Journal of Multivariate Analysis 133 (2015) 173–199

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Influence function of projection-pursuit principal components for functional data

Juan Lucas Bali, Graciela Boente *

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

ARTICLE INFO

Article history:
Received 12 April 2013
Available online 28 September 2014

AMS subject classifications:
62G35
62H25

Keywords:
Elliptical distribution
Fisher-consistency
Functional principal component
Influence function
Robust estimation
Smoothing

ABSTRACT

In the finite-dimensional setting, Li and Chen (1985) proposed a method for principal components analysis using projection-pursuit techniques. This procedure was generalized to the functional setting by Bali et al. (2011), where also different penalized estimators were defined to provide smooth functional robust principal component estimators. This paper completes their study by deriving the influence function of the functional related to the principal direction estimators and their size. As is well known, the influence function is a measure of robustness which can also be used for diagnostic purposes. In this sense, the obtained results can be helpful for detecting influential observations for the principal directions.

1. Introduction

Principal Components Analysis is a standard technique used in the context of multivariate analysis as a dimension-reduction technique. Traditionally, the goal is to determine an orthonormal basis such that each direction maximizes the variability of the random elements.

Two lines of work were developed in parallel from this method. First, the possibility of replacing the variance with a robust dispersion scale is considered, in order to gain resistance when atypical data are present in the sample. This alternative was pursued by Li and Chen [23] who following principles of projection-pursuit defined the first principal direction estimator as that maximizing a robust scale of the projected data. The subsequent directions are obtained by imposing orthogonality conditions. Croux and Ruiz-Gazen [11] obtained the influence function of these projection-pursuit estimators.

Another line of work was to extend the euclidean setting to a functional one. Among them, Dauxois et al. [12] studied the asymptotic properties of the eigenfunctions of the sample covariance operator while Rice and Silverman [29] proposed smooth estimators by penalizing the sample variance using an additive roughness term. Besides, Silverman [33], considered an approach based on penalizing the norm instead of the sample variance, see also Ramsay and Silverman [28] for a review.

There are few works that combine both aspects: robustness and functional setting. One of the first ones is Locantore et al. [24] where spherical principal components are introduced. Their influence function was derived by Gervini [19]. Recently, Bali et al. [3] considered robust estimators of the functional principal directions using a projection-pursuit approach that includes a penalization in the scale or in the norm. Consistency and qualitative robustness are derived therein.

* Correspondence to: Departamento de Matemáticas, FCEyN, UBA, Ciudad Universitaria, Pabellón 2, Buenos Aires, C1428EHA, Argentina.
E-mail addresses: lbali@dm.uba.ar (J.L. Bali), gboente@dm.uba.ar, gboente@fibertel.com.ar (G. Boente).

http://dx.doi.org/10.1016/j.jmva.2014.09.004
0047-259X/

In this paper, we complement the paper by Bali et al. [3] providing an expression for the influence function of the functional related to the "raw" estimators, that is, the unsmoothed estimators obtained without a penalization term included either on the scale or in the norm. We also define a smooth functional to analyse the infinitesimal effect of outliers on the estimators obtained penalizing the scale. Smoothed functionals were also considered in other settings such as nonparametric regression and tolerance intervals to define influence functions of smooth estimators. The paper is organized as follows, Section 2 states some preliminary concepts and results that will be helpful along the paper. To measure robustness with respect to single outliers, the influence function of the raw estimator is studied in Section 3.1, while Section 3.2 considers the situation of a smoothed functional. As a particular case, in Section 4 we derive an expression for the influence function of the projection-pursuit estimators obtained using an M-scale estimator which includes as particular situation the classical estimators. Besides being of theoretical interest, measuring the influence of an observation on the classical estimates can be used as a diagnostic tool to detect influential observations. For that reason, a discussion on diagnostic tools based on the influence function to detect atypical observations is given in Section 5. Proofs are relegated to the Appendix.

2. Preliminaries

2.1. Elliptical families

Elliptical families play an important role when deriving consistency and infinitesimal robustness of the principal component robust estimators, in the finite-dimensional setting. This notion has been extended to the functional setting by Bali and Boente [2]. We recall here their definition for the sake of completeness.

Let $\mathbf{Z} \in \mathbb{R}^d$ be a random vector. We say that $\mathbf{Z}$ has an elliptical distribution, and we denote it as $\mathbf{Z} \sim \mathcal{E}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$, if there exists a vector $\boldsymbol{\mu} \in \mathbb{R}^d$, a positive semidefinite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ and a function $\psi : \mathbb{R}_+ \to \mathbb{R}$, called the characteristic generator, such that the characteristic function of $\mathbf{Z} - \boldsymbol{\mu}$ is given by $\varphi_{\mathbf{Z}-\boldsymbol{\mu}}(\mathbf{t}) = \psi(\mathbf{t}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{t})$, for all $\mathbf{t} \in \mathbb{R}^d$. Let now X be a random element in a separable Hilbert space $\mathcal{H}$ and $\mu \in \mathcal{H}$. Let $\boldsymbol{\Gamma} : \mathcal{H} \to \mathcal{H}$ be a self-adjoint, positive semidefinite and compact operator. The random element X has an elliptical distribution with parameters $(\mu, \boldsymbol{\Gamma})$, denoted as $X \sim \mathcal{E}(\mu, \boldsymbol{\Gamma}, \psi)$, if for any linear and bounded operator $A : \mathcal{H} \to \mathbb{R}^d$, AX has a multivariate elliptical distribution with parameters $A\mu$ and $A\boldsymbol{\Gamma}A^*$, i.e., $AX \sim \mathcal{E}_d(A\mu, A\boldsymbol{\Gamma}A^*, \psi)$, where $A^* : \mathbb{R}^p \to \mathcal{H}$ stands for the adjoint operator of A. For the sake of simplicity, we will omit the symbol ψ and denote $X \sim \mathcal{E}_d(\mu, \boldsymbol{\Gamma})$, when there is no confusion. As in the finite-dimensional setting, if the covariance operator, $\boldsymbol{\Gamma}_X$, of X exists then, $\boldsymbol{\Gamma}_X = a\ \boldsymbol{\Gamma}$, for some $a \in \mathbb{R}$. The elliptical distributions in $\mathcal{H}$ include, among others, the Gaussian distributions and scale mixtures of Gaussian. Recently, Boente et al. [7] provide a characterization of elliptical distributions on separable Hilbert spaces. To be more precise, they show that, when $\boldsymbol{\Gamma}$ has infinite range, the class of elliptical distributions is equivalent to the class of scale mixtures of Gaussian distributions on $\mathcal{H}$.

From now on, denote by $\langle \cdot, \cdot \rangle$ the inner product in $\mathcal{H}$ and by $\|\alpha\|^2 = \langle \alpha, \alpha \rangle$. Moreover, let $P[\alpha]$ be the distribution of $\langle \alpha, X \rangle$ when $X \sim P$.

The following lemma, whose proof is given in the Appendix, states that the projected distributions $P[\alpha]$ belong to the same location–scale family. This property is analogous to Lemma 1 in [11].

Lemma 2.1. *Let $X \in \mathcal{H}$ be such that $X \sim \mathcal{E}(\mu, \boldsymbol{\Gamma}, \psi)$. Then, there exists an univariate symmetric distribution F_0 such that*

$$P[\alpha]((-\infty, y]) = P\left(\langle \alpha, X \rangle \le y\right) = F_0\left(\frac{y - \langle \mu, \alpha \rangle}{\sqrt{\langle \alpha, \boldsymbol{\Gamma}\alpha \rangle}}\right),$$

for any α such that $\alpha \notin \ker(\boldsymbol{\Gamma})$, i.e., such that $\langle \alpha, \boldsymbol{\Gamma}\alpha \rangle \neq 0$. That is, the random variable $Z_\alpha = \langle \alpha, X - \mu \rangle / \sqrt{\langle \alpha, \boldsymbol{\Gamma}\alpha \rangle}$ has distribution F_0. On the other hand, if $\langle \alpha, \boldsymbol{\Gamma}\alpha \rangle = 0$, $\mathbb{P}(\langle \alpha, X - \mu \rangle = 0) = 1$.

2.2. The robust functional

Principal components analysis for general Hilbert spaces can be described as follows.

Let $X \in \mathcal{H}$ be a random element of a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle$ defined in $(\Omega, \mathcal{A}, P)$. 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 $(u \otimes v)w = \langle v, w \rangle u$. When X has finite second moment, i.e., $\mathbb{E}(\|X\|^2) < \infty$, the covariance operator of X, $\boldsymbol{\Gamma}_X$, can be written as $\boldsymbol{\Gamma}_X = \mathbb{E}\{(X - \mu) \otimes (X - \mu)\}$ with $\mu = \mathbb{E}(X)$. In this situation, the operator $\boldsymbol{\Gamma}_X$ which is linear, self-adjoint and continuous is in the trace class, so that, in particular, $\boldsymbol{\Gamma}_X$ is a Hilbert–Schmidt operator.

