale	0.0348	0.0361	0.0776	0.0615	0.7165	0.8350
$C_{4,0.1}$	SD	1.9574	1.9650	1.9460	1.9538	1.9625	1.9711
	MAD	0.0565	0.0513	0.1692	0.1144	0.9961	1.0925
	$M-$scale	0.0549	0.0538	0.0962	0.0999	1.2973	1.3025

Table 5: Mean values of $\|\widehat{\phi}_{\text{PS},1}/\|\widehat{\phi}_{\text{PS},1}\| - \phi_{\text{PS},1}\|^2$, under a proportional model, with trajectories generated from a Wiener process (BM) or from an Ornstein–Uhlenbeck process (OU) using a $K-$fold procedure.

7 Example: notch shape data

To illustrate the proposed procedures, we apply our estimators to notch shape data analyzed in Ramsay and Silverman (2002) where a principal component analysis and a discrimination analysis is considered. Our goal is to use the robust procedure to detect atypical or influential observations in the sample.

The data represent the shape of the knees of different individuals which are classified as healthy or suffering an arthritic condition. For each individual, we have information regarding the shape of the joint. It has been suggested that osteoarthritis can alter this shape. In particular, the inter-condylar notch is considered important by medical specialists. We refer to Ramsay and Silverman (2002) for details. The data set consists of $N = 96$ notch outlines, on each of which we have some concomitant information which provides evidence of arthritic bone damage. In the sample considered, $n_1 = 21$ femur belong to arthritic individuals and $n_2 = 75$ to individuals showing no signs of arthritic bone change. In this section, we perform a robust common functional principal components analysis over these two groups to decide if the sample contains influential observations. As in Ramsay and Silverman (2002), the data is parametrized by arc-length, leading to the plots given in Figure 1.

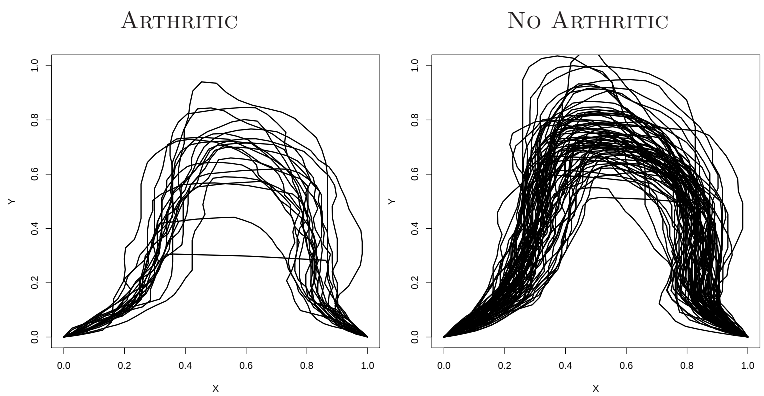

Figure 1: Notch shape of the 96 individuals.

The first four common principal directions are computed, using the standard deviation and an $M-$scale estimator. Since choosing $f = \text{id}$ or $f = \log$, lead to similar results we only report the conclusions obtained when $f = \text{id}$. On the other hand, when $\sigma_{\text{R}} = \text{SD}$ and $f = \text{id}$, the raw estimators of the common principal directions are easily obtained as the eigenfunctions of the pooled sample covariance operator, instead of maximizing (5). The robust estimators are obtained maximizing (5) over a set of candidates $\mathcal{A}$ as described in Section 6. However, to enlarge the set of candidates, we also include the classical directions in $\mathcal{A}$.

Figure 2 presents the parallel boxplots of the scores $\widehat{s}_{i,j,\ell} = \langle X_{i,j} - \widehat{\mu}_i, \widehat{\phi}_\ell \rangle$ when $\widehat{\phi}_\ell$ are the robust estimators. The inner product is taken as the standard dot product in the space of functions from $[0, 1]$ to $\mathbb{R} \times \mathbb{R}$. As is well known, due to a masking effect, the boxplots of the scores over the classical estimators do not reveal any outlier and should not be used. On the other hand, when using the robust projection–pursuit estimators the largest values of $|\widehat{s}_{i,j,\ell}|$ indicates the presence of atypical observations which may influence the estimation of the common principal directions. Figure 2 shows that, in the Arthritic group, one observation appears with a extremely small score in the first direction and another one has a large score in the fourth one. In fact, in both cases, the atypical score correspond to the observation labelled 11 and plotted with a thick line in Figure 3 and corresponds to an observation with a plateau. On the other hand, in the Healthy group, four atypical scores are detected which correspond to the observations plotted with thick lines in Figure 3. The observation with smallest score in the first direction is labelled as 22 and corresponds to the curve with smallest values among the four thick ones in Figure 3. The two observations with the largest values of $\widehat{s}_{2,j,2}$ are labelled as 62 and 63 in the Healthy data set and correspond to the observations showing some torsion to the left. Finally, individual 11 in the Healthy group is the one with the largest value in $\widehat{s}_{2,j,4}$ and is the one with the roundest behaviour among the four atypical data.

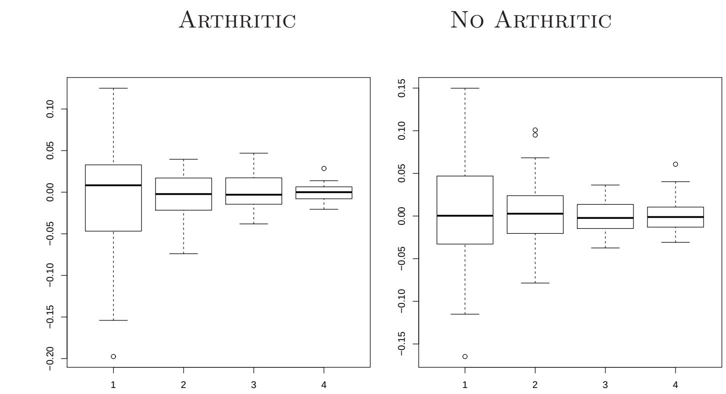

Figure 2: Boxplot of the scores.

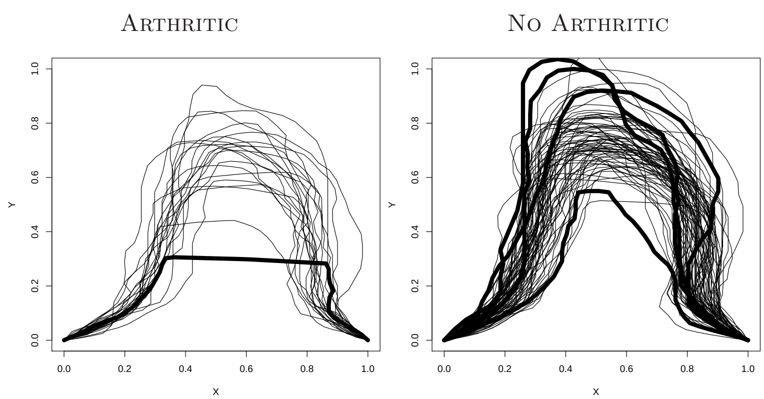

Figure 3: Notch shape of the 96 individuals, with atypical observations plotted in thick lines.

8 Concluding Remarks

In this paper, robust estimators of the common principal directions for functional data are defined using a projection–pursuit approach. In this sense, the estimators in this paper can be seen as a generalization of those defined in Bali *et al* (2011) to the case of k populations satisfying a FCPC model, as well as an extension of the estimators defined by Boente *et al.* (2006), in the multivariate setting. Besides, when considering as scale the standard deviation they provide a smooth alternative to the proposal given in Boente *et al.* (2010).

The estimators defined combine a choice for the scale with a score function and a penalization term, which penalizes the norm or the pooled scales of the projected data. A robust cross-validation procedure is defined to select sequentially the smoothing parameters which allows different penalizations according to the directions to be estimated. Consistency results are derived for the estimators defined.

Finally, the simulation study confirms the expected inadequate behaviour of the classical estimators in the presence of outliers, with the robust procedures performing significantly better. It also shows that it may be preferable to choose the logarithm function to transform the scales instead of maximizing the pooled squared scales of the projected data. In particular, we recommend using an $M-$scale combined with $f = \log$ to improve the efficiency and still obtaining reliable estimators.

A Appendix A: Preliminary results

In this section, we state some preliminary results that will be used in the sequel. Remember that $d_{\rm PR}(P,Q)$ will stand for the Prohorov metric between the probability measures P and Q. Thus, $P_n \stackrel{\omega}{\longrightarrow} P$ if and only if $d_{\rm PR}(P_n,P) \to 0$.

The following lemma, which generalizes the requirement in assumption **C3** to deal with general score functions and sequences of weights $\widehat{\tau}_i$ converging to τ_i, shows that assumption **C5** holds for general continuous score functions defined at 0. It excludes, however, the logarithm which will be treated separately.

Lemma A.1. *Let $\sigma_{\rm R}$ be a continuous scale functional and let $f : [0,+\infty) \to \mathbb{R}$ be a strictly increasing function such that $f : [0,+\infty) \to \mathbb{R}$ is a continuous function.*

a) Let $\{P_n\}_{n\in\mathbb{N}}$ and P be probability measures defined on a separable Hilbert space $\mathcal{H}$, such that $d_{\rm PR}(P_n,P) \to 0$. Then, $\sup_{\|\alpha\|=1} |f(\sigma^2_{\rm R}(P_n[\alpha])) - f(\sigma^2_{\rm R}(P[\alpha]))| \longrightarrow 0$.

b) Let $\{P_{i,n_i}\}_{n_i\in\mathbb{N}}$ and P_i, $1 \le i \le k$, be probability measures defined on a separable Hilbert space $\mathcal{H}$, such that $d_{\rm PR}(P_{i,n_i},P_i) \to 0$ and let τ_{i,n_i} be such that $0 \le \tau_{i,n_i}$ and $\tau_{i,n_i} \to \tau_i$ with $0 \le \tau_i \le 1$, $\sum_{i=1}^k \tau_i = 1$. Then, $\sup_{\|\alpha\|=1} |\sum_{i=1}^k \tau_{i,n_i} f(\sigma^2_{\rm R}(P_{i,n_i}[\alpha])) - \sum_{i=1}^k \tau_i f(\sigma^2_{\rm R}(P_i[\alpha]))| \longrightarrow 0$.

Proof. a) Note that there exists a metric d generating the weak topology in $\mathcal{H}$, the closed ball $\mathcal{V}_r = \{\alpha : \|\alpha\| \le r\}$ is weakly compact and so, compact with respect to d. On the other hand, $\sigma(\alpha) = \sigma_{\rm R}(P[\alpha])$ is a weakly continuous function of α in $\mathcal{H}$, hence continuous with respect to d. These facts entail that the set $\mathcal{A} = \{\sigma^2(\alpha) : \|\alpha\| \le 1\}$ is compact in $[0,+\infty)$, so bounded. Let us assume that $\mathcal{A} \subset [0,A] \subset \mathbb{R}$. The fact that f is continuous in $[0,\infty)$ implies that it is uniformly continuous in $[0,A+1]$. Hence, for any $\epsilon > 0$ there exists $\delta > 0$ such that $u,v \in [0,A+1]$, $|u-v| \le \delta$ entail $|f(u) - f(v)| < \epsilon$.

Theorem 6.2 in Bali *et al.* (2011) implies that $\sup_{\|\alpha\|=1} |\sigma_{\rm R}(P_n[\alpha]) - \sigma_{\rm R}(P[\alpha])| \longrightarrow 0$, thus there exist $n_o \in \mathbb{N}$ such that, for any $n \ge n_o$, $\sup_{\|\alpha\|=1} |\sigma_{\rm R}(P_n[\alpha]) - \sigma_{\rm R}(P[\alpha])| \le \min(\delta/(2(A+1)),\eta)$, where $\eta \le \min(1/\sqrt{A},1)/4$. Thus, using that $\sigma_{\rm R}(P[\alpha]) \in [0,\sqrt{A}]$, for any $\alpha \in \mathcal{S}_1$, we get that, for any $\alpha \in \mathcal{S}_1$, $u_n = \sigma^2_{\rm R}(P_n[\alpha]) \in [0,A+1]$, $v = \sigma^2_{\rm R}(P[\alpha]) \in [0,A+1]$ and $|u_n - v| \le \delta$, which entails that $|f\left(\sigma^2_{\rm R}(P_n[\alpha])\right) - f\left(\sigma^2_{\rm R}(P[\alpha])\right)| < \epsilon$, for any $\alpha \in \mathcal{S}_1$, concluding the proof.

b) Using a) we get easily that $\sup_{\|\alpha\|=1} |\sum_{i=1}^k \tau_{i,n_i} \left(f(\sigma^2_{\rm R}(P_{i,n_i}[\alpha])) - f(\sigma^2_{\rm R}(P_i[\alpha]))\right)| \longrightarrow 0$, since $\tau_{i,n_i} \le 2$, for n_i large enough. It remains to show that $\sup_{\|\alpha\|=1} |\sum_{i=1}^k (\tau_{i,n_i} - \tau_i) f(\sigma^2_{\rm R}(P_i[\alpha]))| \longrightarrow 0$. Noting again that the closed ball $\mathcal{V}_1 = \{\alpha : \|\alpha\| \le 1\}$ is weakly compact and $g_i(\alpha) = f(\sigma^2_i(\alpha)) =$

$f(\sigma^2_{\mathrm{R}}(P_i[\alpha]))$ are weakly continuous functions of α in $\mathcal{H}$, we get that the sets $\mathcal{B}_i = \{f(\sigma_i^2(\alpha)) : \|\alpha\| \le 1\}$ are compact sets, so bounded which together with the fact that $\tau_{i,n_i} \to \tau_i$ concludes the proof of b). □

Using Lemma A.1, we get the following result

Corollary A.1. *Let σ_{R} be a continuous scale functional and let $f : [0, +\infty) \to \mathbb{R}$ be a strictly increasing function such that $f : [0, +\infty) \to \mathbb{R}$ is a continuous function.*

a) *Given P be a probability measure in a separable Hilbert space $\mathcal{H}$ and P_n be the empirical measure of a random sample $X_1, \dots, X_n$ with $X_i \sim P$, we have that $\sup_{\|\alpha\|=1} |f(\sigma^2_{\mathrm{R}}(P_n[\alpha])) - f(\sigma^2_{\mathrm{R}}(P[\alpha]))| \xrightarrow{a.s.} 0$*

b) *Given P_i, $1 \le i \le k$, probability measures defined on a separable Hilbert space $\mathcal{H}$ and $\{P_{i,n_i}\}_{n_i \in \mathbb{N}}$ the empirical measures of independent random samples $X_{i,1}, \dots, X_{i,n_i}$ with $X_{i,1} \sim P_i$, we have that $\sup_{\|\alpha\|=1} |\sum_{i=1}^k \tau_{i,n_i} f(\sigma^2_{\mathrm{R}}(P_{i,n_i}[\alpha])) - \sum_{i=1}^k \tau_i f(\sigma^2_{\mathrm{R}}(P_i[\alpha]))| \xrightarrow{a.s.} 0$, for any sequence τ_{i,n_i} such that $0 \le \tau_{i,n_i}$ and $\tau_{i,n_i} \xrightarrow{a.s.} \tau_i$ with $0 \le \tau_i \le 1$, $\sum_{i=1}^k \tau_i = 1$.*

The following lemma will be used to derive the results stated in Section 4 when considering general continuous score functions defined at 0. Its proof is omitted since it follows using analogous arguments to those considered in the proof of Lemma A.1.