In general, for $Y = \langle \alpha, X \rangle$, we have $\mathrm{var}(Y) = \langle \alpha, \boldsymbol{\Gamma}_X\alpha \rangle$. An important optimality property of the first principal component variable is that it can be defined as the variable $Z_1 = \langle \alpha_1, X \rangle$ such that

$$\mathrm{var}(Z_1) = \sup_{\{\alpha:\|\alpha\|=1\}} \mathrm{var}\left(\langle \alpha, X \rangle\right) = \sup_{\{\alpha:\|\alpha\|=1\}} \langle \alpha, \boldsymbol{\Gamma}_X\alpha \rangle. \tag{1}$$

Any solution to (1), i.e. any α for which the supremum is obtained, corresponds to an eigenfunction associated with the largest eigenvalue of the covariance operator $\boldsymbol{\Gamma}_X$, i.e., $\alpha_1 = \phi_1$ and $\mathrm{var}(Z_1) = \lambda_1$. If $\lambda_1 > \lambda_2$, then α_1 is unique up to a sign change. As in the multivariate setting, the other principal components can be obtained successively via (1), but under the orthogonality condition that $\langle \alpha_j, \alpha_k \rangle = 0$ for $j < k$.

The idea beyond the approach in [3] is to view principal components as in (1), but replacing the variance by a robust scale functional. Denote by $\mathcal{G}$ the set of all univariate distributions. As it is well known, a scale functional $\sigma_{\mathrm{R}} : \mathcal{G} \to [0, +\infty)$ is one 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. Two well known examples of scale functionals are the standard deviation, $\mathrm{SD}(G) = \left\{\mathbb{E}(Y - \mathbb{E}(Y))^2\right\}^{1/2}$, where $Y \sim G$, and the median absolute deviation about the median, $\mathrm{MAD}(G) = c\,\mathrm{median}\,(|Y - \mathrm{median}(Y)|)$. The normalization constant c, used in the MAD, can be chosen so that its empirical or sample version is consistent for a scale parameter of interest. Typically, one chooses $c = 1/\Phi^{-1}(0.75)$ so that the MAD equals the standard deviation at a normal distribution. A broader class which includes the previous ones is the class of the M-scale functionals. An M-scale functional with a bounded and continuous score function can have both a high breakdown point and a continuous and bounded influence function. Given a location parameter μ, an M-scale functionals $\sigma_M(G)$ with a continuous score function $\chi : \mathbb{R} \to \mathbb{R}$ can be defined to be a solution to the equation

$$\mathbb{E}\left[\chi\left(\frac{Y-\mu}{\sigma_{\mathrm{R}}(G)}\right)\right] = \delta. \tag{2}$$

Typically, the score function χ is even with $\chi(0) = 0$, non-decreasing on $\mathbb{R}_+$ and with $0 < \sup_{x\in\mathbb{R}} \chi(x) = \chi(+\infty) = \lim_{x\to+\infty} \chi(x)$. When $\chi(+\infty) = 2\delta$, the M-estimate of scale has a 50% breakdown point, and by choosing χ properly one can also obtain a highly efficient estimate, see [9]. A popular choice, which corresponds to that considered in Section 4, is the score function introduced by Beaton and Tukey [4], namely $\chi_c(y) = \min(3\,(y/c)^2 - 3\,(y/c)^4 + (y/c)^6\,, 1)$, with c being a tuning constant chosen so that the corresponding M-estimator of scale is consistent for a scale parameter of interest. For example, the choice $c = 1.56$ when $\delta = 1/2$ ensures that the M-scale functional is Fisher-consistent at the normal distribution and has a 50% breakdown point.

For a given σ_{R} scale functional, denote as $\sigma(\alpha) = \sigma_{\mathrm{R}}(P[\alpha])$. Bali et al. [3] defined the raw (meaning unsmoothed) robust functional principal component directions as

$$\begin{cases} \phi_{\mathrm{R},1}(P) = \underset{\|\alpha\|=1}{\operatorname{argmax}}\, \sigma(\alpha) \\ \phi_{\mathrm{R},m}(P) = \underset{\|\alpha\|=1,\alpha\in\mathcal{B}_m}{\operatorname{argmax}}\ \sigma(\alpha), \quad 2 \le m, \end{cases} \tag{3}$$

where $\mathcal{B}_m = \{\alpha \in \mathcal{H} : \langle\alpha, \phi_{\mathrm{R},j}(P)\rangle = 0,\ 1 \le j \le m-1\}$. The m-th largest principal value functional is given by

$$\lambda_{\mathrm{R},m}(P) = \sigma^2(\phi_{\mathrm{R},m}) = \max_{\|\alpha\|=1,\alpha\in\mathcal{B}_m} \sigma^2(\alpha)\ . \tag{4}$$

As mentioned in [3], the maximum above is attained if the scale functional σ_{R} is (weakly) continuous. Besides, for elliptical distributions the functionals $\phi_{\mathrm{R},m}(P)$ and $\lambda_{\mathrm{R},m}(P)$ have a simple interpretation. Effectively, when considering a robust scale functional, we have that $\sigma^2(\alpha) = c\langle\alpha, \mathbf{\Gamma}\alpha\rangle$ if X has an elliptical distribution $\mathcal{E}(\mu, \mathbf{\Gamma})$. Hence, Lemma 5.1 in [3] entails that the functionals $\phi_{\mathrm{R},m}(P)$ defined through (3) correspond to the eigenfunctions of $\mathbf{\Gamma}_X$.

2.3. Differentiability

Several notions of differentiability have been defined in normed spaces, the weakest being the Gateaux differentiability, which has been extendedly used in statistics due to its relation to the influence function and the stronger being Fréchet differentiability. In many situations, Fréchet differentiability is too strong for the functionals considered in statistics. In particular, to derive asymptotic normality of estimators obtained from a given functional, Gateaux differentiability is not enough while Fréchet differentiability is many times not satisfied, an intermediate notion is the Hadamard differentiability which turns out to be helpful to apply the delta method. For the sake of completeness we recall their definitions.

Definition 1. Let $\mathbb{D}_1$ and $\mathbb{D}_2$ be two normed spaces and consider a map $\Upsilon : \mathbb{D}_\Upsilon \subset \mathbb{D}_1 \to \mathbb{D}_2$.

(a) The map Υ is said to be *Fréchet differentiable* at $\theta \in \mathbb{D}_\Upsilon$ if there exists a linear and bounded operator $\Upsilon'_\theta : \mathbb{D}_1 \to \mathbb{D}_2$ such that
$$\lim_{h\to 0} \frac{\|\Upsilon(\theta+h) - \Upsilon(\theta) - \Upsilon'_\theta(h)\|}{\|h\|} = 0.$$

(b) The map Υ is said to be *Hadamard differentiable* at $\theta \in \mathbb{D}_\Upsilon$ if there exists a linear and continuous operator $\Upsilon'_\theta : \mathbb{D}_1 \to \mathbb{D}_2$ such that for every compact $K \subset \mathbb{D}_1$
$$\lim_{t\to 0} \sup_{h\in K,\theta+th\in\mathbb{D}_\Upsilon} \left\| \frac{\Upsilon(\theta+th) - \Upsilon(\theta)}{t} - \Upsilon'_\theta(h) \right\| = 0.$$

(c) The map Υ is said to be *Gateaux differentiable* at $\theta \in \mathbb{D}_\Upsilon$, in the direction h, if there exists a quantity $\Upsilon'_\theta(h)$ such that, for any sequence $t_n \to 0$ as $n \to \infty$,
$$\lim_{n\to\infty} \frac{\Upsilon(\theta+t_n h) - \Upsilon(\theta)}{t_n} = \Upsilon'_\theta(h).$$

Higher order derivatives can also be defined similarly. Even if we are using the same notation in the three definitions, it will be clear in each situation, which differentiability is considered.

Remark 2.1. (a) Hadamard differentiability is the weakest kind of differentiability that satisfies the chain rule. When $\mathbb{D}_1 = \mathbb{R}$ the three notions Fréchet, Hadamard and Gateaux differentiability are equivalent. On the other hand, if $\mathbb{D}_1$ is a Banach space, continuous Gateaux differentiability is equivalent to Hadamard differentiability, where continuity is understood as θ varies in $\mathbb{D}_1$. For relations among these notions see, for instance, [18].

(b) Let $B(\mathbb{D}_1, \mathbb{D}_2)$ be the set of linear and bounded (and hence continuous) operators from $\mathbb{D}_1$ to $\mathbb{D}_2$. If $\mathbb{D}_\Upsilon$ is an open set and Υ is *Hadamard differentiable* at each $\theta \in \mathbb{D}_\Upsilon$, we can consider the map $\Upsilon' : \mathbb{D}_\Upsilon \to B(\mathbb{D}_1, \mathbb{D}_2)$ defined as $\Upsilon'(\theta) = \Upsilon'_\theta$ that will be called the *Hadamard derivative* of Υ. When Υ' is continuous, we will say that Υ is $\mathcal{C}^1$-Hadamard.

(c) When $\mathbb{D}_1$ is a separable Hilbert space, as the one we are considering, if $\Upsilon : \mathbb{D}_\Upsilon \to \mathbb{R}$ is *Fréchet differentiable* at $\theta \in \mathbb{D}_\Upsilon$, with $\mathbb{D}_\Upsilon$ an open subset of $\mathbb{D}_1$, the Riesz Theorem allows to represent the linear and bounded operator $\Upsilon'_\theta : \mathbb{D}_1 \to \mathbb{R}$ through an element of the Hilbert space $\mathbb{D}_1$, that will be named the gradient of Υ and denoted $\nabla\Upsilon_\theta$, that is, we have that $\Upsilon'_\theta\alpha = \langle\nabla\Upsilon_\theta, \alpha\rangle$.