Lemma A.2. *Let σ_{R} be a continuous scale functional and let $f : [0, +\infty) \to \mathbb{R}$ be a strictly increasing function such that $f : [0, +\infty) \to \mathbb{R}$ is a continuous function. Let $\{P_{i,n_i}\}_{n_i \in \mathbb{N}}$ and P_i, $1 \le i \le k$, be probability measures defined on a separable Hilbert space $\mathcal{H}$, such that $\sup_{\alpha \in \mathcal{A}_N} \left|\sigma^2_{\mathrm{R}}(P_{i,n_i}[\alpha]) - \sigma^2_{\mathrm{R}}(P_i[\alpha])\right| \longrightarrow 0$, where $\mathcal{A}_N \subset \mathcal{V}_1 = \{\alpha : \|\alpha\| \le 1\}$, and let τ_{i,n_i} be such that $0 \le \tau_{i,n_i}$ and $\tau_{i,n_i} \to \tau_i$ with $0 \le \tau_i \le 1$, $\sum_{i=1}^k \tau_i = 1$. Then, $\sup_{\alpha \in \mathcal{A}_N} |\sum_{i=1}^k \tau_{i,n_i} f(\sigma^2_{\mathrm{R}}(P_{i,n_i}[\alpha])) - \sum_{i=1}^k \tau_i f(\sigma^2_{\mathrm{R}}(P_i[\alpha]))| \longrightarrow 0$.*

B Appendix B: Proofs

Proof of Lemma 2.1. Using that f is a continuous function and σ_i is weakly continuous, we get easily that $f \circ \sigma_i : \mathcal{H} \to \mathbb{R}$ and $\varsigma_f : \mathcal{H} \to \mathbb{R}$ are continuous functions with respect to the weak topology in $\mathcal{H}$. The result follows now from the fact that the unit ball $\{\|\alpha\| = 1\}$ is weakly-compact, since any continuous function reaches its maximum over a compact set. □

Proof of Lemma 2.2. Let $\alpha_n \in \mathcal{S}_1$ be a sequence such that $\varsigma_f(\alpha_n) \to \sup_{\|\alpha\|=1} \varsigma_f(\alpha)$.

Let us begin showing that $\liminf_{n\to\infty} \sigma_i(\alpha_n) = d_i > 0$ for all $1 \le i \le k$. Assume that this assertion does not hold, i.e., that there exists an i and a subsequence α_{n_j} such that $\sigma_i(\alpha_{n_j}) \to 0$ as $n_j \to \infty$. Then, $\varsigma_f(\alpha_{n_j}) \to \sum_{i=1}^k \tau_i f(\sigma_i^2(0)) = -\infty$ which implies that $\sup_{\|\alpha\|=1} \varsigma_f(\alpha) = -\infty$. On the other hand, using that there exists α_0 such that $\sigma_i(\alpha_0) > 0$ for all i, we get $-\infty < \varsigma_f(\alpha_0/\|\alpha_0\|) \le \sup_{\|\alpha\|=1} \varsigma_f(\alpha)$ leading to a contradiction.

Hence, $\liminf_{n\to\infty} \sigma_i(\alpha_n) = d_i > 0$ for all i. Without loss of generality, assume that $\sigma_i(\alpha_n) \to d_i$. Therefore, there exists n_0 such that for $n \ge n_0$, we have that $\sigma_i(\alpha_n) > d_i/2 > 0$. After relabelling the sequence, we can assume that $\sigma_i(\alpha_n) > A > 0$ for all i and n.

Using that $\mathcal{S}_1$ is weakly compact we have that the exists a subsequence α_{n_m} converging to $\beta \in \mathcal{H}$ and $\|\beta\| \leq 1$. Let us show that $\|\beta\| \neq 0$. If $\beta = 0$, we have that $\sigma_i(\beta) = 0$. However, the weak continuity of σ_i entails that $\sigma_i(\beta) = \lim_m \sigma_i(\alpha_{n_m}) > A > 0$ leading to a contradiction. Hence $\|\beta\| \neq 0$. Then, $\alpha_{n_m} \to \beta$ and $\sigma_i(\alpha_{n_m}) > A$ for all i and n. Since $f = \log : [A, \infty] \to \mathbb{R}$ is a continuous function, we get that $\varsigma_f(\alpha_{n_m}) \to \varsigma_f(\beta)$. On the other hand, $\varsigma_f(\alpha_{n_m}) \to \sup_{\|\alpha\|=1} \varsigma_f(\alpha)$, thus $\sup_{\|\alpha\|=1} \varsigma_f(\alpha) = \varsigma_f(\beta)$. It remains to show that $\|\beta\| = 1$. Assume that $\|\beta\| < 1$ and define $\gamma = \beta/\|\beta\|$. We have that $\log(\|\beta\|^2) < 0$, so $\varsigma_f(\gamma) = \varsigma_f(\beta) - \sum_{i=1}^k \tau_i \log(\|\beta\|^2) > \varsigma_f(\beta) = \sup_{\|\alpha\|=1} \varsigma_f(\alpha)$, leading to a contradiction. Therefore, $\|\beta\| = 1$ and the supremum is reached at β. □

Proof of Lemma 2.3. The fact that $P_1, \dots, P_k$ are weakly–FCPC under σ_{R} entails that $\phi_{\text{R},j}(P_i) = \phi_{\text{R},j}(P_m) = \phi_{\text{R},j}$ for all $j \geq 1$ and $1 \leq i, m \leq k$. So, for any $\alpha \in \mathcal{S}_1$, we have that $\sigma^2_{\text{R}}(P_i[\phi_{\text{R},1}]) \geq \sigma^2_{\text{R}}(P_i[\alpha])$ which together with the fact that f is strictly increasing entails that $f(\sigma^2_{\text{R}}(P_i[\phi_{\text{R},1}])) \geq f(\sigma^2_{\text{R}}(P_i[\alpha]))$. Hence, $\varsigma_f(\alpha) \leq \varsigma_f(\phi_{\text{R},1})$, which implies that $\phi_{f,1} = \phi_{\text{R},1}$.

The proof follows now by an induction argument. Assume that $\phi_{f,j}(P) = \phi_{\text{R},j}(P_1)$ for $1 \leq j \leq m$, we want to show that $\phi_{f,m+1}(P) = \phi_{\text{R},m+1}(P_1)$. First note that $\mathcal{B}_{m+1} = \mathcal{B}_{f,m+1}$, so for any $\alpha \in \mathcal{S}_1 \cap \mathcal{B}_{f,m+1}$, we have that $\sigma_i(\alpha) \leq \sigma_i(\phi_{\text{R},m+1})$, so $\varsigma_f(\alpha) \leq \varsigma_f(\phi_{\text{R},m+1})$, concluding the proof. □

Proof of Lemma 2.4. Let us begin by showing the result for $j = 1$. Note that **A1** implies that $P_1, \dots, P_k$ are strongly–FCPC and so, weakly–FCPC under σ_{R}. Thus, for any $\alpha \in \mathcal{S}_1$, we have that $\sigma^2_{\text{R}}(P_i[\phi_1]) = \sigma^2_i(\phi_1) = c_i\lambda_{i,1} \geq c_i\langle\alpha, \mathbf{\Gamma}_0\alpha\rangle = \sigma^2_i(\alpha) = \sigma^2_{\text{R}}(P_i[\alpha])$ and the inequality is strict when $i = i_0$ and $\alpha \neq \phi_1$. Hence, $f(\sigma^2_i(\phi_1)) \geq f(\sigma^2_i(\alpha))$ for any $1 \leq i \leq k$ and $f(\sigma^2_{i_0}(\phi_1)) > f(\sigma^2_{i_0}(\alpha))$ since f is strictly increasing which together with the fact that $\tau_i \geq 0$ and $\tau_{i_0} > 0$ imply that $\varsigma_f(\phi_1)) > \varsigma_f(\alpha)$ for any $\alpha \in \mathcal{S}_1$. Thus, $\phi_{f,1}(P) = \phi_1$.

The proof follows easily using an induction argument. Assume that $\phi_{f,j}(P) = \phi_j$, for $1 \leq j \leq m-1$, $m \leq q$, we want to show that $\phi_{f,m}(P) = \phi_m$. Now the set $\mathcal{B}_{f,m}$ equals $\{\alpha : \langle\alpha, \phi_j\rangle = 0, 1 \leq j \leq m-1\}$, hence, for any $\alpha \in \mathcal{S}_1 \cap \mathcal{B}_{f,m}$, we have that $\langle\alpha, \mathbf{\Gamma}_{i,0}\alpha\rangle \leq \langle\phi_m, \mathbf{\Gamma}_{i,0}\phi_m\rangle = \lambda_{i,m}$ with strict inequality when $i = i_0$ and $\alpha \neq \phi_m$. This implies that for any $\alpha \in \mathcal{S}_1 \cap \mathcal{B}_{f,m}$, we have $f(\sigma^2_i(\phi_m)) = f\left(c_i\langle\phi_m, \mathbf{\Gamma}_{i,0}\phi_m\rangle\right) \geq f\left(c_i\langle\alpha, \mathbf{\Gamma}_{i,0}\alpha\rangle\right) = f(\sigma^2_i(\alpha))$ for any $1 \leq i \leq k$ and $f(\sigma^2_{i_0}(\phi_m)) > f(\sigma^2_{i_0}(\alpha))$ which entails that $\varsigma_f(\phi_m)) > \varsigma_f(\alpha)$, so $\phi_{f,m}(P) = \phi_m$, concluding the proof.

The result regarding the eigenvalues follow easily since $\lambda_{f,i,j} = \sigma^2_{\text{R}}\left(P_i[\phi_{f,j}]\right) = \sigma^2_i(\phi_j) = c_i\langle\phi_j, \mathbf{\Gamma}_{i,0}\phi_j\rangle = c_i\lambda_{i,j}$. □

Proof of Lemma 4.1. First of all, note that **C1** and **C2** imply that $\varsigma_f : \mathcal{H} \to \mathbb{R}$ is a weakly continuous function. Hence, it is a weakly uniformly continuous function on $\mathcal{V}_1$ which is weakly compact.

a) Let $\mathcal{N} = \{\omega : \varsigma_f(\widehat{\phi}_1(\omega)) \not\to \varsigma_f(\phi_{f,1})\}$ and fix $\omega \notin \mathcal{N}$, then $\varsigma_f(\widehat{\phi}_1(\omega)) \to \varsigma_f(\phi_{f,1})$. Using $\{\|\alpha\| \leq 1\}$ is weakly compact, we have that for any subsequence γ_ℓ of the sequence $\widehat{\phi}_1(\omega)$ there exists a subsequence γ_{ℓ_s} such that $\gamma_{\ell_s} \to \gamma \in \mathcal{H}$ such that that $\|\gamma\| \leq 1$. Besides, using that $\varsigma_f(\widehat{\phi}_1(\omega)) \to \varsigma_f(\phi_{f,1})$, we get that $\varsigma_f(\gamma_{\ell_s}) \to \varsigma_f(\phi_{f,1})$ while on the other hand, the weak continuity of ς_f entails that $\varsigma_f(\gamma_{\ell_s}) \to \varsigma_f(\gamma)$, as $s \to \infty$. Hence, $\varsigma_f(\gamma) = \varsigma_f(\phi_{f,1})$ which entails that $\gamma \neq 0$. Effectively, assume that $\gamma = 0$. Then, we have that $\sigma_{\text{R}}(P_i[\gamma]) = \sigma_{\text{R}}(P_i[0]) = 0$ which implies that $\varsigma_f(\gamma) = f(0)$, since $\sum_{i=1}^k \tau_i = 1$. Therefore, $\varsigma_f(\phi_{f,1}) = f(0)$ and $\varsigma_f(\phi_{f,1}) = \sum_{i=1}^k \tau_i f(\lambda_{f,i,1})$. Using that f is strictly increasing and the fact that **C0** implies that there exist i such that $\lambda_{f,i,1} > 0$, we get that $\sum_{i=1}^k \tau_i f(\lambda_{f,i,1}) > f(0)$ leading to a contradiction. Hence, $\gamma \neq 0$.