Denote as $\mathbb{D}_1^\star = B(\mathbb{D}_1, \mathbb{R})$ the dual space. If $\Upsilon : \mathbb{D}_\Upsilon \to \mathbb{R}$ is twice *Fréchet differentiable* at $\theta \in \mathbb{D}_\Upsilon$, $\mathbb{D}_\Upsilon \subset \mathbb{D}_1$, using that $B(\mathbb{D}_1, \mathbb{D}_1^\star)$ can be identified with the space of bilinear and continuous operator from $\mathbb{D}_1 \times \mathbb{D}_1$ to $\mathbb{R}$, denoted $BL(\mathbb{D}_1 \times \mathbb{D}_1, \mathbb{R})$, we get that Υ''_θ defines a continuous bilinear operator from $\mathbb{D}_1 \times \mathbb{D}_1$ to $\mathbb{R}$, through the identification $\Upsilon''_\theta(\alpha, \beta) = (\Upsilon''_\theta(\alpha))\,\beta$ for $(\alpha, \beta) \in \mathbb{D}_1 \times \mathbb{D}_1$. In particular, if $\mathbb{D}_1$ is a Hilbert space, Υ''_θ can be represented through a bounded operator, $H\Upsilon_\theta : \mathbb{D}_1 \to \mathbb{D}_1$ as $\Upsilon''_\theta(\alpha, \beta) = \langle H\Upsilon_\theta\alpha, \beta\rangle$. Besides, if we define $\Psi(\theta) = \nabla\Upsilon_\theta$, then $H\Upsilon_\theta = \Psi'_\theta$.

Similar arguments can be used for Hadamard differentiable maps.

Remark 2.2. (a) Let $\mathcal{H}$ be a separable Hilbert space. As a simple example, that will be used in the sequel, let us consider the maps $\Upsilon : \mathcal{H} \to \mathbb{R}$ defined as $\Upsilon(\alpha) = \langle\alpha, \mathbf{\Gamma}\alpha\rangle$ and $\Psi : \mathcal{H} \to \mathbb{R}$, defined as $\Psi(\alpha) = \langle\beta_0, \alpha\rangle$ with $\mathbf{\Gamma}$ a self-adjoint, semi-definite and compact operator and $\beta_0 \in \mathcal{H}$ fixed. Note that Ψ is a linear and bounded operator. The Gateaux derivatives of Υ and Ψ at a given α_0 can easily be computed and they result to be the linear and bounded operators $\Upsilon'_{\alpha_0} : \mathcal{H} \to \mathbb{R}$ and $\Psi'_{\alpha_0} : \mathcal{H} \to \mathbb{R}$ given by $\Upsilon'_{\alpha_0}\alpha = 2\langle\mathbf{\Gamma}\alpha_0, \alpha\rangle$ and $\Psi'_{\alpha_0}\alpha = \langle\beta_0, \alpha\rangle$. In fact easy calculations allow to show that Υ and Ψ are Fréchet differentiable with Fréchet derivatives Υ'_{α_0} and Ψ'_{α_0}, respectively, and gradients $\nabla\Upsilon_{\alpha_0} = 2\mathbf{\Gamma}\alpha_0$ and $\nabla\Upsilon_{\alpha_0} = \beta_0$. Moreover, $H\Upsilon_{\alpha_0} = 2\mathbf{\Gamma}$.

(b) We will also use the following property.

Let us consider maps $\Upsilon : \mathcal{U} \subset \mathcal{H} \to \mathbb{R}$ and $\Psi : \mathcal{U} \subset \mathcal{H} \to \mathbb{R}$, where $\mathcal{H}$ is a separable Hilbert space and $\mathcal{U}$ is an open set. Define $\widetilde{\Upsilon}(\theta) = \Upsilon(\theta)\Psi(\theta)$. Then, if Υ and Ψ are Fréchet (Hadamard or Gateaux) differentiable at $\theta_0 \in \mathcal{U}$, we have that $\widetilde{\Upsilon}$ is Fréchet (Hadamard or Gateaux) differentiable at θ_0 and $\nabla\widetilde{\Upsilon}_{\theta_0} = \Psi(\theta_0)\nabla\Upsilon_{\theta_0} + \Upsilon(\theta_0)\nabla\Psi_{\theta_0}$.

In product spaces, partial differentiability can also be defined.

Definition 2. Let $\mathbb{D}_1$, $\mathbb{D}_2$ and $\mathbb{D}$ be Banach spaces, $\mathcal{U}_i \subset \mathbb{D}_i$, $i = 1, 2$ open subsets, $\Upsilon : \mathcal{U}_1 \times \mathcal{U}_2 \to \mathbb{D}$ and $\theta_0 = (\theta_{01}, \theta_{02}) \in \mathcal{U}_1 \times \mathcal{U}_2$ fixed. We say that Υ is partially Fréchet differentiable with respect to the first coordinate at θ_0 if $\Upsilon_{\theta_{02}} : \mathcal{U}_1 \to \mathbb{D}$ defined as $\Upsilon_{\theta_{02}}(\theta_1) = \Upsilon(\theta_1, \theta_{02})$, is Fréchet differentiable at θ_{01}. We will denote its *partial Fréchet derivative* at θ_0, $D_{1,\theta_0}\Upsilon = (\partial\Upsilon/\partial\mathbb{D}_1)(\theta_0)$. Hence, $D_{1,\theta_0}\Upsilon$ is a bounded and linear operator $D_{1,\theta_0}\Upsilon : \mathbb{D}_1 \to \mathbb{D}$ that satisfies

$$\lim_{h_1\to 0} \frac{\|\Upsilon(\theta_0 + (h_1, 0)) - f(\theta_0) - (D_{1,\theta_0}\Upsilon)h_1\|}{\|h_1\|} = 0.$$

Analogously we define $D_{2,\theta_0}\Upsilon$. In particular, the map $D_1\Upsilon : \mathcal{U}_1 \times \mathcal{U}_2 \to B(\mathbb{D}_1, \mathbb{D})$ defined as $(D_1\Upsilon)\theta = D_{1,\theta}\Upsilon$ is defined when Υ is partially Fréchet differentiable at any $\theta \in \mathcal{U}_1$.

Remark 2.3. (a) It is easy to show that if Υ is Fréchet differentiable at θ_0, then it is partially differentiable with respect to each coordinate at θ_0 and for any $h = (h_1, h_2) \in \mathbb{D}_1 \times \mathbb{D}_2$, we have $\Upsilon'_{\theta_0}h = (D_{1,\theta_0}\Upsilon)h_1 + (D_{2,\theta_0}\Upsilon)h_2$. Moreover, if we define $J_i : \mathbb{D}_1 \to \mathbb{D}$ as $J_1(h_1) = (h_1, 0)$ and $J_2(h_2) = (0, h_2)$, then $D_{i,\theta_0}\Upsilon = \Upsilon'_{\theta_0} \circ J_i$.

(b) Partial Hadamard differentials and Gateaux differentials can be defined in the same way, as well as higher order partial derivatives.

For instance, assume that $D_{1,\theta}\Upsilon$ exists for θ in a neighbourhood of θ_0 and define $\Psi : \mathcal{U}_1 \times \mathcal{U}_2 \to B(\mathbb{D}_1, \mathbb{D})$ as $\Psi(\theta) = D_{1,\theta}\Upsilon$. Then, we say that Υ is twice partially Fréchet differentiable with respect to the first coordinate at θ_0 if there exists a linear and bounded operator $D^2_{11,\theta_0}\Upsilon : \mathbb{D}_1 \to B(\mathbb{D}_1, \mathbb{D})$ such that

$$\lim_{h_1\to 0} \frac{\|\Psi(\theta_0 + (h_1, 0)) - \Psi(\theta_0) - (D^2_{11,\theta_0}\Upsilon)h_1\|}{\|h_1\|} = 0,$$

where the norm in the numerator is the operator norm in $B(\mathbb{D}_1, \mathbb{D}_2)$.

Hence, $D^2_{11,\theta_0}\Upsilon \in B(\mathbb{D}_1, B(\mathbb{D}_1, \mathbb{D}))$. Noting that $B(\mathbb{D}_1, B(\mathbb{D}_1, \mathbb{D}))$ can be identified with the space of bilinear and continuous operator from $\mathbb{D}_1 \times \mathbb{D}_1$ to $\mathbb{D}$, denoted $BL(\mathbb{D}_1 \times \mathbb{D}_1, \mathbb{D})$, we get that $D^2_{11,\theta_0}\Upsilon$ defines a continuous bilinear operator from $\mathbb{D}_1 \times \mathbb{D}_1$ to $\mathbb{D}$, through the identification $D^2_{11,\theta_0}\Upsilon(h_1, \upsilon_1) = ((D^2_{11,\theta_0}\Upsilon)h_1)\,\upsilon_1$ for $(h_1, \upsilon_1) \in \mathbb{D}_1 \times \mathbb{D}_1$. In particular, if $\mathbb{D}_i$ are Hilbert spaces and $\mathbb{D} = \mathbb{R}$, $D^2_{11,\theta_0}\Upsilon$ can be represented through a bounded operator, $\mathbf{G}_{11} : \mathbb{D}_1 \to \mathbb{D}_1$ as $D^2_{11,\theta_0}\Upsilon(\alpha, \beta) = \langle \mathbf{G}_{11}\alpha, \beta\rangle$.