Assume that $\|\gamma\| < 1$ and let $\widetilde{\gamma} = \gamma/\|\gamma\|$, then $\widetilde{\gamma} \in \mathcal{S}$ which implies that $\varsigma_f(\widetilde{\gamma}) \leq \varsigma_f(\phi_{f,1})$. On the other hand, using that $\sigma_{\rm R}$ is a scale functional, $\|\gamma\| < 1$ and f is strictly increasing, we get

$$\varsigma_f(\widetilde{\gamma}) = \sum_{i=1}^{k} \tau_i f\left(\sigma_{\rm R}^2\left(P_i\left[\widetilde{\gamma}\right]\right)\right) = \sum_{i=1}^{k} \tau_i f\left(\frac{\sigma_{\rm R}^2(P_i[\gamma])}{\|\gamma\|^2}\right) > \sum_{i=1}^{k} \tau_i f\left(\sigma_{\rm R}^2(P_i[\gamma])\right) = \varsigma_f(\gamma) = \varsigma_f(\phi_{f,1})$$

which contradicts the fact that $\varsigma_f(\phi_{f,1}) = \max_{\|\alpha\|=1} \varsigma_f(\alpha)$. Hence, $\|\gamma\| = 1$ and **C0** implies that $\gamma = \phi_{f,1}$ except maybe for a sign change, that is, $\langle \gamma, \phi_{f,1}\rangle^2 = 1$. Thus, any subsequence of $\widehat{\phi}_1(\omega)$ will have a limit converging either to $\phi_{\rm R,1}$ or $-\phi_{\rm R,1}$, concluding the proof of a).

b) Write $\widehat{\phi}_m$ as $\widehat{\phi}_m = \sum_{j=1}^{m-1} \widehat{a}_j \phi_{f,j} + \widehat{\gamma}_m$, with $\langle \widehat{\gamma}_m, \phi_{f,j}\rangle = 0$, $1 \le j \le m-1$. To obtain b) we only have to show that $\langle \widehat{\gamma}_m, \phi_{f,m}\rangle^2 \xrightarrow{a.s.} 1$. Note that $\langle \widehat{\phi}_m, \widehat{\phi}_j\rangle \xrightarrow{a.s.} 0$, for $j \neq m$, implies that $\widehat{a}_j = \langle \widehat{\phi}_m, \phi_{f,j}\rangle = \langle \widehat{\phi}_m, \phi_{f,j} - \widehat{\phi}_j\rangle + \langle \widehat{\phi}_m, \widehat{\phi}_j\rangle = \langle \widehat{\phi}_m, \phi_{f,j} - \widehat{\phi}_j\rangle + o_{\rm a.s.}(1)$. Thus, using that $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{f,j}$, $1 \le j \le m-1$, and $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$, we get that $\widehat{a}_j \xrightarrow{a.s.} 0$ for $1 \le j \le m-1$. Therefore, $\|\widehat{\phi}_m - \widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 0$. Moreover, using that $\|\widehat{\phi}_m\|^2 = \sum_{j=1}^{m-1} \widehat{a}_j^2 + \|\widehat{\gamma}_m\|^2$ and $\|\widehat{\phi}_m\|^2 \xrightarrow{a.s.} 1$, we get that $\|\widehat{\gamma}_m\|^2 \le 1$ and $\|\widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 1$, which implies that $\|\widehat{\phi}_m - \widetilde{\gamma}_m\| \xrightarrow{a.s.} 0$, where $\widetilde{\gamma}_m = \widehat{\gamma}_m/\|\widehat{\gamma}_m\|$.

Using now that $\varsigma_f(\alpha)$ is a weakly uniformly continuous function of α in $\mathcal{V}_1$, we obtain that $\varsigma_f(\widetilde{\gamma}_m) - \varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ which together with the fact that $\varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$ implies that $\varsigma_f(\widetilde{\gamma}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$. The proof follows now as in a) using the fact that $\widetilde{\gamma}_m \in \mathcal{C}_m$, with $\mathcal{C}_m = \{\alpha \in \mathcal{S} : \langle \alpha, \phi_{f,j}\rangle = 0,\ 1 \le j \le m-1\}$ and $\phi_{f,m}$ is the unique maximizer of $\varsigma_f(\alpha)$ over $\mathcal{C}_m$. □

Proof of Lemma 4.2. The proof is quite similar to that of Lemma 4.1 avoiding the problems caused by the singularity at 0. Again, **C1** imply that $\sigma_i : \mathcal{H} \to \mathbb{R}$ is a weakly uniformly continuous function on $\mathcal{V}_1$ which is weakly compact.

a) As in Lemma 4.1, let $\mathcal{N} = \{\omega : \varsigma_f(\widehat{\phi}_1(\omega)) \not\to \varsigma_f(\phi_{f,1})\}$ and fix $\omega \notin \mathcal{N}$, then $\varsigma_f(\widehat{\phi}_1(\omega)) \to \varsigma_f(\phi_{f,1})$. Using $\{\|\alpha\| \le 1\}$ is weakly compact, we have that for any subsequence γ_ℓ of the sequence $\widehat{\phi}_1(\omega)$ there exists a subsequence γ_{ℓ_s} such that $\gamma_{\ell_s} \to \gamma \in \mathcal{H}$ such that that $\|\gamma\| \le 1$. Besides, using that $\varsigma_f(\widehat{\phi}_1(\omega)) \to \varsigma_f(\phi_{f,1})$, we get that $\varsigma_f(\gamma_{\ell_s}) \to \varsigma_f(\phi_{f,1})$. On the other hand, the weak continuity of σ_i entails that $\sigma_i(\gamma_{\ell_s}) \to \sigma_i(\gamma)$, as $s \to \infty$, for $1 \le i \le k$. If there exist $1 \le i \le k$ such that $\sigma_i(\gamma) = 0$, (which includes the situation $\gamma = 0$), the fact that $\tau_i \ge 0$ implies that $\varsigma_f(\gamma_{\ell_s}) \to -\infty$ and so, $\varsigma_f(\phi_{f,1}) = -\infty$ which contradicts the fact that $\varsigma_f(\phi_{f,1}) = \sum_{i=1}^k \tau_i f(\lambda_{f,i,1})$ and $\lambda_{f,i,1} > 0$ for all $1 \le i \le k$. Thus $\sigma_i(\gamma) \neq 0$, for all $1 \le i \le k$, (which entails that $\gamma \neq 0$), the continuity of $f = \log$ implies that $\varsigma_f(\gamma_{\ell_s}) \to \varsigma_f(\gamma)$ and so, $\varsigma_f(\gamma) = \varsigma_f(\phi_{f,1})$. The proof follows now as in Lemma 4.1.

b) Write $\widehat{\phi}_m$ as $\widehat{\phi}_m = \sum_{j=1}^{m-1} \widehat{a}_j \phi_{f,j} + \widehat{\gamma}_m$, with $\langle \widehat{\gamma}_m, \phi_{f,j}\rangle = 0$, $1 \le j \le m-1$. To obtain b) we only have to show that $\langle \widehat{\gamma}_m, \phi_{f,m}\rangle^2 \xrightarrow{a.s.} 1$. Note that $\langle \widehat{\phi}_m, \widehat{\phi}_j\rangle \xrightarrow{a.s.} 0$, for $j \neq m$, implies that $\widehat{a}_j = \langle \widehat{\phi}_m, \phi_{f,j}\rangle = \langle \widehat{\phi}_m, \phi_{f,j} - \widehat{\phi}_j\rangle + \langle \widehat{\phi}_m, \widehat{\phi}_j\rangle = \langle \widehat{\phi}_m, \phi_{f,j} - \widehat{\phi}_j\rangle + o_{\rm a.s.}(1)$. Thus, using that $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{f,j}$, $1 \le j \le m-1$, and $\|\widehat{\phi}_m\| \xrightarrow{a.s.} 1$, we get that $\widehat{a}_j \xrightarrow{a.s.} 0$ for $1 \le j \le m-1$. Therefore, $\|\widehat{\phi}_m - \widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 0$. Moreover, using that $\|\widehat{\phi}_m\|^2 = \sum_{j=1}^{m-1} \widehat{a}_j^2 + \|\widehat{\gamma}_m\|^2$ and $\|\widehat{\phi}_m\|^2 \xrightarrow{a.s.} 1$, we get that $\|\widehat{\gamma}_m\|^2 \le 1$ and $\|\widehat{\gamma}_m\|^2 \xrightarrow{a.s.} 1$, which implies that $\|\widehat{\phi}_m - \widetilde{\gamma}_m\| \xrightarrow{a.s.} 0$, where $\widetilde{\gamma}_m = \widehat{\gamma}_m/\|\widehat{\gamma}_m\|$.

Using now that, for all $1 \le i \le k$, $\sigma_i^2(\alpha)$ is a weakly uniformly continuous function of α in $\mathcal{V}_1$, we obtain that $\sigma_i^2(\widetilde{\gamma}_m) - \sigma_i^2(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ which together with the fact that $\sigma_i^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma_i^2(\phi_{f,m})$ implies that $\sigma_i^2(\widetilde{\gamma}_m) \xrightarrow{a.s.} \sigma_i^2(\phi_{f,m})$. Therefore, noting that $\lambda_{f,i,m} > 0$ implies that $\sigma_i(\phi_{f,m}) > 0$, for all $1 \le i \le k$, we get from the continuity of $f = \log$ on $(0, \infty)$ that $\varsigma_f(\widetilde{\gamma}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$.

The proof follows now as in a) using the fact that $\widetilde{\gamma}_m \in \mathcal{C}_m$, with $\mathcal{C}_m = \{\alpha \in \mathcal{S} : \langle \alpha, \phi_{f,j}\rangle = 0,\ 1 \le j \le m-1\}$ and $\phi_{f,m}$ is the unique maximizer of $\varsigma_f(\alpha)$ over $\mathcal{C}_m$. □

We will need first some auxiliary definitions. Denote by $\mathcal{L}_k$ the linear space spanned by $\{\phi_{f,1}, \dots, \phi_{f,k}\}$ and let $\widehat{\mathcal{L}}_k$ be the linear space spanned by the first common principal directions $\widehat{\phi}_1, \dots \widehat{\phi}_k$. For any linear space $\mathcal{L}$, let $\pi_{\mathcal{L}} : \mathcal{H} \to \mathcal{L}$ be the orthogonal projection onto $\mathcal{L}$, which is well defined if $\mathcal{L}$ is a closed space, for instance, if $\mathcal{L}$ is a finite–dimensional linear space.

Denote by $\mathcal{T}_k = \mathcal{L}_k^{\perp}$ the linear space orthogonal to $\mathcal{L}_k$ and by $\pi_k = \pi_{\mathcal{T}_k}$ the orthogonal projection with respect to the inner product defined in $\mathcal{H}$. On the other hand, let $\widehat{\pi}_{\nu,k}$ be the projection onto the linear space orthogonal to $\widehat{\phi}_1, \dots, \widehat{\phi}_k$ in the space $\mathcal{H}_{\mathrm{S}}$ in the inner product $\langle \cdot, \cdot \rangle_\nu$. That is, $\widehat{\pi}_{\tau,k}(\alpha) = \alpha - \sum_{j=1}^k \langle \alpha, \widehat{\phi}_j \rangle_\nu \widehat{\phi}_j$. Moreover, $\widehat{\mathcal{T}}_{\nu,k}$ will stand for the linear space orthogonal to $\widehat{\mathcal{L}}_k$ with the inner product $\langle \cdot, \cdot \rangle_\nu$. Thus, $\widehat{\pi}_{\nu,k}$ is the orthogonal projection onto $\widehat{\mathcal{T}}_{\nu,k}$ with respect to this inner product.

Proof of Theorem 4.1. The proof uses similar arguments to those considered in the proof of Theorem 4.1 in Bali *et al.* (2011).

First note that as in the proof of Lemma A.1, assumption ii), **C2** and the fact that $\tau_{i,n_i} \xrightarrow{a.s.} \tau_i$ imply that

$$\sup_{\|\alpha\|=1} |\varsigma_N(\alpha) - \varsigma_f(\alpha)| \xrightarrow{a.s.} 0 \,. \tag{B.1}$$

Besides, the fact that σ_{R} is a scale functional entails that $\sigma_{i,n_i}(\alpha) = \|\alpha\| \; \sigma_{i,n_i}(\alpha/\|\alpha\|)$. Thus, from assumption ii) and the fact that $\|\alpha\| \leq \|\alpha\|_\nu$ we get that

$$\sup_{\|\alpha\|\leq 1} \left|\sigma_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha)\right| \xrightarrow{a.s.} 0 \quad \text{and} \quad \sup_{\|\alpha\|_\nu \leq 1} \left|\sigma_{i,n_i}^2(\alpha) - \sigma_i^2(\alpha)\right| \xrightarrow{a.s.} 0 \,. \tag{B.2}$$

Hence using that the sets $\mathcal{B}_i = \{\sigma_i^2(\alpha), \|\alpha\| \leq 1\}$ are compact sets on $[0, \infty)$ and the fact that f is a continuous function on $[0, \infty)$ and so, uniformly continuous on any closed neighbourhood of $\mathcal{B}_i$, we get easily from Lemma A.2 that

$$\sup_{\|\alpha\|\leq 1} |\varsigma_N(\alpha) - \varsigma_f(\alpha)| \xrightarrow{a.s.} 0 \quad \text{and} \quad \sup_{\|\alpha\|_\nu \leq 1} |\varsigma_N(\alpha) - \varsigma_f(\alpha)| \xrightarrow{a.s.} 0 \,. \tag{B.3}$$

Note also that since f is defined at 0 we can assume without loss of generality that $f(0) = 0$ which entails that $\varsigma_f(\phi_{f,m}) > 0$ for $1 \leq m \leq q$.

a) To prove $\varsigma_f(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$, using (B.3) and the fact that $\|\widehat{\phi}_1\| \leq \|\widehat{\phi}_1\|_\nu = 1$, it suffices to show that $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$, that will follow if we show that

$$\varsigma_f(\phi_{f,1}) \;\geq\; \varsigma_N(\widehat{\phi}_1) + o_{\mathrm{a.s.}}(1) \,, \tag{B.4}$$

$$\varsigma_f(\phi_{f,1}) \;\leq\; \varsigma_N(\widehat{\phi}_1) + o_{\mathrm{a.s.}}(1) \,. \tag{B.5}$$

Using (B.3), we get that $\widehat{a}_{N,1} = \varsigma_N(\widehat{\phi}_1) - \varsigma_f(\widehat{\phi}_1) \xrightarrow{a.s.} 0$, $\widehat{b}_{N,1} = \varsigma_N(\phi_{f,1}) - \varsigma_f(\phi_{f,1}) \xrightarrow{a.s.} 0$ and $\widehat{c}_{N,1} = \varsigma_N(\phi_{f,1}/\|\phi_{f,1}\|_\nu) - \varsigma_f(\phi_{f,1}/\|\phi_{f,1}\|_\nu) \xrightarrow{a.s.} 0$. Using that σ_{R} is a scale functional, f is strictly increasing, $\varsigma_f(\phi_{f,1}) = \sup_{\alpha \in \mathcal{S}} \varsigma_f(\alpha)$ and the fact that $\|\widehat{\phi}_1\| \leq \|\widehat{\phi}_1\|_\nu = 1$, we obtain easily that $\varsigma_f(\phi_{f,1}) \geq \varsigma_f\left(\widehat{\phi}_1/\|\widehat{\phi}_1\|\right) \geq \varsigma_f(\widehat{\phi}_1) = \varsigma_N(\widehat{\phi}_1) - \widehat{a}_{N,1} = \varsigma_N(\widehat{\phi}_1) + o_{\mathrm{a.s.}}(1)$, concluding the proof of (B.4).

To derive (B.5), note that since $\phi_{f,1} \in \mathcal{H}_{\mathrm{S}}$, $\|\phi_{f,1}\|_\nu < \infty$ and $\|\phi_{f,1}\|_\nu \geq \|\phi_{f,1}\| = 1$. Then, using that $\widehat{\phi}_1 = \mathrm{argmax}_{\|\alpha\|_\nu = 1} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\}$ and defining $\beta_1 = \phi_{f,1}/\|\phi_{f,1}\|_\nu$, we have that $\|\beta_1\|_\nu = 1$ and

$$\varsigma_N(\widehat{\phi}_1) \geq \varsigma_N(\widehat{\phi}_1) - \rho\Psi(\widehat{\phi}_1) \geq \varsigma_N(\beta_1) - \rho\Psi(\beta_1) = \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) - \rho\Psi\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) \,.$$

Hence, using that $\Psi(a\alpha) = a^2\Psi(\alpha)$, for any $a \in \mathbb{R}$, we get

$$\begin{aligned}\varsigma_N(\widehat{\phi}_1) &\geq \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) - \rho\Psi\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) = \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) - \rho\frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|_\nu^2} \\ &\geq \varsigma_f\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) + \widehat{c}_{N,1} - \rho\frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|_\nu^2}\,.\end{aligned}$$

When $\rho = 0$, we have defined $\rho\Psi(\phi_{f,1}) = 0$ and similarly when $\nu = 0$ in which case $\|\phi_{f,1}\|_\nu = \|\phi_{f,1}\| = 1$. So from now on, we will assume that $\nu_N > 0$ and $\rho_N > 0$. Since $\nu \xrightarrow{a.s.} 0$, we have that and $\|\phi_{f,1}\|_\nu \xrightarrow{a.s.} \|\phi_{f,1}\| = 1$. Using that $\varsigma_f : \mathcal{H} \to \mathbb{R}$ is weakly continuous, we get that $\varsigma_f\left(\phi_{f,1}/\|\phi_{f,1}\|_\nu\right) \xrightarrow{a.s.} \varsigma_f\left(\phi_{f,1}\right)$, which together with the fact that $\widehat{c}_{N,1} = o_{\text{a.s.}}(1)$, $\rho \xrightarrow{a.s.} 0$, so $\rho\Psi(\phi_{f,1}) \xrightarrow{a.s.} 0$, entails that

$$\varsigma_N(\widehat{\phi}_1) \geq \varsigma_f\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) + \widehat{c}_{N,1} - \rho\frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|_\nu^2} \geq \varsigma_f\left(\phi_{f,1}\right) + o_{\text{a.s.}}(1)\,,$$

concluding the proof of (B.5). Thus, $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f\left(\phi_{f,1}\right)$.

As mentioned above, from (B.3) and the fact that $\|\widehat{\phi}_1\| \leq 1$, we obtain that $\varsigma_N(\widehat{\phi}_1) - \varsigma_f(\widehat{\phi}_1) \xrightarrow{a.s.} 0$. Therefore, using that $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f\left(\phi_{f,1}\right)$, we get that

$$\varsigma_f(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})\,. \tag{B.6}$$

Moreover, the inequalities $\varsigma_f(\phi_{f,1}) \geq \varsigma_f\left(\widehat{\phi}_1/\|\widehat{\phi}_1\|\right) \geq \varsigma_f(\widehat{\phi}_1)$ obtained above imply that

$$\varsigma_f\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})\,. \tag{B.7}$$

We have to show that $\rho\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$, which follows easily from the fact that $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$, $\|\phi_{f,1}\|_\nu \xrightarrow{a.s.} 1$, $\varsigma_f(\phi_{f,1}/\|\phi_{f,1}\|_\nu) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$, $\widehat{c}_{N,1} \xrightarrow{a.s.} 0$ and $\rho \xrightarrow{a.s.} 0$ since

$$\varsigma_N(\widehat{\phi}_1) \geq \varsigma_N(\widehat{\phi}_1) - \rho\Psi(\widehat{\phi}_1) \geq \varsigma_f\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) + \widehat{c}_{N,1} - \rho\frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|_\nu^2}\,.$$

It remains to show that $\nu\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$. Using that $\|\widehat{\phi}_1\|_\nu = 1$, we get that $\nu\Psi(\widehat{\phi}_1) = 1 - \|\widehat{\phi}_1\|^2$. Hence, we only have to show that $\|\widehat{\phi}_1\| \xrightarrow{a.s.} 1$. Note that $0 < \|\widehat{\phi}_1\| \leq 1$. Let $\mathcal{N} = \{\varsigma_f(\widehat{\phi}_1) \not\to \varsigma_f(\phi_{f,1}), \varsigma_f\left(\widehat{\phi}_1/\|\widehat{\phi}_1\|\right) \not\to \varsigma_f(\phi_{f,1})\}$ and fix $\omega \notin \mathcal{N}$ and denote $\gamma_N = \widehat{\phi}_1(\omega)$, so that $0 < y_N = \|\gamma_N\| \leq 1$,

$$\varsigma_f(\gamma_N) \to \varsigma_f(\phi_{f,1}) \quad \text{and} \quad \varsigma_f\left(\gamma_N/\|\gamma_N\|\right) \to \varsigma_f(\phi_{f,1})\,. \tag{B.8}$$

We want to show that $y_N \to 1$. Given any subsequence $\{y_{N'}\}_{N'\in\mathbb{N}}$ of $\{y_N\}_{N\in\mathbb{N}}$, there exists a subsequence $\{y_{N'_\ell}\}_{\ell\in\mathbb{N}}$ such that $y_{N'_\ell} \to y$. Clearly, $y \neq 0$. Effectively, if $y = 0$, $\gamma_{N'_\ell} \to 0$ and from the weak continuity of ς_f, we obtain that $\varsigma_f(\gamma_{N'_\ell}) \to \varsigma_f(0)$. However, $\varsigma_f(\gamma_N) \to \varsigma_f(\phi_{f,1}) > \varsigma_f(0)$ since f is strictly increasing and $\lambda_{f,i,1} > 0$ for some i, leading to a contradiction. Thus, $y_{N'_\ell} \to y \neq 0$. Using that the unit ball $\mathcal{V}_1$ is weakly compact and that $\gamma_{N'_\ell} \in \mathcal{V}_1$, there exists a subsequence such that $\gamma_{N'_{\ell_s}}$ converges weakly to $\gamma \in \mathcal{V}_1$. Note that $\gamma \neq 0$, since the weak continuity of ς_f implies that $\varsigma_f(\gamma_{N'_{\ell_s}}) \to \varsigma_f(\gamma)$ while on the other hand, $\varsigma_f(\gamma_{N'_{\ell_s}}) \to \varsigma_f(\phi_{f,1}) > \varsigma_f(0)$. On the other hand, $\|\gamma_{N'_{\ell_s}}\| \to y$, so the weak continuity of ς_f and the fact that $\gamma_{N'_{\ell_s}}/\|\gamma_{N'_{\ell_s}}\|$ converges weakly to γ/y implies that $\varsigma_f\left(\gamma_{N'_{\ell_s}}/\|\gamma_{N'_{\ell_s}}\|\right) \to \varsigma_f\left(\gamma/y\right)$ and $\varsigma_f\left(\gamma_{N'_{\ell_s}}\right) \to \varsigma_f\left(\gamma\right)$. Using (B.8), we get that

$\varsigma_f(\gamma/y) = \varsigma_f(\gamma) = \varsigma_f(\phi_{f,1})$, which implies that $y = 1$ since f is strictly increasing, concluding the proof of a).

b) The proof of $\langle \widehat{\phi}_1, \phi_{f,1}\rangle^2 \xrightarrow{a.s.} 1$ follows immediately from Lemma 4.1. Since both σ^2_{i,n_i} and σ^2_i are invariant under sign changes, we can assume that $\widehat{\phi}_1 \xrightarrow{a.s.} \phi_{f,1}$.

Using that $\|\widehat{\phi}_1\| \le 1$ and (B.2), we get that $\widehat{\lambda}_{i,1} - \sigma^2_i(\widehat{\phi}_1) = \sigma^2_{i,n_i}(\widehat{\phi}_1) - \sigma^2_i(\widehat{\phi}_1) \xrightarrow{a.s.} 0$. On the other hand, the fact that $\widehat{\phi}_1 \xrightarrow{a.s.} \phi_{f,1}$ together with the weak continuity of σ_i implies that $\sigma^2_i(\widehat{\phi}_1) \xrightarrow{a.s.} \sigma^2_i(\phi_{f,1}) = \lambda_{f,i,1}$, concluding the proof of b).

c) The proof will be carried in several steps. We begin proving that $\varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$ by proving that

(i)

$$\sup_{\|\alpha\|_\nu \le 1} |\varsigma_f(\pi_{m-1}\alpha) - \varsigma_N(\widehat{\pi}_{\nu,m-1}\alpha)| \xrightarrow{a.s.} 0. \tag{B.9}$$

(ii)

$$\varsigma_f(\phi_{f,m}) \ge \varsigma_N(\widehat{\phi}_m) + o_{\text{a.s.}}(1). \tag{B.10}$$

(iii)

$$\varsigma_f(\phi_{f,m}) \le \varsigma_N(\widehat{\phi}_m) + o_{\text{a.s.}}(1). \tag{B.11}$$

(i) We begin by showing (B.9). Note that

$$\sup_{\|\alpha\|_\nu \le 1} |\varsigma_f(\pi_{m-1}\alpha) - \varsigma_N(\widehat{\pi}_{\nu,m-1}\alpha)| \le \sup_{\|\alpha\|_\nu \le 1} |\varsigma_f(\pi_{m-1}\alpha) - \varsigma_f(\widehat{\pi}_{\nu,m-1}\alpha)| + \sup_{\|\alpha\|_\nu \le 1} |\varsigma_N(\widehat{\pi}_{\nu,m-1}\alpha) - \varsigma_f(\widehat{\pi}_{\nu,m-1}\alpha)|.$$

Using (B.3) and the fact that $\|\alpha\|_\nu \le 1$ implies $\|\widehat{\pi}_{\nu,m-1}\alpha\| \le 1$, we get that the second term on the right hand side of the above inequality converges to zero almost surely. Hence, we only have to prove that $\sup_{\|\alpha\|_\nu \le 1} |\varsigma_f(\pi_{m-1}\alpha) - \varsigma_f(\widehat{\pi}_{\nu,m-1}\alpha)| \xrightarrow{a.s.} 0$.

For any $\alpha \in \mathcal{H}_{\text{S}}$, we have that $\langle \alpha, \phi_{f,j}\rangle \phi_{f,j} = \langle \alpha, \phi_{f,j}\rangle(\phi_{f,j} - \widehat{\phi}_j) + \langle \alpha, \phi_{f,j}\rangle \widehat{\phi}_j$. Therefore, if $\|\alpha\|^2_\nu = \|\alpha\|^2 + \nu\Psi(\alpha) \le 1$, we get that

$$\begin{aligned}
|\langle \alpha, \phi_{f,j}\rangle \phi_{f,j} - \langle \alpha, \widehat{\phi}_j\rangle_\nu \widehat{\phi}_j| &\le \|\alpha\| \, \|\phi_{f,j} - \widehat{\phi}_j\| + \|\widehat{\phi}_j\| |\langle \alpha, \phi_{f,j}\rangle - \langle \alpha, \widehat{\phi}_j\rangle_\nu| \\
&\le \|\phi_{f,j} - \widehat{\phi}_j\| + |\langle \alpha, \phi_{f,j} - \widehat{\phi}_j\rangle - \nu\lceil \alpha, \widehat{\phi}_j\rceil| \\
&\le \|\phi_{f,j} - \widehat{\phi}_j\| + \left\{ \|\phi_{f,j} - \widehat{\phi}_j\| + (\nu\Psi(\alpha))^{\frac{1}{2}} (\nu\Psi(\widehat{\phi}_j))^{\frac{1}{2}} \right\} \\
&\le \|\phi_{f,j} - \widehat{\phi}_j\| + \left\{ \|\phi_{f,j} - \widehat{\phi}_j\| + (\nu\Psi(\widehat{\phi}_j))^{\frac{1}{2}} \right\} .
\end{aligned}$$

Using that for $1 \le j \le m-1$, $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{f,j}$ and $\nu\Psi(\widehat{\phi}_j) \xrightarrow{a.s.} 0$, we obtain that

$$\sup_{\|\alpha\|_\nu \le 1} \|\langle \alpha, \phi_{f,j}\rangle \phi_{f,j} - \langle \alpha, \widehat{\phi}_j\rangle_\nu \widehat{\phi}_j\| \xrightarrow{a.s.} 0\,,$$

which implies that

$$\sup_{\|\alpha\|_\nu \le 1} \|\widehat{\pi}_{\nu,m-1}\alpha - \pi_{m-1}\alpha\| \xrightarrow{a.s.} 0\;. \tag{B.12}$$

Thereafter, using that **C1** and **C2** entail that ς_f is weakly uniformly continuous over $\mathcal{V}_1$, we obtain that $\sup_{\|\alpha\|_\nu \le 1} |\varsigma_f(\pi_{m-1}\alpha) - \varsigma_f(\widehat{\pi}_{\nu,m-1}\alpha)| \xrightarrow{a.s.} 0$, concluding the proof of (B.9).