We can also define $D^2_{11}\Upsilon : \mathcal{U}_1 \times \mathcal{U}_2 \to BL(\mathbb{D}_1 \times \mathbb{D}_1, \mathbb{D})$ as $(D^2_{11}\Upsilon)\theta_0 = D^2_{11,\theta_0}\Upsilon$. Similar arguments can be used for the other partial second derivatives $D^2_{12,\theta_0}\Upsilon$, $D^2_{21,\theta_0}\Upsilon$ and $D^2_{22,\theta_0}\Upsilon$ where to avoid confusion in the notation, $D^2_{21,\theta_0}\Upsilon$ will stand for the derivative in the second coordinate of $\Psi(\theta) = D_{1,\theta}\Upsilon$, while $D^2_{12,\theta_0}\Upsilon$ will stand for the derivative in the first coordinate of $D_{2,\theta}\Upsilon$.

(c) As in real analysis, if Υ is twice continuously Fréchet differentiable at θ_0, then Υ''_{θ_0} is symmetric, which entails that $D^2_{12,\theta_0}\Upsilon(\theta_1, \theta_2) = D^2_{21,\theta_0}\Upsilon(\theta_2, \theta_1)$.
Similar results hold for Hadamard differentiable maps.

(d) In particular, we have the following property.
Given $\Upsilon : \mathbb{D}_1 \times \mathbb{D}_2 \to \mathbb{F}$, $\eta : \mathbb{D} \to \mathbb{D}_1$, $\upsilon : \mathbb{D} \to \mathbb{D}_2$ define $\Lambda : \mathbb{D} \to \mathbb{F}$ as $\Lambda(\theta) = \Upsilon(\eta(\theta), \upsilon(\theta))$ and $\boldsymbol{\zeta} : \mathbb{D} \to \mathbb{D}_1 \times \mathbb{D}_2$ as $\boldsymbol{\zeta}(\theta) = (\eta(\theta), \upsilon(\theta))$. If Υ and $\boldsymbol{\zeta}$ are Fréchet (Hadamard) differentiable at $\alpha_0 = \zeta(\theta_0) = (\eta(\theta_0), \upsilon(\theta_0))$ and θ_0, respectively, then we have that Λ is Fréchet (Hadamard) differentiable at θ_0 and for any $h \in \mathbb{D}$

$$\Lambda'_{\theta_0}h = \Upsilon'_{\alpha_0} \circ \boldsymbol{\zeta}'_{\alpha_0}h = (D_{1,\alpha_0}\Upsilon)\eta'_{\theta_0}h + (D_{2,\alpha_0}\Upsilon)\upsilon'_{\theta_0}h.$$

For the sake of completeness we recall, in Theorems 2.1 and 2.2, the Implicit Function Theorem and the *Lagrange multipliers method* that will be used in the sequel. Their proof can be found, for instance, in [18].

Theorem 2.1. *Let $\mathbb{D}_1$ and $\mathbb{D}_2$ be Banach spaces and $\Upsilon : \mathbb{D}_1 \times \mathbb{D}_2 \to \mathbb{D}$ be a $\mathcal{C}^1$-Fréchet differentiable map. Let $\theta_0 = (x_0, y_0) \in \mathbb{D}_1 \times \mathbb{D}_2$ be such that $\Upsilon(\theta_0) = 0$ and $D_{2,\theta_0}\Upsilon : \mathbb{D}_2 \to \mathbb{D}$ is an invertible operator. Then, there exist open neighbourhoods $\mathcal{U}_{x_0} \subset \mathbb{D}_1$ of x_0 and $\mathcal{V}_{y_0} \subset \mathbb{D}_2$ of y_0 and a $\mathcal{C}^1$-Fréchet differentiable map $\upsilon : \mathcal{U}_{x_0} \to \mathcal{V}_{y_0}$ such that $y_0 = \upsilon(x_0)$ and for any $x \in \mathcal{U}_{x_0}$ and $y \in \mathcal{V}_{y_0}$, we have that $\Upsilon(x, y) = 0$ if and only if $y = \upsilon(x)$. Moreover, for any $x \in \mathcal{U}_{x_0}$, if $\theta = (x, \upsilon(x))$ and if $D_{2,\theta}\Upsilon : \mathbb{D}_2 \to \mathbb{D}$ is an invertible operator, the derivative of υ is given by $\upsilon'_x = \left(D_{2,\theta}\Upsilon\right)^{-1} D_{1,\theta}\Upsilon$, and υ'_x is continuous at x_0.*

Let $\mathbb{D}$ and $\mathbb{F}$ be Banach spaces, $\mathcal{U} \subset \mathbb{D}$ an open set, $\Upsilon : \mathcal{U} \times \mathbb{R}$ and $\Gamma : \mathcal{U} \to \mathbb{F}$. We say that Υ has a local maximum at $\theta_0 \in \mathcal{U}$ subject to the condition $\Psi(\theta) = 0$ if θ_0 belongs to the surface level $S = \Psi^{-1}(\{0\}) = \{\theta : \Psi(\theta) = 0\}$ and there exists a neighbourhood $\mathcal{U}_{\theta_0} \subset \mathcal{U}$ of θ_0 such that $\Upsilon(\theta) \le \Upsilon(\theta_0)$ for any $\theta \in S \cap \mathcal{U}_{\theta_0}$. Local conditional minimum are defined similarly.

Given $\Psi : \mathcal{U} \to \mathbb{F}$, we say that Ψ is onto $\mathbb{F}$ if for any $y \in \mathbb{F}$ there exists $\theta \in \mathcal{U}$ such that $\Psi(\theta) = y$.

Theorem 2.2. *Let $\mathbb{D}$ and $\mathbb{F}$ be Banach spaces, $\mathcal{U} \subset \mathbb{D}$ an open set, $\Upsilon : \mathcal{U} \times \mathbb{R}$ and $\Psi : \mathcal{U} \to \mathbb{F}$. Let $\theta_0 \in \mathcal{U}$ be such that $\Psi(\theta_0) = 0$. Assume that*

(i) *Υ has a local maximum (or minimum) at θ_0 subject to the condition $\Psi(\theta) = 0$.*
(ii) *Υ is Hadamard differentiable at θ_0.*
(iii) *Ψ is Hadamard differentiable on a neighbourhood of θ_0, Ψ'_θ is onto $\mathbb{F}$ and is continuous at θ_0.*

Then, $\ker(\Psi'_{\theta_0}) \subset \ker(\Upsilon'_{\theta_0})$ *and there exists $\Lambda \in \mathbb{F}^\star$, where $\mathbb{F}^\star$ stands for the dual space, such that $\Upsilon'_{\theta_0} = \Lambda \circ \Psi'_{\theta_0}$.*

Remark 2.4. In particular, if $\mathbb{F} = \mathbb{R}^q$, we have that there exists $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^{\mathrm{T}} \in \mathbb{R}^q$, such that $\Upsilon'_{\theta_0} = \boldsymbol{\gamma}^{\mathrm{T}}\Psi'_{\theta_0}$.

Moreover, if $\mathbb{F} = \mathbb{R}$ and $\mathbb{D}_1$ is a separable Hilbert space, there exists $\gamma \in \mathbb{R}$, such that $\Upsilon'_{\theta_0} = \gamma\Psi'_{\theta_0}$ and $\nabla\Upsilon_{\theta_0} = \gamma\nabla\Psi_{\theta_0}$.

3. Influence function

Usually, in robustness, there are two popular measures of the resistance to outliers of a given estimator: the breakdown point and the influence function of the related functional. Loosely speaking, the breakdown point of an estimator is the smallest fraction of outliers that can take the estimate beyond any bound. On the other hand, the influence function is a measure of robustness with respect to single outliers that allows us to study the local robustness and the asymptotic efficiency of the estimators, providing a rationale for choosing appropriate weight functions and tuning parameters. It can be thought as the first derivative of the functional version of the estimator which, under mild conditions, enables the derivation of the asymptotic covariance matrix of the corresponding estimator. Let $\mathcal{M}$ be the set of all probability measures over $\mathcal{H}$ and a functional $T : \mathcal{M} \to \mathbb{D}$, where $\mathbb{D}$ is a Banach normed space. The influence function of T at P is defined as $\mathrm{IF}(x_0; T, P) = \lim_{\epsilon\to 0}(T(P_{x_0,\epsilon}) - T(P))/\epsilon$, where $P_{x_0,\epsilon} = (1-\epsilon)P + \epsilon\delta_{x_0}$ and δ_{x_0} denotes the probability measure which puts mass 1 at the point x_0 and represents the contaminated model. With the notations given in Section 2, we have that $\mathrm{IF}(x; T, P)$, is the Gateaux derivative at P in the direction $P - \delta_x$ of the functional T.