(ii) As in a), we will now prove that (B.10) holds. Recall that $\varsigma_f(\phi_{f,m}) = \sup_{\alpha\in\mathcal{S}_1\cap\mathcal{T}_{m-1}} \varsigma_f(\alpha)$. As in a), using that for any $c \geq 1$ we have $\varsigma_f(c\alpha) \geq \varsigma_f(\alpha)$, we get that for any $\alpha \in \mathcal{H}$ such that $\|\alpha\| \leq 1$, we have $\varsigma_f(\alpha/\|\alpha\|) \geq \varsigma_f(\alpha)$. Thus, $\sup_{\alpha\in\mathcal{S}_1\cap\mathcal{T}_{m-1}} \varsigma_f(\alpha) = \sup_{\alpha\in\{\|\alpha\|\leq 1\}\cap\mathcal{T}_{m-1}} \varsigma_f(\alpha)$. Moreover, for any $\alpha \in \mathcal{V}_1$, we have that $\|\pi_{m-1}\alpha\| \leq 1$ and $\pi_{m-1}\alpha \in \mathcal{T}_{m-1}$, so

$$\sup_{\alpha\in\{\|\alpha\|\leq 1\}\cap\mathcal{T}_{m-1}} \varsigma_f(\alpha) \geq \sup_{\|\alpha\|\leq 1} \varsigma_f(\pi_{m-1}\alpha) \geq \sup_{\alpha\in\mathcal{S}_1} \varsigma_f(\pi_{m-1}\alpha)\,. \tag{B.13}$$

Now using that $\widehat{\phi}_m/\|\widehat{\phi}_m\| \in \mathcal{S}_1$, we get that

$$\sup_{\alpha\in\{\|\alpha\|\leq 1\}\cap\mathcal{T}_{m-1}} \varsigma_f(\alpha) \geq \varsigma_f\left(\pi_{m-1}\left(\frac{\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right)\right) \geq \varsigma_f\left(\frac{1}{\|\widehat{\phi}_m\|}\pi_{m-1}\widehat{\phi}_m\right) \geq \varsigma_f(\pi_{m-1}\widehat{\phi}_m)$$

since $\|\widehat{\phi}_m\| \leq 1$. Summing up together, we have that $\varsigma_f(\phi_{f,m}) \geq \varsigma_f(\pi_{m-1}\widehat{\phi}_m)$. Using (B.9) and the fact that $\|\widehat{\phi}_m\|_\nu = 1$, we obtain that $b_m = \varsigma_f(\pi_{m-1}\widehat{\phi}_m) - \varsigma_N(\widehat{\pi}_{\nu,m-1}\widehat{\phi}_m) \xrightarrow{a.s.} 0$. Noticing that $\widehat{\pi}_{\nu,m-1}\widehat{\phi}_m = \widehat{\phi}_m$, we get that

$$\varsigma_f(\phi_{f,m}) \geq \varsigma_f(\pi_{m-1}\widehat{\phi}_m) = \varsigma_N(\widehat{\pi}_{\nu,m-1}\widehat{\phi}_m) + o_{\text{a.s.}}(1) = \varsigma_N(\widehat{\phi}_m) + o_{\text{a.s.}}(1)\,,$$

concluding the proof of (B.10).

(iii) Let us derive (B.11). Since $\phi_{f,m} \in \mathcal{H}_{\text{S}}$, we have that $1 \leq \|\phi_{f,m}\|_\nu < \infty$ and using that $\nu \xrightarrow{a.s.} 0$, we also have that $\|\phi_{f,m}\|_\nu \xrightarrow{a.s.} \|\phi_{f,m}\| = 1$. Hence,

$$\begin{aligned}\varsigma_N(\widehat{\phi}_m) &\geq \varsigma_N(\widehat{\phi}_m) - \rho\Psi(\widehat{\phi}_m) = \sup_{\|\alpha\|_\nu=1,\alpha\in\widehat{\mathcal{T}}_{\nu,m-1}} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \\ &\geq \varsigma_N\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu}\right) - \rho\Psi\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu}\right)\,.\end{aligned}$$

Using that $\varsigma_N(c\alpha) \geq \varsigma_N(\alpha)$ if $c \geq 1$ and the fact that $\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu \leq \|\phi_{f,m}\|_\nu$, we easily obtain that

$$\begin{aligned}\varsigma_N(\widehat{\phi}_m) &\geq \varsigma_N\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_\nu}\right) - \rho\Psi\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu}\right) \\ &\geq \varsigma_N\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_\nu}\right) - \rho\frac{\Psi\left(\widehat{\pi}_{\nu,m-1}\phi_{f,m}\right)}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu^2}\,.\end{aligned}$$

Note that (B.9) entail that $c_m = \varsigma_f(\pi_{m-1}\phi_{f,m}/\|\phi_{f,m}\|_\nu) - \varsigma_N(\widehat{\pi}_{\nu,m-1}\phi_{f,m}/\|\phi_{f,m}\|_\nu) \xrightarrow{a.s.} 0$, thus,

$$\begin{aligned}\varsigma_N(\widehat{\phi}_m) &\geq \varsigma_N\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_\nu}\right) - \rho\Psi\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu}\right) \\ &\geq \varsigma_f\left(\frac{\pi_{m-1}\phi_{f,m}}{\|\phi_{f,m}\|_\nu}\right) + o_{\text{a.s.}}(1) - \rho\frac{\Psi\left(\widehat{\pi}_{\nu,m-1}\phi_{f,m}\right)}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu^2}\,.\end{aligned}$$

On the other hand, the weak continuity of ς_f, the fact that $\|\phi_{f,m}\|_\nu \xrightarrow{a.s.} \|\phi_{f,m}\| = 1$ and $\pi_{m-1}\phi_{f,m} = \phi_{f,m}$, imply that

$$\varsigma_N(\widehat{\phi}_m) \geq \varsigma_f\left(\phi_{f,m}\right) + o_{\text{a.s.}}(1) - \frac{\rho\Psi\left(\widehat{\pi}_{\nu,m-1}\phi_{f,m}\right)}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu^2}\,. \tag{B.14}$$

Using that $\nu\Psi(\widehat{\phi}_j) \xrightarrow{a.s.} 0$ and $\rho\Psi(\widehat{\phi}_j) \xrightarrow{a.s.} 0$, analogous arguments to those considered in Pezzulli and Silverman (1993), allow to show that

$$\rho\Psi(\widehat{\pi}_{\nu,m-1}\phi_{f,m}) = \rho\lceil\widehat{\pi}_{\nu,m-1}\phi_{f,m}, \widehat{\pi}_{\nu,m-1}\phi_{f,m}\rceil \xrightarrow{a.s.} 0\,, \tag{B.15}$$

$$\nu\Psi(\widehat{\pi}_{\nu,m-1}\phi_{f,m}) = \nu\lceil\widehat{\pi}_{\nu,m-1}\phi_{f,m}, \widehat{\pi}_{\nu,m-1}\phi_{f,m}\rceil \xrightarrow{a.s.} 0\,. \tag{B.16}$$

Besides, as in the proof of (B.12), we have that

$$\begin{aligned}\|\langle\phi_{f,m},\phi_{f,j}\rangle\phi_{f,j}-\langle\phi_{f,m},\widehat{\phi}_j\rangle_\nu\widehat{\phi}_j\| &\leq \|\phi_{f,j}-\widehat{\phi}_j\|+\|\widehat{\phi}_j\||\langle\phi_{f,m},\phi_{f,j}\rangle-\langle\phi_{f,m},\widehat{\phi}_j\rangle_\nu|| \\ &\leq \|\phi_{f,j}-\widehat{\phi}_j\|+|\langle\phi_{f,m},\phi_{f,j}-\widehat{\phi}_j\rangle-\nu\lceil\phi_{f,m},\widehat{\phi}_j\rceil| \\ &\leq \|\phi_{f,j}-\widehat{\phi}_j\|+\left\{\|\phi_{f,j}-\widehat{\phi}_j\|+(\nu\Psi(\phi_{f,m}))^{\frac{1}{2}}(\nu\Psi(\widehat{\phi}_j))^{\frac{1}{2}}\right\}\,.\end{aligned}$$

Using that for $1\leq j\leq m-1$, $\widehat{\phi}_j\xrightarrow{a.s.}\phi_{f,j}$, $\nu\Psi(\widehat{\phi}_j)\xrightarrow{a.s.}0$ and $\nu\Psi(\phi_{f,m})\xrightarrow{a.s.}0$, we obtain that $\|\langle\phi_{f,m},\phi_{f,j}\rangle\phi_{f,j}-\langle\phi_{f,m},\widehat{\phi}_j\rangle_\nu\widehat{\phi}_j\|\xrightarrow{a.s.}0$. The fact that $\langle\phi_{f,m},\phi_{f,j}\rangle=0$ for $j<m$ and that $\|\phi_{f,m}\|=1$ entails that $\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|=\|\phi_{f,m}-\sum_{j=1}^{m-1}\langle\phi_{f,m},\widehat{\phi}_j\rangle_\nu\widehat{\phi}_j\|\xrightarrow{a.s.}1$, which together with (B.16) implies $\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu^2=\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|^2+\nu\Psi(\widehat{\pi}_{\nu,m-1}\phi_{f,m})\xrightarrow{a.s.}1$. Therefore, (B.14) together with (B.15) entail that $\varsigma_N(\widehat{\phi}_m)\geq\varsigma_f(\phi_{f,m})+o_{\text{a.s.}}(1)$, concluding the proof of (B.11).

Note that (B.10) and (B.11) imply that

$$\varsigma_N(\widehat{\phi}_m)\xrightarrow{a.s.}\varsigma_f(\phi_{f,m}),\tag{B.17}$$

as desired. On the other hand, the fact that $\|\widehat{\phi}_m\|\leq 1$ and (B.3) entail that $\varsigma_N(\widehat{\phi}_m)-\varsigma_f(\widehat{\phi}_m)\xrightarrow{a.s.}0$, which together with (B.17) lead us to $\varsigma_f(\widehat{\phi}_m)\xrightarrow{a.s.}\varsigma_f(\phi_{f,m})$.

(iv) We will show that $\rho\Psi(\widehat{\phi}_m)\xrightarrow{a.s.}0$. Note that the above inequalities entail that

$$\varsigma_N(\widehat{\phi}_m)\geq\varsigma_N(\widehat{\phi}_m)-\rho\Psi(\widehat{\phi}_m)\geq\varsigma_f(\phi_{f,m})+o_{\text{a.s.}}(1)\,.$$

Hence, (B.17) entail that $0\leq\rho\Psi(\widehat{\phi}_m)\leq\varsigma_N(\widehat{\phi}_m)-\varsigma_f(\phi_{f,m})+o_{\text{a.s.}}(1)\xrightarrow{a.s.}0$, as desired.

(v) To conclude the proof of c) it remains to show that

$$\nu\Psi(\widehat{\phi}_m)\xrightarrow{a.s.}0.\tag{B.18}$$

Recall that as shown above

$$\varsigma_f(\phi_{f,m})\geq\varsigma_f\left(\frac{\pi_{m-1}\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right)\geq\varsigma_f\left(\pi_{m-1}\widehat{\phi}_m\right)=\varsigma_N(\widehat{\phi}_m)+o_{\text{a.s.}}(1)\,.$$

Hence,

$$\varsigma_f\left(\frac{\pi_{m-1}\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right)\xrightarrow{a.s.}\varsigma_f(\phi_{f,m})\qquad\text{and}\qquad\varsigma_f\left(\pi_{m-1}\widehat{\phi}_m\right)\xrightarrow{a.s.}\varsigma_f(\phi_{f,m})\,.$$

Now the proof follows as in a) taking $\gamma_N=\pi_{m-1}\widehat{\phi}_m(\omega)$, $y_N=\|\widehat{\phi}_m(\omega)\|\leq 1$ for $\omega\notin\mathcal{N}$ with $\mathcal{N}=\{\omega\in\Omega:\varsigma_f(\widehat{\phi}_m(\omega))\not\to\varsigma_f(\phi_{f,m}),\varsigma_f\left(\pi_{m-1}\widehat{\phi}_m(\omega)/\|\widehat{\phi}_m(\omega)\|\right)\not\to\varsigma_f(\phi_{f,m})\text{ and }\varsigma_f\left(\pi_{m-1}\widehat{\phi}_m(\omega)\right)\not\to\varsigma_f(\phi_{f,m})\}$ and using that the eigenvalues are ordered in a strictly decreasing order and $\lambda_{f,i,m}>0$ for some i, since $m<q$ and **C0** holds.

d) We have already proved that when $m=1$ the result holds. We proceed by induction and assume that $\langle\widehat{\phi}_\ell,\phi_{f,\ell}\rangle^2\to 1$, $\nu\Psi(\widehat{\phi}_\ell)\xrightarrow{a.s.}0$ and $\rho\Psi(\widehat{\phi}_\ell)\xrightarrow{a.s.}0$ for $1\leq\ell\leq m-1$, to show that $\langle\widehat{\phi}_m,\phi_{f,m}\rangle^2\to 1$. Without loss of generality, we can assume that $\widehat{\phi}_\ell\xrightarrow{a.s.}\phi_{f,\ell}$, for $1\leq\ell\leq m-1$. Using c) we have that $\varsigma_f(\widehat{\phi}_m)\xrightarrow{a.s.}\varsigma_f(\phi_{\text{R},m})$ and that $\|\widehat{\phi}_m\|\xrightarrow{a.s.}1$ and so, from Lemma 4.1 we get that $\langle\widehat{\phi}_m,\phi_{f,m}\rangle^2\xrightarrow{a.s.}1$. Without loss of generality we can assume that $\widehat{\phi}_m\xrightarrow{a.s.}\phi_{f,m}$, since σ_{i,n_i}^2 and σ_i^2 are invariant under sign changes. Hence, as in b) using that $\|\widehat{\phi}_m\|\leq 1$ and (B.2), we get that $\widehat{\lambda}_{i,m}-\sigma_i^2(\widehat{\phi}_m)=\sigma_{i,n_i}^2(\widehat{\phi}_m)-\sigma_i^2(\widehat{\phi}_m)\xrightarrow{a.s.}0$. On the other hand, the fact that $\widehat{\phi}_m\xrightarrow{a.s.}\phi_{f,m}$ together with the weak continuity of σ_i implies that $\sigma_i^2(\widehat{\phi}_m) \xrightarrow{a.s.} \sigma_i^2(\phi_{f,m}) = \lambda_{f,i,m}$, concluding the proof of d). $\square$

Proof of Theorem 4.2. First note that as in Theorem 4.1 from assumption ii) and the fact that $\|\alpha\| \le \|\alpha\|_\nu$ we get that (B.2) holds.

a) As in Theorem 4.1, to prove that $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$, it is enough to show

$$\begin{aligned} \varsigma_f(\phi_{f,1}) &\ge \varsigma_N(\widehat{\phi}_1) + o_{\text{a.s.}}(1)\ , \qquad (B.19)\\ \varsigma_f(\phi_{f,1}) &\le \varsigma_N(\widehat{\phi}_1) + o_{\text{a.s.}}(1)\ . \qquad (B.20)\end{aligned}$$