When $\mathbb{D} = \mathbb{R}^q$, under mild conditions, see [15], it allows to provide a Bahadur expansion for the estimators, i.e., $\sqrt{n}\,\{T(P_n) - T(P)\} = (1/\sqrt{n})\sum_{i=1}^n \mathrm{IF}\,(X_i; T, P) + o_p(1)$, where P_n denotes the empirical probability measure of the observations X_i, $1 \le i \le n$. Therefore, the asymptotic variance of the estimates can be evaluated as $\mathrm{asvar}\,(T, P) = \mathbb{E}_P\,\{\mathrm{IF}\,(X_1, T, P) \otimes \mathrm{IF}\,(X_1, T, P)\}$. Besides being of theoretical interest and helpful to calibrate the efficiency of the robust estimates, measuring the influence of an observation on the classical estimates can be used as a diagnostic tool to detect influential observations, see for instance, [8,32,10,5].

3.1. Raw functional principal components

The following theorem, whose proof is given in the Appendix, provides an expression for the influence function of the functional defined through (3) at an elliptical distribution $\mathcal{E}(\mu, \mathbf{\Gamma})$. It also proves the existence of the influence function. In this case, if $\lambda_1 \geq \lambda_2 \geq \cdots$ denote the eigenvalues of $\mathbf{\Gamma}$ and ϕ_j its associated eigenfunctions, we have that $\phi_{\text{R},j}(P) = \phi_j$ and $\lambda_{\text{R},j}(P) = c\lambda_j$, for some constant c related to the scale functional. So, if $\lambda_1 > \lambda_2 > \cdots > \lambda_q > \lambda_{q+1}$, for $1 \leq j \leq q$, $\phi_{\text{R},j}(P)$ are unique up to a sign change. Without loss of generality, we assume that the location parameter μ equals 0.

Theorem 3.1. *Let P be an elliptical probability measure $P = \mathcal{E}(0, \mathbf{\Gamma})$, where $\mathbf{\Gamma}$ is a self-adjoint, positive semidefinite and compact operator with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$. Denote by ϕ_j the eigenfunction associated to λ_j. Assume that $\lambda_1 > \lambda_2 > \cdots > \lambda_q > \lambda_{q+1}$.*

Let F_0 be the univariate measure defined in Lemma 2.1. Suppose that the map $S : [0, 1] \times \mathbb{R} \to \mathbb{R}$ defined by $S(\epsilon, y) = \sigma_{\text{R}}((1-\epsilon)F_0 + \epsilon\delta_y)$ is twice continuously differentiable at any $(0, y)$. In particular, $\text{IF}(y; \sigma_{\text{R}}, F_0)$ is differentiable and its derivative with respect to y will be denoted $\text{DIF}(y; \sigma_{\text{R}}, F_0)$.

Besides, assume that there exists $\epsilon_0 < 1$, such that for each fixed $\epsilon \in [0, \epsilon_0]$ and $x \in \mathcal{H}$, the map $\Upsilon_{x,\epsilon} : \mathcal{H} \to \mathbb{R}$ defined as $\Upsilon_{x,\epsilon}(\alpha) = \sigma^2_{\text{R}}(P_{x,\epsilon}[\alpha])$ is Hadamard differentiable, where $P_{x,\epsilon} = (1-\epsilon)P + \epsilon\delta_x$ is the contaminated probability measure. Assume further that, for $k \leq q$, the map $\Upsilon : [0, \epsilon_0] \times \mathcal{H} \to \mathbb{R}$ defined as $\Upsilon(\epsilon, \alpha) = \sigma^2_{\text{R}}(P_{x,\epsilon}[\alpha])$ is twice continuously Hadamard differentiable at $(0, \phi_k)$.

Without loss of generality, assume that $\lambda_{\text{R},j} = \lambda_j$ which means that σ_{R} is Fisher-consistent at F_0. Then, for any $k \leq q$, the influence function of the principal direction functional defined through (3) exist and is given by

$$\text{IF}(x; \phi_{\text{R},k}, P) = \sum_{j \geq k+1} \frac{\sqrt{\lambda_k}}{\lambda_k - \lambda_j} \text{DIF}\left(\frac{\langle x, \phi_k\rangle}{\sqrt{\lambda_k}}; \sigma_{\text{R}}, F_0\right) \langle x, \phi_j\rangle \phi_j + \sum_{j=1}^{k-1} \frac{\sqrt{\lambda_j}}{\lambda_k - \lambda_j} \text{DIF}\left(\frac{\langle x, \phi_j\rangle}{\sqrt{\lambda_j}}; \sigma_{\text{R}}, F_0\right) \langle x, \phi_k\rangle \phi_j, \tag{5}$$

while that of the principal values defined through (4) is given by

$$\text{IF}(x; \lambda_{\text{R},k}, P) = 2\, \lambda_k \text{IF}\left(\frac{\langle x, \phi_k\rangle}{\sqrt{\lambda_k}}; \sigma_{\text{R}}, F_0\right). \tag{6}$$

3.2. A functional related to the penalized functional principal components

In this section, we study the influence function of a smoothed functional related to the estimators defined in [3] obtained by penalizing the robust scale. As mentioned in the Introduction, smoothed functionals were also considered in other settings such as nonparametric regression and tolerance intervals. In nonparametric regression, a smoothed functional approach to nonparametric kernel estimators was introduced by Aït Sahalia [1] and used by Tamine [36] to define a smoothed influence function. On the other hand, Fernholz [17] studied the influence function of the smoothed corrected content of tolerance intervals. A general discussion on kernel-smoothed versions of functionals is given in Fernholz [16].

Let us denote $\mathcal{H}_{\text{S}}$, the subset of "smooth elements" of $\mathcal{H}$. Let $D : \mathcal{H}_{\text{S}} \to \mathcal{H}$ be a linear operator, which we will refer to as the "differentiator". Using D, we define the symmetric positive semidefinite bilinear form $\lceil \cdot, \cdot \rceil : \mathcal{H}_{\text{S}} \times \mathcal{H}_{\text{S}} \to \mathbb{R}$, where $\lceil \alpha, \beta \rceil = \langle D\alpha, D\beta\rangle$. The "penalization operator" is then defined as $L : \mathcal{H}_{\text{S}} \to \mathbb{R}$, $L(\alpha) = \lceil \alpha, \alpha \rceil$, and the penalized inner product as $\langle \alpha, \beta\rangle_\tau = \langle \alpha, \beta\rangle + \tau\lceil \alpha, \beta\rceil$. Therefore, $\|\alpha\|^2_\tau = \|\alpha\|^2 + \tau L(\alpha)$.

We want to study the sensitivity to single outliers of the robust scale penalized estimators defined in [3]. For a sample $X_1, \ldots, X_n$ of i.i.d. observations $X_i \in \mathcal{H}$, $X_i \sim X \sim P$, these estimators are given by

$$\begin{cases} \widehat{\phi}_{\text{S},1} = \underset{\|\alpha\|=1}{\operatorname{argmax}} \left\{ s^2_n(\alpha) - \rho L(\alpha) \right\} \\ \widehat{\phi}_{\text{S},m} = \underset{\alpha \in \widehat{\mathcal{B}}_{m,\text{S}}}{\operatorname{argmax}} \left\{ s^2_n(\alpha) - \rho L(\alpha) \right\} \quad 2 \leq m, \end{cases} \tag{7}$$

with $\widehat{\mathcal{B}}_{m,\text{S}} = \{\alpha \in \mathcal{H} : \|\alpha\| = 1, \langle \alpha, \widehat{\phi}_{\text{S},j}\rangle = 0, \ \forall\, 1 \leq j \leq m-1\}$ and $s^2_n(\alpha) = \sigma^2_{\text{R}}(P_n[\alpha])$, where $\sigma_{\text{R}}(P_n[\alpha])$ stands for the functional σ_{R} computed at the empirical distribution of $\langle \alpha, X_1\rangle, \ldots, \langle \alpha, X_n\rangle$. Note that the "raw" estimators correspond to the choice $\rho = 0$. Bali et al. [3] have shown that, under mild condition, if the smoothing parameter ρ varies with the sample size and is such that $\rho = \rho_n \to 0$, then the estimators are consistent to the functional defined through (3). However, the influence function derived in Theorem 3.1, does not allow to measure the sensitivity of the smoothed estimators to anomalous data, since this functional does not depend on the penalization parameter. For that purpose, even if the estimators are consistent to the functional defined through (3), to strength the dependence on the penalization, we will derive, for a fixed value of ρ the influence function of the functional defined through

$$\begin{cases} \phi_{\text{R},\text{S},1}(P) = \underset{\|\alpha\|=1}{\operatorname{argmax}}\, \sigma(\alpha) - \rho L(\alpha) \\ \phi_{\text{R},\text{S},m}(P) = \underset{\|\alpha\|=1, \alpha \in \mathcal{B}_{m,\text{S}}}{\operatorname{argmax}} \sigma(\alpha) - \rho L(\alpha) \quad 2 \leq m, \end{cases} \tag{8}$$

with $\mathcal{B}_{m,s} = \{\alpha \in \mathcal{H} : \langle \alpha, \phi_{\mathrm{R},s,j}(P)\rangle = 0,\ 1 \le j \le m-1\}$. The corresponding principal values are defined as $\lambda_{\mathrm{R},s,m}(P) = \sigma^2_{\mathrm{R}}(P[\phi_{\mathrm{R},s,m}])$. Note that to allow for a proper definition of $\phi_{\mathrm{R},s,m}(P)$ and to derive the influence function of the resulting functionals, we assume that there exists α such that $L(\alpha) < \infty$, i.e., $\mathcal{H}_s \neq \emptyset$. The idea beyond this approach is that the influence of the smooth functional defined in (8) may allow to approximate the sensitivity of the smoothed estimators to anomalous data, since it may be thought as an approximation of the finite-sample version of the influence function introduced by Tukey [37].