Using (B.2), we get that $\widehat{a}_{N,i} = \sigma^2_{i,n_i}(\widehat{\phi}_1) - \sigma_i^2(\widehat{\phi}_1) \xrightarrow{a.s.} 0$, $\widehat{b}_{N,i} = \sigma^2_{i,n_i}(\phi_{f,1}) - \sigma_i^2(\phi_{f,1}) \xrightarrow{a.s.} 0$. On the other hand, using that $\sigma_i^2(\phi_{f,1}) > 0$ we get that $\log(\sigma^2_{i,n_i}(\phi_{f,1})) \xrightarrow{a.s.} \log\left(\sigma_i^2(\phi_{f,1})\right)$ which implies that $\varsigma_N(\phi_{f,1}) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})$. Besides, using that $\nu \xrightarrow{a.s.} 0$ and $\|\phi_{f,1}\|_\nu \xrightarrow{a.s.} 1$ we get that

$$\log\left(\sigma^2_{i,n_i}\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right)\right) = \log\left(\frac{\sigma^2_{i,n_i}(\phi_{f,1})}{\|\phi_{f,1}\|^2_\nu}\right) = \log\left(\sigma^2_{i,n_i}(\phi_{f,1})\right) - 2\log\left(\|\phi_{f,1}\|_\nu\right) \xrightarrow{a.s.} \log\left(\sigma_i^2(\phi_{f,1})\right)$$

so,

$$\varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1})\ . \qquad (B.21)$$

To derive (B.20), note that since $\phi_{f,1} \in \mathcal{H}_{\text{S}}$, $\|\phi_{f,1}\|_\nu < \infty$ and $\|\phi_{f,1}\|_\nu \ge \|\phi_{f,1}\| = 1$. Then, using that $\widehat{\phi}_1 = \text{argmax}_{\|\alpha\|_\nu=1}\{\varsigma_N(\alpha) - \rho\Psi(\alpha)\}$ and defining $\beta_1 = \phi_{f,1}/\|\phi_{f,1}\|_\nu$, we have that $\|\beta_1\|_\nu = 1$ and

$$\varsigma_N(\widehat{\phi}_1) \ge \varsigma_N(\widehat{\phi}_1) - \rho\Psi(\widehat{\phi}_1) \ge \varsigma_N(\beta_1) - \rho\Psi(\beta_1) = \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) - \rho\Psi\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right)\ .$$

Hence, using that $\Psi(a\alpha) = a^2\Psi(\alpha)$, for any $a \in \mathbb{R}$, and (B.21), we get

$$\begin{aligned} \varsigma_N(\widehat{\phi}_1) &\ge \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) - \rho\Psi\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) = \varsigma_N\left(\frac{\phi_{f,1}}{\|\phi_{f,1}\|_\nu}\right) - \rho\frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|^2_\nu}\\ &\ge \varsigma_f(\phi_{f,1}) + o_{\text{a.s.}}(1) - \rho\frac{\Psi(\phi_{f,1})}{\|\phi_{f,1}\|^2_\nu}\ . \qquad (B.22)\end{aligned}$$

When $\rho = 0$, we have defined $\rho\Psi(\phi_{f,1}) = 0$ and similarly when $\nu = 0$ in which case $\|\phi_{f,1}\|_\nu = \|\phi_{f,1}\| = 1$. So from now on, we will assume that $\nu_N > 0$ and $\rho_N > 0$. Since $\nu \xrightarrow{a.s.} 0$, we have that and $\|\phi_{f,1}\|_\nu \xrightarrow{a.s.} \|\phi_{f,1}\| = 1$. Hence, using that $\rho \xrightarrow{a.s.} 0$, and so, $\rho\Psi(\phi_{f,1})/\|\phi_{f,1}\|^2_\nu \xrightarrow{a.s.} 0$, we get (B.20).

Let $\omega \notin \mathcal{N}$ where $\mathcal{N}$ is the set of probability 0 where the almost sure convergences of assumptions (ii) to (iv) do not hold and let us show first that, for any $1 \le i \le k$, $\liminf \sigma_i^2(\widehat{\phi}_1(\omega)) > 0$. Effectively, assume there exists i_0 such that $\liminf \sigma^2_{i_0}(\widehat{\phi}_1(\omega)) = 0$. Then, there exists a subsequence of $\gamma_N = \widehat{\phi}_1(\omega)$ such that $\sigma^2_{i_0}(\gamma_{N_\ell}) \to 0$, using that $\widehat{a}_{N,i_0} = \sigma^2_{i_0,n_{i_0}}(\widehat{\phi}_1(\omega)) - \sigma^2_{i_0}(\widehat{\phi}_1(\omega)) \to 0$, we get that $\sigma^2_{i_0,\widehat{\tau}_{i_0}N_\ell}(\gamma_{N_\ell}) \to 0$ and so, $\varsigma_{N_\ell}(\gamma_{N_\ell}) \to -\infty$. Thus, using (B.22), we get that $\varsigma_f(\phi_{f,1}) = -\infty$ which contradicts the fact that $\lambda_{f,i,1} > 0$. Hence, for any $1 \le i \le k$, $\liminf \sigma_i^2(\widehat{\phi}_1(\omega)) > 0$, thus there exists $\epsilon > 0$ such that $\sigma_i^2(\widehat{\phi}_1(\omega)) \in [\epsilon, +\infty)$. Moreover, using that σ_i is weakly continuous and the unit ball is weakly compact, we have that $\{\sigma_i(\alpha) : \alpha \in \mathcal{V}_1\}$ is a bounded set. Hence there exists $\epsilon > 0$ and $A > 0$ such that $\sigma_i^2(\widehat{\phi}_1(\omega)) \in [\epsilon, A]$, for any $1 \le i \le k$. Using that

$\widehat{a}_{N,i}(\omega) = \sigma^2_{i,n_i}(\widehat{\phi}_1(\omega)) - \sigma^2_i(\widehat{\phi}_1(\omega)) \to 0$, we get that for N large enough, $\sigma^2_{i,n_i}(\widehat{\phi}_1(\omega)) \in [\epsilon/2, 2A]$ which together with the uniform continuity of the logarithm on $[\epsilon/2, 2A]$ entails that

$$\widehat{a}_N = \varsigma_N(\widehat{\phi}_1) - \varsigma_f(\widehat{\phi}_1) \xrightarrow{a.s.} 0 \, . \tag{B.23}$$

Using that σ_{R} is a scale functional, f is strictly increasing, $\varsigma_f(\phi_{f,1}) = \sup_{\alpha\in\mathcal{S}} \varsigma_f(\alpha)$ and the fact that $\|\widehat{\phi}_1\| \le \|\widehat{\phi}_1\|_\nu = 1$, we obtain easily that

$$\varsigma_f(\phi_{f,1}) \ge \varsigma_f\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \ge \varsigma_f(\widehat{\phi}_1) = \varsigma_N(\widehat{\phi}_1) - \widehat{a}_N = \varsigma_N(\widehat{\phi}_1) + o_{\mathrm{a.s.}}(1)\,,$$

concluding the proof of (B.19) and so, $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f\,(\phi_{f,1})$.

As mentioned above, from (B.23), and using that $\varsigma_N(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f\,(\phi_{f,1})$, we get that

$$\varsigma_f(\widehat{\phi}_1) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1}) \, . \tag{B.24}$$

Moreover, the inequalities $\varsigma_f(\phi_{f,1}) \ge \varsigma_f\left(\widehat{\phi}_1/\|\widehat{\phi}_1\|\right) \ge \varsigma_f(\widehat{\phi}_1)$, obtained above, imply that

$$\varsigma_f\left(\frac{\widehat{\phi}_1}{\|\widehat{\phi}_1\|}\right) \xrightarrow{a.s.} \varsigma_f(\phi_{f,1}) \, . \tag{B.25}$$

We have to show that $\rho\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$, which follows as in the proof of Theorem 4.1.

It remains to show that $\nu\Psi(\widehat{\phi}_1) \xrightarrow{a.s.} 0$. Using that $\|\widehat{\phi}_1\|_\nu = 1$, we get that $\nu\Psi(\widehat{\phi}_1) = 1 - \|\widehat{\phi}_1\|^2$. Hence, we only have to show that $\|\widehat{\phi}_1\| \xrightarrow{a.s.} 1$. The fact that $\varsigma_f\left(\widehat{\phi}_1/\|\widehat{\phi}_1\|\right) = \varsigma_f(\widehat{\phi}_1) - \log(\|\widehat{\phi}_1\|^2)$ leads to $\varsigma_f(\widehat{\phi}_1) - \varsigma_f\left(\widehat{\phi}_1/\|\widehat{\phi}_1\|\right) = \log(\|\widehat{\phi}_1\|^2)$. Using (B.24) and (B.25), we get that $\log(\|\widehat{\phi}_1\|^2) \xrightarrow{a.s.} 0$, so $\|\widehat{\phi}_1\|^2 \xrightarrow{a.s.} 1$ concluding the proof of a).

b) The proof of $\langle\widehat{\phi}_1, \phi_{f,1}\rangle^2 \xrightarrow{a.s.} 1$ follows immediately from Lemma 4.2, while the fact that $\widehat{\lambda}_{i,1} \xrightarrow{a.s.} \lambda_{f,i,1}$ follows as in Theorem 4.1.

c) First note that as in the proof of Theorem 4.1, using that for $1 \le j \le m-1$, $\widehat{\phi}_j \xrightarrow{a.s.} \phi_{f,j}$ and $\nu\Psi(\widehat{\phi}_j) \xrightarrow{a.s.} 0$, we obtain that (B.12) holds.

Using (B.2), (B.12) and the fact that $\|\widehat{\phi}_m\|_\nu = 1$ and $\|\phi_{f,m}\| = 1$, we get that $\widehat{a}_{N,i} = \sigma^2_{i,n_i}(\widehat{\phi}_m) - \sigma^2_i(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, $\widehat{b}_{N,i} = \sigma^2_{i,n_i}(\pi_{m-1}\widehat{\phi}_m) - \sigma^2_i(\pi_{m-1}\widehat{\phi}_m) \xrightarrow{a.s.} 0$, $\widehat{c}_{N,i} = \sigma^2_{i,n_i}(\phi_{f,m}) - \sigma^2_i(\phi_{f,m}) \xrightarrow{a.s.} 0$ and $\widehat{A}_{N,i} = \sigma^2_{i,n_i}(\widehat{\pi}_{\nu,m-1}\phi_{f,m}) - \sigma^2_i(\phi_{f,m}) \xrightarrow{a.s.} 0$. Moreover, using (B.12), we get that

$$\|\pi_{m-1}\widehat{\phi}_m - \widehat{\phi}_m\| = \|\pi_{m-1}\widehat{\phi}_m - \widehat{\pi}_{\nu,m-1}\widehat{\phi}_m\| \le \sup_{\|\alpha\|_\nu \le 1} \|\pi_{m-1}\alpha - \widehat{\pi}_{\nu,m-1}\alpha\| \xrightarrow{a.s.} 0.$$

Thus, $\pi_{m-1}\widehat{\phi}_m - \widehat{\phi}_m \xrightarrow{a.s.} 0$. Using that $\pi_{m-1}\widehat{\phi}_m \in \mathcal{V}_1$ and $\widehat{\phi}_m \in \mathcal{V}_1$ and the fact that σ_i is uniformly weakly continuous in $\mathcal{V}_1$, we get that $\widehat{B}_{N,i} = \sigma^2_i(\pi_{m-1}\widehat{\phi}_m) - \sigma^2_i(\widehat{\phi}_m) \xrightarrow{a.s.} 0$.

Arguing as in a), and using that, for all $1 \le i \le k$, $\sigma^2_i(\phi_{f,m}) > 0$ we get that

$$\widehat{a}_N = \varsigma_N(\widehat{\phi}_m) - \varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} 0\,, \tag{B.26}$$
$$\widehat{A}_N = \varsigma_N(\widehat{\pi}_{\nu,m-1}\phi_{f,m}) - \varsigma_f(\phi_{f,m}) \xrightarrow{a.s.} 0\,, \tag{B.27}$$
$$\widehat{b}_N = \varsigma_N(\pi_{m-1}\widehat{\phi}_m) - \varsigma_f(\pi_{m-1}\widehat{\phi}_m) \xrightarrow{a.s.} 0\,, \tag{B.28}$$
$$\widehat{B}_N = \varsigma_f(\pi_{m-1}\widehat{\phi}_m) - \varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} 0\,. \tag{B.29}$$

Moreover, we also obtain that $\log(\sigma^2_{i,n_i}(\phi_{f,m})) \xrightarrow{a.s.} \log\left(\sigma^2_i(\phi_{f,m})\right)$ which implies that $\varsigma_N(\phi_{f,m}) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$.

As in a) and as in Theorem 4.1, we will first show that

$$\varsigma_f(\phi_{f,m}) \geq \varsigma_N(\widehat{\phi}_m) + o_{\text{a.s.}}(1) \,, \tag{B.30}$$
$$\varsigma_f(\phi_{f,m}) \leq \varsigma_N(\widehat{\phi}_m) + o_{\text{a.s.}}(1) \,, \tag{B.31}$$

which ensures that $\varsigma_N(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$, as desired.

Let us show (B.30). As in the proof of Theorem 4.1 (see (c ii)), we have that $\varsigma_f(\phi_{f,m}) = \sup_{\alpha\in\mathcal{S}_1\cap\mathcal{T}_{m-1}} \varsigma_f(\alpha) = \sup_{\alpha\in\mathcal{V}_1\cap\mathcal{T}_{m-1}} \varsigma_f(\alpha)$. Therefore, using that $\|\widehat{\phi}_m\| \leq 1$ and that $f(t) = \log(t)$ is increasing, we get

$$\varsigma_f(\phi_{f,m}) = \sup_{\alpha\in\mathcal{V}_1\cap\mathcal{T}_{m-1}} \varsigma_f(\alpha) \geq \varsigma_f\left(\pi_{m-1}\left(\frac{\widehat{\phi}_m}{\|\widehat{\phi}_m\|}\right)\right) \geq \varsigma_f\left(\frac{1}{\|\widehat{\phi}_m\|}\pi_{m-1}\widehat{\phi}_m\right) \geq \varsigma_f(\pi_{m-1}\widehat{\phi}_m)\,.$$

On the other hand, (B.26) and (B.29) entail that $\varsigma_f(\pi_{m-1}\widehat{\phi}_m) = \varsigma_N(\widehat{\phi}_m) - \widehat{a}_N - \widehat{B}_N = \varsigma_N(\widehat{\phi}_m) + o_{\text{a.s.}}(1)$ concluding the proof of (B.30).