Theorem 3.2. *Let P be an elliptical probability measure $P = \mathcal{E}(0, \mathbf{\Gamma})$, where $\mathbf{\Gamma}$ is a self-adjoint, positive semidefinite and compact operator with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots$. Denote as ϕ_j the eigenfunction associated to λ_j. Let F_0 be the univariate measure defined in Lemma 2.1 and assume, without loss of generality, that $\sigma_{\mathrm{R}}(F_0) = 1$, so that $\sigma(\alpha) = \langle \alpha, \mathbf{\Gamma}\alpha\rangle$.*

Denote $\widetilde{\lambda}_k = \lambda_{\mathrm{R},s,k}(P) = \langle \phi_{\mathrm{R},s,k}, \mathbf{\Gamma}\phi_{\mathrm{R},s,k}\rangle$, assume that there exists $q \ge 2$ such that $\widetilde{\lambda}_1 > \widetilde{\lambda}_2 > \cdots \widetilde{\lambda}_q > \widetilde{\lambda}_{q+1}$.

Suppose that the map $S : [0, 1] \times \mathbb{R} \to \mathbb{R}$ defined by $S(\epsilon, y) = \sigma_{\mathrm{R}}((1-\epsilon)F_0 + \epsilon\delta_y)$ is twice continuously differentiable at any $(0, y)$. In particular, $\mathrm{IF}(y; \sigma_{\mathrm{R}}, F_0)$ is differentiable and denote $\mathrm{DIF}(y; \sigma_{\mathrm{R}}, F_0)$ its derivative with respect to y.

Besides, assume that there exists $\epsilon_0 < 1$, such that for each fixed $\epsilon \in [0, \epsilon_0]$ and $x \in \mathcal{H}$, the map $\Upsilon_{x,\epsilon} : \mathcal{H} \to \mathbb{R}$ defined as $\Upsilon_{x,\epsilon}(\alpha) = \sigma^2_{\mathrm{R}}(P_{x,\epsilon}[\alpha])$ is Hadamard differentiable with respect to the norm $\|\cdot\|_\rho$, where $P_{x,\epsilon} = (1-\epsilon)P + \epsilon\delta_x$ is the contaminated probability measure. Assume further that, for $k \le q$, the map $\Upsilon : [0, \epsilon_0] \times \mathcal{H} \to \mathbb{R}$ defined as $\Upsilon(\epsilon, \alpha) = \sigma^2_{\mathrm{R}}(P_{x,\epsilon}[\alpha])$ is twice continuously Hadamard differentiable at $(0, \phi_{\mathrm{R},s,k})$ with respect to the norm $\|\cdot\|_\rho$. Then, for any $k \le q$, the influence function of the principal direction functional $\phi_{\mathrm{R},s,k}(P)$ defined through (8) is given through the implicit solution of

$$\begin{aligned}
&\langle \pi_k A_k, \alpha\rangle - 2\rho\lceil \mathrm{IF}_k, \alpha\rceil + 2\rho\sum_{\ell=1}^{k}\lceil \mathrm{IF}_k, \phi_{\ell,0}\rceil\langle \phi_{\ell,0}, \alpha\rangle \\
&\quad = 2\tilde{\lambda}_k\langle \mathrm{IF}_k, \alpha\rangle + 2\sum_{\ell<k}\langle \mathbf{\Gamma}\phi_{k,0}, \phi_{\ell,0}\rangle\langle \mathrm{IF}_\ell, \alpha\rangle + 2\sum_{\ell=1}^{k}\langle \mathbf{\Gamma}\phi_{k,0}, \mathrm{IF}_\ell\rangle\langle \phi_{\ell,0}, \alpha\rangle \\
&\qquad - 2\rho\sum_{\ell=1}^{k}\lceil \mathrm{IF}_\ell, \phi_{k,0}\rceil\langle \phi_{\ell,0}, \alpha\rangle - 4\rho\sum_{\ell=1}^{k}\lceil \phi_{k,0}, \phi_{\ell,0}\rceil\langle \mathrm{IF}_\ell, \alpha\rangle
\end{aligned} \tag{9}$$

for any $\alpha \in \mathcal{H}$, with $\mathrm{IF}_k = \mathrm{IF}(x, \phi_{\mathrm{R},s,k}, P)$, $\phi_{k,0} = \phi_{\mathrm{R},s,k}(P)$, $A_k = 2\mathbf{\Gamma}\mathrm{IF}_k + \nabla_k$ and

$$\nabla_k = 2\ \mathrm{IF}\left(\frac{\langle x, \phi_{k,0}\rangle}{\sqrt{\widetilde{\lambda}_k}}; \sigma^2_{\mathrm{R}}, F_0\right)\mathbf{\Gamma}\phi_{\mathrm{R},s,k} + \widetilde{\lambda}_k\mathrm{DIF}\left(\frac{\langle x, \phi_{k,0}\rangle}{\sqrt{\widetilde{\lambda}_k}}; \sigma^2_{\mathrm{R}}, F_0\right)\left(\frac{x - \langle x, \phi_{k,0}\rangle\frac{\mathbf{\Gamma}\phi_{k,0}}{\widetilde{\lambda}_k}}{\sqrt{\widetilde{\lambda}_k}}\right).$$

Remark 3.1. As an application of Theorem 3.2, we provide an example in which the operator $\mathbf{\Gamma}$ has as eigenfunctions the elements of the Fourier basis. For that purpose, assume that $\mathcal{H} = L^2(0, 1)$ with its standard inner product. Let $X \sim \mathcal{E}(0, \mathbf{\Gamma})$ with $\mathbf{\Gamma} = \sum_{i\ge1}\lambda_i\phi_i \otimes \phi_i$ where $\lambda_1 > \lambda_2 > \cdots > \lambda_j > \lambda_{j+1} > \cdots$, that is, λ_j converge to 0 in a strictly decreasing way and $\{\phi_i\}_{i\ge1}$ are elements of the Fourier basis. Thus, $\phi_i(x)$ is either $\sin(a_i\pi x)$ or $\cos(a_i\pi x)$, with $a_i \to \infty$, in a non-decreasing way. When the first eigenfunction equals 1, we take $a_1 = 0$. Define $Df = f''$, then, $L(\alpha) = \int_0^1(\alpha''(t))^2dt$. Note that the bilinear form of the penalization is closable, that is there exists a symmetric, non-negative operator $D^{(2)}$ defined on a subspace $\mathcal{V} \subset \mathcal{H}_s$ such that $\lceil \alpha, \beta\rceil = \langle \alpha, D^{(2)}\beta\rangle$, for any $\alpha \in \mathcal{H}_s$ and $\beta \in \mathcal{V}$. Moreover, in this situation $\phi_{\mathrm{R},s,k}(P) = \phi_k$. For technical reasons, we will also require that $\lambda_1 - \lambda_2 - 2\rho a_1^4\pi^4 \neq 0$ and $\lambda_1 - 4\rho a_1^4\pi^4 \neq 0$.

Under these conditions, in the Appendix we derive the following explicit formula for the influence function of the first principal direction,

$$\mathrm{IF}_1 = \mathrm{IF}(\phi_{\mathrm{R},s,1}, x, P) = -\sqrt{\lambda_1}\mathrm{DIF}\left(\frac{\langle x, \phi_1\rangle}{\sqrt{\lambda_1}}; \sigma_{\mathrm{R}}, F_0\right)\sum_{\ell\ge2}\frac{\langle x, \phi_\ell\rangle}{\lambda_\ell - \lambda_1 - \rho(a_\ell^4 - 2\,a_1^4)\pi^4}\phi_\ell. \tag{10}$$

Hence,

$$\|\mathrm{IF}_1\|^2 = \lambda_1\mathrm{DIF}^2\left(\frac{\langle x, \phi_1\rangle}{\sqrt{\lambda_1}}; \sigma_{\mathrm{R}}, F_0\right)\sum_{\ell\ge2}\frac{\langle x, \phi_\ell\rangle^2}{(\lambda_\ell - \lambda_1 - \rho(a_\ell^4 - 2\,a_1^4)\pi^4)^2}.$$

Notice that, when $\rho \to \infty$, we have that $\|\mathrm{IF}_1\|^2 \to 0$. In this sense, the estimators will be more stable when ρ is large. However, as is well known over-smoothing introduces a bias in the estimation procedure. In particular, consistency results require that the smoothing parameter $\rho_n \to 0$ as the sample size n increases, leading to a compromise between consistency and resistance in this setting. Also notice that, as is to be expected, when $\rho = 0$ we get the expression for the influence function given in (5).

4. The influence function for *M*-scale based estimators

To illustrate the performance of the projection-pursuit estimators when considering M-scales, in this section, we give an explicit formula for the influence function of the functionals $\phi_{\mathrm{R},k}$ and $\lambda_{\mathrm{R},k}$ for Gaussian distributions, when the scale functional is an M-estimator. In particular, the expression includes the standard deviation and an M-scale estimator using the bisquare Tukey's function.

Let $\{\phi_j\}_{j\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$ and assume that $X \sim P$, is a Gaussian process with principal components ϕ_j and mean $\mu = 0$, so that $X = \sum_{j\geq 1} Z_j\phi_j$, with $Z_j \sim N(0, \lambda_j)$ independent. Moreover, $\mathbf{\Gamma}_X = \sum_{j\geq 1} \lambda_j\phi_j\otimes\phi_j$ and for any α such that $\alpha \notin \ker(\mathbf{\Gamma}_X)$, we have that $Z_\alpha = \langle\alpha, X-\mu\rangle/\sqrt{\langle\alpha, \mathbf{\Gamma}\alpha\rangle} \sim N(0, 1)$. Hence, the distribution F_0 in Lemma 2.1 equals $F_0 = \Phi$ the standard normal distribution. In this case, the functional defined through (3) is Fisher-consistent, so that $\phi_{\mathrm{R},k}(P) = \phi_k$, while $\lambda_{\mathrm{R},k}(P) = \lambda_k$ if the scale functional is Fisher-consistent at F_0.

On the other hand, the influence function of scale estimators $\sigma_{\mathrm{R}}(G)$ defined through (2) is given by

$$\mathrm{IF}(x; \sigma_{\mathrm{R}}, G) = \sigma_{\mathrm{R}}(G)\left\{\chi\left(\frac{x-\mu(G)}{\sigma_{\mathrm{R}}(G)} - b\right)\right\}\left\{\int \chi'\left(\frac{y-\mu(G)}{\sigma_{\mathrm{R}}(G)}\right)\frac{y-\mu(G)}{\sigma_{\mathrm{R}}(G)}dG(y)\right\}^{-1},$$

see [21]. Thus,

$$\mathrm{DIF}(x; \sigma_{\mathrm{R}}, G) = \chi'\left(\frac{x-\mu(G)}{\sigma_{\mathrm{R}}(G)}\right)\left\{\int \chi'\left(\frac{y-\mu(G)}{\sigma_{\mathrm{R}}(G)}\right)\frac{y-\mu(G)}{\sigma_{\mathrm{R}}(G)}dG(y)\right\}^{-1},$$

with χ' the derivative of χ. Usually, $\sigma_{\mathrm{R}}(G)$ is calibrated to attain Fisher-consistency at the normal distribution, i.e., $\sigma_{\mathrm{R}}(F_0) = 1$ and the location functional is also Fisher-consistency. Therefore, if we denote by $Y \sim N(0, 1)$, we have that the above expressions reduce to

$$\mathrm{IF}(t; \sigma_{\mathrm{R}}, F_0) = \{\chi(t) - b\}\left\{\mathbb{E}_{F_0}\chi'(Y)\,Y\right\}^{-1},$$
$$\mathrm{DIF}(t; \sigma_{\mathrm{R}}, F_0) = \chi'(t)\left\{\mathbb{E}_{F_0}\chi'(Y)\,Y\right\}^{-1}.$$

By plugging these expressions into (5) and (6), we observe that, as in the euclidean setting, the principal values influence function remains bounded if σ_{R} has a bounded influence function, i.e., if χ is bounded. On the other hand, as noted by Croux and Ruiz-Gazen [11], the influence function of the principal directions is related to the behaviour of $\mathrm{DIF}(x; \sigma_{\mathrm{R}}, F_0)$, so that robust scale estimators having a smooth bounded derivative should be preferred. When considering M-scales, this means that the practitioner should prefer a bounded χ function with continuous and bounded derivative χ'. However, even in this case, the influence function for the principal directions may still become unbounded as in the finite-dimensional situation. To be more precise, denote as $x_j = \langle x, \phi_j\rangle$ the components of x in the orthonormal basis $\{\phi_j\}_{j\geq 1}$. Thus, Theorem 3.1 leads to

$$\|\mathrm{IF}(x; \phi_{\mathrm{R},k}, P)\|^2 = \lambda_k\,\mathrm{DIF}^2\left(\frac{x_k}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}, F_0\right)\sum_{j\geq k+1}\frac{x_j^2}{\left(\lambda_k-\lambda_j\right)^2} + x_k^2\sum_{j=1}^{k-1}\frac{\lambda_j}{\left(\lambda_k-\lambda_j\right)^2}\mathrm{DIF}^2\left(\frac{x_j}{\sqrt{\lambda_j}}; \sigma_{\mathrm{R}}, F_0\right). \quad (11)$$

As mentioned in [11], for most robust scale estimators $\mathrm{DIF}(y; \sigma_{\mathrm{R}}, F_0)$ is bounded or even tends to or becomes 0 when $|y|$ converges to ∞. For instance, these assumptions are fulfilled for M-scales when χ' is bounded or $\lim_{t\to\infty}\chi'(t) = 0$, respectively. The choice of χ as the bisquare's Tukey function satisfies both conditions. However, the term x_j^2 in the first term of the right hand side of (11) can still make the influence function to go beyond any limit. As in the finite-dimensional situation, assume that the scale functional σ_{R} has an influence function with bounded derivative redescending to zero. Then, for any principal direction, large values of $|x_1| = |\langle x, \phi_1\rangle|$ will have bounded influence. On the other hand, for any fixed $j > 1$, two situations arise. Large absolute values of $x_j = \langle x, \phi_j\rangle$ have bounded influence on the principal directions $\phi_{\mathrm{R},k}$ when $k > j$, while for $k < j$ they may still yield to a huge influence on the principal direction $\phi_{\mathrm{R},k}$ if they are combined with a small absolute value of x_k. Finally, if $k = j$, large values of $|x_k|$ may have large influence on the principal direction $\phi_{\mathrm{R},k}$ if they are combined with small values of $|x_\ell|$, for $\ell < k$. In this sense, the raw robust estimators of the principal directions based on robust scales, such as M-scales behave as those defined in the finite-dimensional setting. The same comments apply to influence function of the smooth functional.

However, it is worth noting, that if the Gaussian process has finite range, or more generally, if $X \sim \mathcal{E}(0, \mathbf{\Gamma})$, where $\mathbf{\Gamma}$ has finite range, the influence function of the unsmoothed functional will not have a finite expansion as in the finite-dimensional case. To be more precise, assume that $\mathbf{\Gamma}$ has range p, so that $\lambda_i > 0$ for $i \leq p$ and $\lambda_i = 0$ for $i \geq p+1$ Moreover, as in Theorem 3.1 assume that, for some $q \leq p$, $\lambda_1 > \lambda_2 > \cdots > \lambda_q > \lambda_{q+1}$. Then, for any $k \leq q$, (5) reduces to

$$\mathrm{IF}(x; \phi_{\mathrm{R},k}, P) = \mathrm{DIF}\left(\frac{x_k}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}, F_0\right)\frac{1}{\sqrt{\lambda_k}}\sum_{j\geq p+1} x_j\phi_j$$
$$+\,\mathrm{DIF}\left(\frac{x_k}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}, F_0\right)\sum_{k+1\leq j\leq p}\frac{\sqrt{\lambda_k}}{\lambda_k-\lambda_j}x_j\phi_j + x_k\sum_{j=1}^{k-1}\frac{\sqrt{\lambda_j}}{\lambda_k-\lambda_j}\mathrm{DIF}\left(\frac{x_j}{\sqrt{\lambda_j}}; \sigma_{\mathrm{R}}, F_0\right)\phi_j, \quad (12)$$