We now proceed to prove (B.31). As in Theorem 4.1 (see (c iii)), we have that

$$\begin{aligned}
\varsigma_N(\widehat{\phi}_m) &\geq \varsigma_N(\widehat{\phi}_m) - \rho\Psi(\widehat{\phi}_m) = \sup_{\|\alpha\|_\nu=1,\alpha\in\widehat{\mathcal{T}}_{\nu,m-1}} \{\varsigma_N(\alpha) - \rho\Psi(\alpha)\} \\
&\geq \varsigma_N\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu}\right) - \rho\Psi\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|_\nu}\right) \\
&\geq \varsigma_N\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_\nu}\right) - \rho\frac{\Psi\left(\widehat{\pi}_{\nu,m-1}\phi_{f,m}\right)}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|^2_\nu}\,.
\end{aligned}$$

Note that $\varsigma_N\left(\widehat{\pi}_{\nu,m-1}\phi_{f,m}/\|\phi_{f,m}\|_\nu\right) = \varsigma_N\left(\widehat{\pi}_{\nu,m-1}\phi_{f,m}\right) - 2\log\left(\|\phi_{f,m}\|_\nu\right)$, so that (B.27) and the fact that $\|\phi_{f,m}\|_\nu \xrightarrow{a.s.} 1$ entails that

$$\begin{aligned}
\varsigma_N(\widehat{\phi}_m) &\geq \varsigma_N\left(\frac{\widehat{\pi}_{\nu,m-1}\phi_{f,m}}{\|\phi_{f,m}\|_\nu}\right) - \rho\frac{\Psi\left(\widehat{\pi}_{\nu,m-1}\phi_{f,m}\right)}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|^2_\nu} \\
&\geq \varsigma_f\left(\phi_{f,m}\right) - \widehat{A}_N - 2\log\left(\|\phi_{f,m}\|_\nu\right) - \rho\frac{\Psi\left(\widehat{\pi}_{\nu,m-1}\phi_{f,m}\right)}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|^2_\nu} \\
&\geq \varsigma_f\left(\phi_{f,m}\right) + o_{\text{a.s.}}(1) - \rho\frac{\Psi\left(\widehat{\pi}_{\nu,m-1}\phi_{f,m}\right)}{\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|^2_\nu}\,.
\end{aligned}$$

Hence, the proof of (B.31) follows now as in Theorem 4.1 from (B.15), (B.16) and the fact that $\|\widehat{\pi}_{\nu,m-1}\phi_{f,m}\|^2_\nu \xrightarrow{a.s.} 1$.

From $\varsigma_N(\widehat{\phi}_N) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$ and $\varsigma_N(\widehat{\phi}_m) - \varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} 0$, see (B.26), we easily get that $\varsigma_f(\widehat{\phi}_m) \xrightarrow{a.s.} \varsigma_f(\phi_{f,m})$.

It remains to show that $\rho\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ and $\nu\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$. The proof of $\rho\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ follows as in (c iv) in the proof of Theorem 4.1, while that of $\nu\Psi(\widehat{\phi}_m) \xrightarrow{a.s.} 0$ follows using analogous arguments to those considered in STEP 5 in the proof of Theorem 4.1

d) The proof follows as that of Theorem 4.1 using Lemma 4.2 instead of Lemma 4.1. □

References

[1] Bali, J. L. (2012). *Métodos Robustos de estimación de componentes principales funcionales y el modelo de componentes principales comunes*. PhD. Thesis (in spanish), Universidad de Buenos Aires. Available on `http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_5153_Bali.pdf`.

[2] Bali, J. L. and Boente, G. (2009). Principal points and elliptical distributions from the multivariate setting to the functional case. *Statist. Probab. Lett.*, **79**, 1858-1865.

[3] Bali, J. L., Boente, G., Tyler, D. E. and Wang, J. L. (2011). Robust Functional Principal Components: a projection-pursuit approach. *Annals of Statistics*, **39**,2852-2882.

[4] Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band–spectroscopic data. *Technometrics*, **16**, 147-185.

[5] Benko, M.; Härdle, P. and Kneip, A. (2009). Common Functional Principal Components. *Ann. Statist.*, **37**, 1-34.

[6] Boente, G. and Fraiman, R. (2000). Kernel-based functional principal components. *Statist. Probab. Lett.*, **48**, 335-345.

[7] Boente, G. and Orellana L.C. (2001), A robust approach to common principal components. *Statistics in Genetics and in Environmental Sciences*. pp 117-46.

[8] Boente G., Pires A. and Rodrigues I. (2006) General projection-pursuit estimators for the common principal components model: influence functions and Monte Carlo study. *Journal of Multivariate Analysis*, **97**, 124-147.

[9] Boente, G., Rodriguez, D. and Sued, M. (2010). Inference under functional proportional and common principal components models. *Journal of Multivariate Analysis*, **101**, 464-475.

[10] Croux, C., Filzmoser, P. and Oliveira M.R. (2007). Algorithms for Projection-Pursuit Robust Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **87**, 218-225.

[11] Croux, C. and Ruiz–Gazen, A. (1996). A fast algorithm for robust principal components based on projection pursuit. In *Compstat: Proceedings in Computational Statistics*, ed. A. Prat, Heidelberg: Physica–Verlag, 211-217.

[12] Croux, C. and Ruiz–Gazen, A. (2005). High Breakdown Estimators for Principal Components: the Projection–Pursuit Approach Revisited. *J. Multivar. Anal.*, **95**, 206-226.

[13] Dauxois, J., Pousse, A. and 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.

[14] Flury, B. (1984) Common principal components in k groups. *Journal of the American Statistical Association* **79**, 892-898.

[15] Gervini, D. (2006). Free-knot spline smoothing for functional data. *J. Roy. Statist. Soc. Ser. B*, **68**, 671-87.

[16] Gervini, D. (2008). Robust functional estimation using the spatial median and spherical principal components. *Biometrika*, **95**, 587-600.

[17] Hampel, F. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.*, **42**, 1887-1896.

[18] Hyndman, R. J. and S. Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Comp. Statist. Data Anal.*, **51**, 4942-4956.

[19] Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. and Cohen, K. L. (1999). Robust principal components for functional data (with Discussion). *Test*, **8**, 1-73.

[20] Li, G. and 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.

[21] Pezzulli, S. D. and Silverman, B. W. (1993). Some properties of smoothed principal components analysis for functional data. *Comput. Statist.*, **8**, 1-16.

[22] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. New York: Springer-Verlag.

[23] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer-Verlag.

[24] Rice, J. and 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.

[25] Sawant, P., Billor, N. and Shin, H. (2011) Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, **27**, 83-102.

[26] Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.*, **24**, 1-24.

Supplementary file

Scale estimator	a	$\widehat{\phi}_{\mathrm{PS},j}$			$\widehat{\phi}_{\mathrm{PN},j}$		
		$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
SD	0	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M-scale	0	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.1	0.0032	0.0033	0.0033	0.0025	0.0035	0.0055
MAD	0.1	0.0418	0.0494	0.0231	0.0291	0.0439	0.0664
M-scale	0.1	0.0103	0.0111	0.0051	0.0068	0.0096	0.0093
SD	0.5	0.0029	0.0029	0.0032	0.0017	0.0053	0.0220
MAD	0.5	0.0384	0.0384	0.0110	0.0144	0.0475	0.1448
M-scale	0.5	0.0089	0.0088	0.0038	0.0033	0.0105	0.0275
SD	1	0.0027	0.0028	0.0032	0.0015	0.0085	0.0534
MAD	1	0.0333	0.0326	0.0074	0.0091	0.0541	0.1986
M-scale	1	0.0079	0.0078	0.0043	0.0023	0.0137	0.0613
SD	1.5	0.0026	0.0026	0.0032	0.0014	0.0122	0.0863
MAD	1.5	0.0307	0.0300	0.0060	0.0069	0.0605	0.2464
M-scale	1.5	0.0070	0.0070	0.0035	0.0021	0.0180	0.0939
SD	2	0.0024	0.0025	0.0032	0.0014	0.0158	0.1165
MAD	2	0.0271	0.0265	0.0053	0.0057	0.0657	0.2738
M-scale	2	0.0064	0.0064	0.0034	0.0019	0.0221	0.1270
SD	2.5	0.0023	0.0024	0.0032	0.0014	0.0195	0.1447
MAD	2.5	0.0255	0.0250	0.0049	0.0049	0.0701	0.3050
M-scale	2.5	0.0059	0.0059	0.0034	0.0018	0.0257	0.1576
SD	3	0.0022	0.0023	0.0032	0.0014	0.0230	0.1688
MAD	3	0.0240	0.0235	0.0046	0.0046	0.0784	0.3341
M-scale	3	0.0053	0.0054	0.0034	0.0017	0.0290	0.1849
SD	3.5	0.0021	0.0022	0.0032	0.0013	0.0261	0.1921
MAD	3.5	0.0222	0.0217	0.0044	0.0043	0.0832	0.3537
M-scale	3.5	0.0050	0.0051	0.0034	0.0017	0.0326	0.2061
SD	4	0.0020	0.0022	0.0032	0.0013	0.0293	0.2132
MAD	4	0.0212	0.0208	0.0044	0.0039	0.0886	0.3784
M-scale	4	0.0046	0.0047	0.0034	0.0016	0.0366	0.2265
SD	4.5	0.0020	0.0021	0.0032	0.0013	0.0325	0.2326
MAD	4.5	0.0201	0.0198	0.0042	0.0036	0.0949	0.3994
M-scale	4.5	0.0044	0.0045	0.0034	0.0016	0.0398	0.2430
SD	5	0.0019	0.0021	0.0032	0.0013	0.0362	0.2481
MAD	5	0.0187	0.0184	0.0041	0.0035	0.1017	0.4205
M-scale	5	0.0042	0.0043	0.0034	0.0016	0.0428	0.2607
SD	5.5	0.0019	0.0020	0.0032	0.0013	0.0395	0.2630
MAD	5.5	0.0173	0.0171	0.0040	0.0034	0.1028	0.4300
M-scale	5.5	0.0040	0.0041	0.0034	0.0016	0.0460	0.2759

Table 6: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under C_0 when ρ or $\nu = aN^{-2}$ and $f = \mathrm{id}$.

Scale	a	$\widehat{\phi}_{\mathrm{PS},j}$			$\widehat{\phi}_{\mathrm{PN},j}$		
		$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
SD	0	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M-scale	0	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.1	0.0033	0.0039	0.0037	0.0033	0.0038	0.0037
MAD	0.1	0.0436	0.0634	0.0368	0.0436	0.0637	0.0377
M-scale	0.1	0.0108	0.0140	0.0077	0.0108	0.0140	0.0076
SD	0.15	0.0033	0.0039	0.0037	0.0033	0.0038	0.0037
MAD	0.15	0.0436	0.0634	0.0368	0.0436	0.0635	0.0377
M-scale	0.15	0.0108	0.0140	0.0076	0.0107	0.0139	0.0076
SD	0.25	0.0033	0.0038	0.0037	0.0033	0.0038	0.0037
MAD	0.25	0.0436	0.0633	0.0368	0.0434	0.0628	0.0376
M-scale	0.25	0.0108	0.0140	0.0076	0.0107	0.0138	0.0076
SD	0.5	0.0033	0.0038	0.0037	0.0032	0.0038	0.0037
MAD	0.5	0.0436	0.0632	0.0367	0.0423	0.0611	0.0376
M-scale	0.5	0.0108	0.0140	0.0076	0.0106	0.0136	0.0076
SD	0.75	0.0033	0.0038	0.0037	0.0032	0.0037	0.0037
MAD	0.75	0.0436	0.0629	0.0364	0.0419	0.0606	0.0381
M-scale	0.75	0.0108	0.0139	0.0075	0.0106	0.0134	0.0076
SD	1	0.0033	0.0038	0.0037	0.0032	0.0037	0.0037
MAD	1	0.0436	0.0628	0.0363	0.0419	0.0603	0.0389
M-scale	1	0.0108	0.0139	0.0075	0.0104	0.0132	0.0076
SD	1.5	0.0033	0.0038	0.0036	0.0032	0.0037	0.0037
MAD	1.5	0.0436	0.0624	0.0360	0.0417	0.0590	0.0390
M-scale	1.5	0.0108	0.0138	0.0074	0.0103	0.0129	0.0075
SD	2	0.0033	0.0038	0.0036	0.0032	0.0036	0.0037
MAD	2	0.0436	0.0624	0.0360	0.0415	0.0581	0.0395
M-scale	2	0.0107	0.0136	0.0073	0.0102	0.0126	0.0075
SD	2.5	0.0033	0.0038	0.0036	0.0032	0.0036	0.0038
MAD	2.5	0.0436	0.0620	0.0355	0.0410	0.0577	0.0406
M-scale	2.5	0.0107	0.0135	0.0072	0.0100	0.0123	0.0074
SD	3	0.0033	0.0038	0.0036	0.0031	0.0036	0.0038
MAD	3	0.0435	0.0611	0.0345	0.0407	0.0566	0.0410
M-scale	3	0.0107	0.0135	0.0072	0.0099	0.0121	0.0073
SD	3.5	0.0033	0.0037	0.0036	0.0031	0.0035	0.0038
MAD	3.5	0.0435	0.0609	0.0344	0.0406	0.0558	0.0416
M-scale	3.5	0.0107	0.0134	0.0071	0.0097	0.0118	0.0074
SD	4	0.0033	0.0037	0.0036	0.0031	0.0035	0.0038
MAD	4	0.0435	0.0606	0.0340	0.0404	0.0552	0.0423
M-scale	4	0.0107	0.0133	0.0070	0.0096	0.0117	0.0074
SD	4.5	0.0033	0.0037	0.0036	0.0031	0.0035	0.0039
MAD	4.5	0.0434	0.0602	0.0336	0.0403	0.0548	0.0432
M-scale	4.5	0.0107	0.0133	0.0069	0.0095	0.0116	0.0076

Table 7: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under C_0 when ρ or $\nu = aN^{-3}$ and $f = \mathrm{id}$.

Scale	a	$\widehat{\phi}_{\text{PS},j}$			$\widehat{\phi}_{\text{PN},j}$		
SD	5	0.0033	0.0037	0.0036	0.0030	0.0034	0.0039
MAD	5	0.0434	0.0596	0.0330	0.0399	0.0543	0.0440
M-scale	5	0.0107	0.0132	0.0069	0.0095	0.0115	0.0076
SD	5.5	0.0033	0.0037	0.0035	0.0030	0.0034	0.0039
MAD	5.5	0.0432	0.0591	0.0326	0.0397	0.0536	0.0441
M-scale	5.5	0.0107	0.0131	0.0068	0.0094	0.0114	0.0076
SD	6	0.0032	0.0037	0.0035	0.0030	0.0034	0.0039
MAD	6	0.0432	0.0587	0.0321	0.0394	0.0532	0.0449
M-scale	6	0.0107	0.0130	0.0068	0.0093	0.0113	0.0076
SD	6.5	0.0032	0.0036	0.0035	0.0030	0.0034	0.0040
MAD	6.5	0.0432	0.0582	0.0317	0.0393	0.0530	0.0458
M-scale	6.5	0.0107	0.0130	0.0067	0.0092	0.0112	0.0077
SD	7	0.0032	0.0036	0.0035	0.0030	0.0034	0.0040
MAD	7	0.0431	0.0575	0.0310	0.0391	0.0525	0.0457
M-scale	7	0.0107	0.0129	0.0067	0.0092	0.0111	0.0076
SD	7.5	0.0032	0.0036	0.0035	0.0030	0.0034	0.0040
MAD	7.5	0.0428	0.0568	0.0306	0.0385	0.0517	0.0458
M-scale	7.5	0.0107	0.0128	0.0066	0.0091	0.0110	0.0076
SD	8	0.0032	0.0036	0.0035	0.0029	0.0034	0.0041
MAD	8	0.0428	0.0566	0.0304	0.0382	0.0514	0.0465
M-scale	8	0.0106	0.0128	0.0065	0.0090	0.0110	0.0076
SD	8.5	0.0032	0.0036	0.0035	0.0029	0.0034	0.0041
MAD	8.5	0.0427	0.0564	0.0303	0.0380	0.0510	0.0472
M-scale	8.5	0.0106	0.0127	0.0065	0.0090	0.0109	0.0075

Table 8: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under C_0 when ρ or $\nu = aN^{-3}$ and $f = \text{id}$.

Scale	a	$\widehat{\phi}_{\mathrm{PS},j}$			$\widehat{\phi}_{\mathrm{PN},j}$		
		$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
SD	0	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M-scale	0	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.1	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0.1	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M-scale	0.1	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.15	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0.15	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M-scale	0.15	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.25	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0.25	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M-scale	0.25	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.5	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0.5	0.0436	0.0634	0.0368	0.0436	0.0634	0.0368
M-scale	0.5	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	0.75	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	0.75	0.0436	0.0634	0.0368	0.0436	0.0634	0.0369
M-scale	0.75	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	1	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	1	0.0436	0.0634	0.0368	0.0436	0.0634	0.0369
M-scale	1	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	1.5	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	1.5	0.0436	0.0634	0.0368	0.0436	0.0634	0.0369
M-scale	1.5	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	2	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	2	0.0436	0.0634	0.0368	0.0436	0.0634	0.0369
M-scale	2	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077
SD	2.5	0.0033	0.0039	0.0037	0.0033	0.0039	0.0037
MAD	2.5	0.0436	0.0634	0.0368	0.0436	0.0634	0.0369
M-scale	2.5	0.0108	0.0141	0.0077	0.0108	0.0141	0.0077

Table 9: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under C_0 when ρ or $\nu = aN^{-4}$ and $f = \mathrm{id}$.

Scale estimator	a	$\widehat{\phi}_{\mathrm{PS},j}$			$\widehat{\phi}_{\mathrm{PN},j}$		
		$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
SD	0	0.0026	0.0033	0.0041	0.0026	0.0033	0.0041
MAD	0	0.0352	0.0528	0.0341	0.0352	0.0528	0.0341
M-scale	0	0.0078	0.0119	0.0088	0.0078	0.0119	0.0088
SD	0.1	0.0026	0.0032	0.0040	0.0026	0.0032	0.0041
MAD	0.1	0.0351	0.0523	0.0334	0.0351	0.0526	0.0341
M-scale	0.1	0.0078	0.0117	0.0086	0.0078	0.0117	0.0087
SD	0.5	0.0025	0.0031	0.0040	0.0025	0.0031	0.0041
MAD	0.5	0.0349	0.0515	0.0327	0.0349	0.0519	0.0353
M-scale	0.5	0.0077	0.0113	0.0082	0.0077	0.0114	0.0087
SD	1	0.0025	0.0031	0.0039	0.0025	0.0031	0.0041
MAD	1	0.0347	0.0496	0.0306	0.0347	0.0508	0.0358
M-scale	1	0.0077	0.0110	0.0077	0.0077	0.0111	0.0086
SD	1.5	0.0025	0.0030	0.0038	0.0025	0.0030	0.0041
MAD	1.5	0.0340	0.0480	0.0288	0.0340	0.0490	0.0358
M-scale	1.5	0.0076	0.0105	0.0073	0.0076	0.0106	0.0085
SD	2	0.0025	0.0029	0.0038	0.0025	0.0030	0.0041
MAD	2	0.0338	0.0465	0.0271	0.0338	0.0480	0.0364
M-scale	2	0.0075	0.0102	0.0069	0.0075	0.0104	0.0084
SD	2.5	0.0024	0.0028	0.0037	0.0024	0.0029	0.0041
MAD	2.5	0.0334	0.0458	0.0261	0.0334	0.0475	0.0372
M-scale	2.5	0.0074	0.0099	0.0066	0.0074	0.0101	0.0084
SD	3	0.0024	0.0028	0.0037	0.0024	0.0029	0.0041
MAD	3	0.0331	0.0444	0.0244	0.0331	0.0466	0.0373
M-scale	3	0.0073	0.0096	0.0063	0.0073	0.0099	0.0084
SD	3.5	0.0024	0.0027	0.0036	0.0024	0.0028	0.0042
MAD	3.5	0.0328	0.0435	0.0232	0.0328	0.0460	0.0378
M-scale	3.5	0.0072	0.0092	0.0060	0.0072	0.0096	0.0083
SD	4	0.0023	0.0027	0.0036	0.0023	0.0028	0.0042
MAD	4	0.0323	0.0422	0.0220	0.0324	0.0450	0.0380
M-scale	4	0.0071	0.0090	0.0058	0.0072	0.0095	0.0084
SD	4.5	0.0023	0.0026	0.0035	0.0023	0.0027	0.0042
MAD	4.5	0.0321	0.0413	0.0210	0.0321	0.0445	0.0387
M-scale	4.5	0.0070	0.0088	0.0056	0.0071	0.0093	0.0082
SD	5	0.0023	0.0026	0.0035	0.0023	0.0027	0.0042
MAD	5	0.0318	0.0406	0.0204	0.0319	0.0441	0.0398
M-scale	5	0.0070	0.0086	0.0055	0.0070	0.0091	0.0083
SD	5.5	0.0023	0.0025	0.0035	0.0023	0.0027	0.0042
MAD	5.5	0.0315	0.0396	0.0194	0.0315	0.0433	0.0398
M-scale	5.5	0.0069	0.0085	0.0053	0.0070	0.0090	0.0083

Table 10: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$, under C_0 when ρ or $\nu = an^{-3}$ and $f = \log$.

Model	Scale Estimator	$\widehat{\phi}_{\text{RAW},j}$ $(a=0)$			$\widehat{\phi}_{\text{PS},j}$ $(a=3)$			$\widehat{\phi}_{\text{PN},j}$ $(a=3)$		
		$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
C_0	SD	0.0033	0.0039	0.0037	0.0033	0.0038	0.0036	0.0031	0.0036	0.0038
	MAD	0.0436	0.0634	0.0368	0.0435	0.0611	0.0345	0.0407	0.0566	0.0410
	M-scale	0.0108	0.0141	0.0077	0.0107	0.0135	0.0072	0.0099	0.0121	0.0073
$C_{2,0.2}$	SD	1.4496	1.4496	0.0023	1.4470	1.4470	0.0022	1.3987	1.4060	0.0022
	MAD	0.3780	0.3903	0.0271	0.3779	0.3899	0.0267	0.3678	0.3853	0.0306
	M-scale	0.4347	0.4354	0.0058	0.4340	0.4347	0.0057	0.4225	0.4325	0.0061
$C_{3,a,0.1}$	SD	1.1438	1.9082	1.9172	1.1096	1.9059	1.9155	0.4586	1.8630	1.9160
	MAD	0.0929	0.2257	0.2339	0.0919	0.2145	0.2229	0.0747	0.1976	0.2627
	M-scale	0.0847	0.2123	0.2628	0.0832	0.2020	0.2528	0.0597	0.1778	0.2843
$C_{3,b,0.1}$	SD	0.0063	0.9221	0.9459	0.0061	0.7742	0.7973	0.0047	0.3208	0.4022
	MAD	0.0545	0.1691	0.1547	0.0540	0.1581	0.1442	0.0469	0.1329	0.1599
	M-scale	0.0200	0.1237	0.1294	0.0196	0.1106	0.1163	0.0154	0.0773	0.1132
$C_{3,a,0.2}$	SD	1.8275	1.9238	1.9405	1.8247	1.9239	1.9402	1.7071	1.9241	1.9437
	MAD	0.2598	0.7858	0.8071	0.2572	0.7619	0.7855	0.2081	0.7528	0.8699
	M-scale	0.2997	1.0419	1.0977	0.2960	1.0138	1.0722	0.2391	1.0190	1.1590
$C_{3,b,0.2}$	SD	0.0140	1.7596	1.7905	0.0136	1.7409	1.7725	0.0089	1.5634	1.6750
	MAD	0.0874	0.5092	0.5013	0.0854	0.4850	0.4774	0.0651	0.4010	0.4910
	M-scale	0.0427	0.4782	0.4966	0.0413	0.4513	0.4704	0.0253	0.3441	0.4699
$C_{23,0.1}$	SD	1.1762	1.3131	1.5685	1.1477	1.3071	1.5515	0.5761	1.1000	1.1980
	MAD	0.1808	0.3205	0.2415	0.1785	0.3131	0.2349	0.1534	0.2877	0.2706
	M-scale	0.1850	0.3601	0.2862	0.1834	0.3525	0.2795	0.1524	0.3156	0.3141
$C_{23,0.2}$	SD	1.8303	0.1930	1.8255	1.8295	0.1964	1.8233	1.7296	0.3014	1.7622
	MAD	0.9350	1.0709	0.6005	0.9305	1.0663	0.5936	0.8700	1.0133	0.6702
	M-scale	1.0599	1.2001	0.7009	1.0564	1.1928	0.6944	0.9940	1.0974	0.7602

Table 11: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$ when ρ or $\nu = an^{-3}$ for $a = 0, 3$ and $f = \text{id}$.

Model	Scale Estimator	$\widehat{\phi}_{\text{RAW},j}$			$\widehat{\phi}_{\text{PS},j}$			$\widehat{\phi}_{\text{PN},j}$		
		$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$	$j=1$	$j=2$	$j=3$
C_0	SD	0.0026	0.0033	0.0041	0.0024	0.0028	0.0037	0.0024	0.0029	0.0041
	MAD	0.0352	0.0528	0.0341	0.0331	0.0444	0.0244	0.0331	0.0466	0.0373
	M-scale	0.0078	0.0119	0.0088	0.0073	0.0096	0.0063	0.0073	0.0099	0.0084
$C_{2,0.2}$	SD	1.5317	1.5316	0.0023	1.4808	1.4808	0.0024	1.4828	1.4898	0.0024
	MAD	0.3724	0.3820	0.0250	0.3596	0.3656	0.0189	0.3596	0.3768	0.0295
	M-scale	0.4350	0.4359	0.0062	0.4193	0.4196	0.0050	0.4193	0.4295	0.0065
$C_{3,a,0.1}$	SD	1.2444	1.9202	1.9291	0.3998	1.8529	1.8860	0.4865	1.8703	1.9251
	MAD	0.0858	0.2198	0.2400	0.0631	0.1417	0.1568	0.0636	0.1776	0.2533
	M-scale	0.0838	0.2088	0.2630	0.0523	0.1285	0.1709	0.0531	0.1590	0.2713
$C_{3,a,0.2}$	SD	1.8608	1.9362	1.9453	1.7376	1.9369	1.9229	1.7766	1.9372	1.9495
	MAD	0.2649	0.8389	0.8763	0.2111	0.5327	0.5904	0.2138	0.8055	0.9377
	M-scale	0.3058	1.1085	1.1667	0.2429	0.6651	0.7634	0.2480	1.0856	1.2310
$C_{3,b,0.1}$	SD	0.0043	0.8526	0.8813	0.0036	0.1738	0.1894	0.0036	0.2166	0.2883
	MAD	0.0425	0.1551	0.1491	0.0364	0.0993	0.0906	0.0364	0.1115	0.1421
	M-scale	0.0135	0.1177	0.1267	0.0104	0.0662	0.0720	0.0104	0.0694	0.1070
$C_{3,b,0.2}$	SD	0.0096	1.7664	1.8026	0.0067	1.4635	1.5061	0.0067	1.5608	1.6813
	MAD	0.0608	0.4761	0.4801	0.0444	0.3102	0.3156	0.0451	0.3543	0.4597
	M-scale	0.0274	0.4591	0.4827	0.0159	0.3099	0.3316	0.0159	0.3351	0.4722
$C_{23,0.1}$	SD	1.2966	1.4357	1.6548	0.5249	1.1194	1.1114	0.6028	1.1933	1.3057
	MAD	0.1664	0.3169	0.2501	0.1394	0.2471	0.1837	0.1398	0.2857	0.2830
	M-scale	0.1688	0.3476	0.2947	0.1336	0.2605	0.2130	0.1342	0.2972	0.3210
$C_{23,0.2}$	SD	1.8585	0.1266	1.8542	1.7549	0.2603	1.7420	1.7857	0.1954	1.8153
	MAD	0.9219	1.1274	0.5867	0.8667	1.0157	0.4736	0.8699	1.0779	0.6581
	M-scale	1.0565	1.2364	0.6868	1.0028	1.1224	0.5483	1.0052	1.1709	0.7590

Table 12: Mean values of $\|\widehat{\phi}_j/\|\widehat{\phi}_j\| - \phi_j\|^2$ when ρ or $\nu = an^{-3}$ for $a = 0, 3$ and $f = \log$.