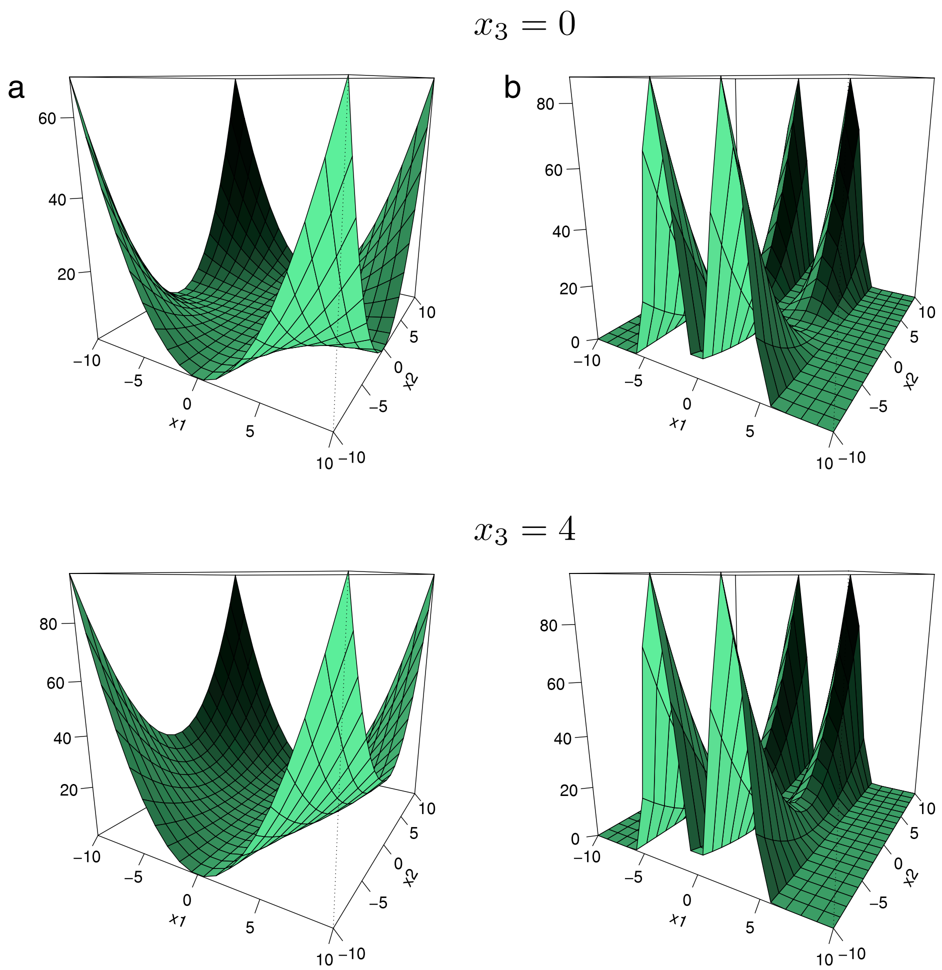

Fig. 1. Norm of the influence function of the largest eigenfunction functional for (a) the classical estimator and (b) the projection-pursuit estimator based on an M-scale when $x_3 = 0, 4$.

where the middle term equals 0 if $k = p$. The last two terms of the above expression correspond to the influence function obtained in Theorem 1 of Croux and Ruiz-Gazen [11]. However, in the infinite-dimensional setting, even in the finite-range situation, $\text{IF}(x; \phi_{\text{R},k}, P)$ has an infinite expansion since the first term on the right hand side of (12) may not be 0. To be more precise, let $\ker(\boldsymbol{\Gamma})^{\perp}$ be the orthogonal of $\ker(\boldsymbol{\Gamma})$. Hence, any element $x \in \ker(\boldsymbol{\Gamma})^{\perp}$ with large values of some of its components $x_j = \langle x, \phi_j \rangle$, for $j > p$ will still have influence on the first p principal directions.

For instance, when considering the standard deviation, using (5), we get that

$$\text{IF}(x; \phi_{\text{SD},k}, P) = \langle x, \phi_k \rangle \sum_{j \neq k} \frac{1}{\lambda_k - \lambda_j} \langle x, \phi_j \rangle \phi_j = x_k \sum_{j \neq k} \frac{1}{\lambda_k - \lambda_j} x_j \phi_j \tag{13}$$

$$\text{IF}(x; \lambda_{\text{SD},k}, P) = \langle x, \phi_k \rangle^2 - \lambda_k = x_k^2 - \lambda_k, \tag{14}$$

so that the classical estimators will have unbounded influence functions as in the finite-dimensional case. On the other hand, as mentioned above, if x is collinear with ϕ_k then $\text{IF}(x; \phi_{\text{SD},k}, P) = 0$. Besides, any point mass contamination on the direction of a principal component will lead to a zero influence, so in order to produce significant modifications on the classical estimators it is necessary for the contamination to propagate in several components. The above expressions are the functional counterpart of those obtained by Croux and Ruiz-Gazen [11] in the finite-dimensional setting and allow to define outlier detection rules by plugging-in a robust estimator if the direction and their sizes in (13) and (14).

As an example, let us consider $X = Z_1\phi_1 + Z_2\phi_2 + Z_3\phi_3$ where $\{\phi_k\}_{k \geq 1}$ is an orthonormal basis and Z_j are independent random variables $Z_j \sim N(0, \lambda_j)$, $\lambda_1 = 16$, $\lambda_2 = 4$ y $\lambda_3 = 1$. To plot the influence function, we consider the classical estimator, based on the standard deviation and the robust one computed with an M-scale estimator with bisquare Tukey's function and tuning constant $c = 1.56$ and breakdown point 1/2. Fig. 1 gives the plots of $\|\text{IF}(x; \phi_{\text{R},1}, P)\|^2$ as a function of x_1 and x_2 when $\sum_{j \geq 4} x_j^2 = 0$ and x_3 equals 0 and 3, respectively where $x_j = \langle \phi_j, x \rangle$, that is, we consider values of x such that $x = x_1\phi_1 + x_2\phi_2 + x_3\phi_3$. Other values of x will lead to similar shapes, when varying x_1 and x_2 since for the selected process

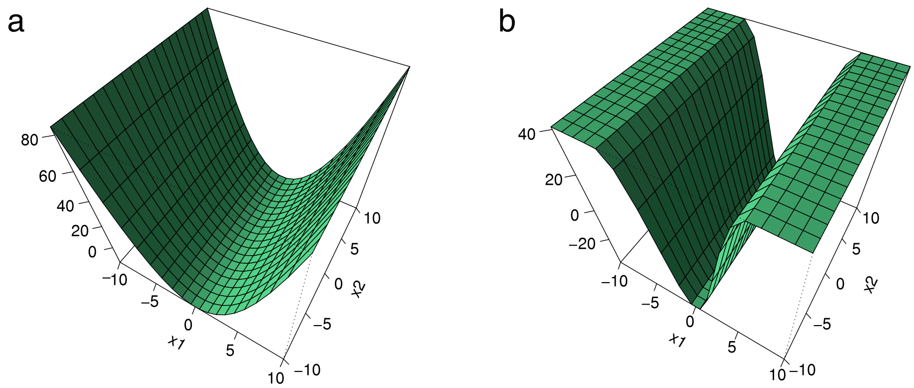

Fig. 2. Influence function of the largest eigenvalue for (a) the classical estimator and (b) the projection-pursuit estimator based on an M-scale when $x_3 = 0$.

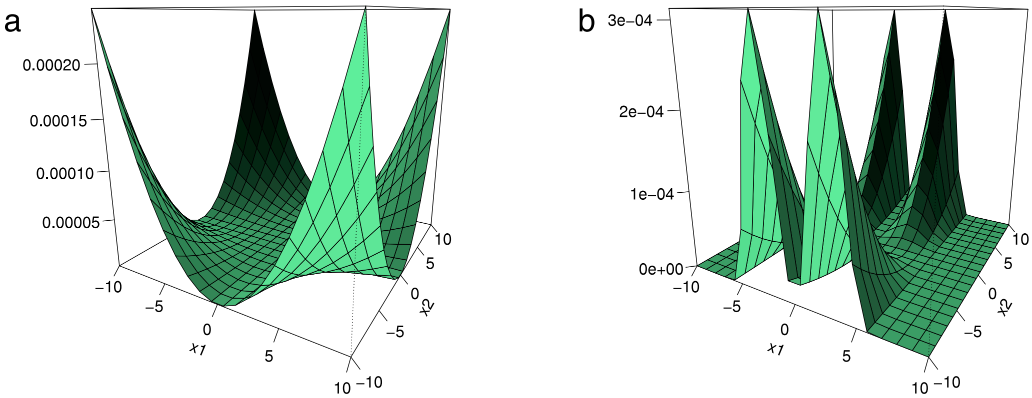

Fig. 3. Norm of the influence function of the largest eigenfunction functional for (a) the classical estimator and (b) the projection-pursuit estimator based on an M-scale when $x_3 = 0$ and $\rho = 0.005$.

$\|\text{IF}(x; \phi_{\text{R},1}, P)\|^2$ and $\|\text{IF}(x; \phi_{\text{SD},k}, P)\|^2$ equal

$$\|\text{IF}(x; \phi_{\text{R},1}, P)\|^2 = \text{DIF}^2\left(\frac{x_1}{\sqrt{\lambda_1}}; \sigma_{\text{R}}, F_0\right) \frac{1}{\lambda_1} \sum_{j\geq 4} x_j^2 + \text{DIF}^2\left(\frac{x_1}{\sqrt{\lambda_1}}; \sigma_{\text{R}}, F_0\right) \sum_{j=2}^{3} \frac{\lambda_1}{(\lambda_1 - \lambda_j)^2} x_j^2,$$

$$\|\text{IF}(x; \phi_{\text{SD},1}, P)\|^2 = x_1^2 \frac{1}{\lambda_1} \sum_{j\geq 4} x_j^2 + x_1^2 \sum_{j=2}^{3} \frac{1}{\lambda_1 - \lambda_j} x_j^2.$$

As expected these plots are analogous to those obtained in the finite-dimensional setting. The shape of the influence function for the projection pursuit estimator based on the M-scale with Tukey's function is comparable to the classical estimator at the centre of the distribution. Observations far away from the centre of the distribution have a much smaller influence by using the robust estimator. However, for the principal directions the squared norm $\|\text{IF}(x; \phi_{\text{R},1}, P)\|^2$ can still attain huge values, but only for smaller values of x_1 combined with huge values of x_2. On the other hand, Fig. 2 gives the influence function of the eigenvalue functional $\lambda_{\text{R},1}(P)$ and confirms the boundedness of $\text{IF}(x; \lambda_{\text{R},k}, P)$ when using the robust scale.

Fig. 3 shows the plots of $\|\text{IF}(\phi_{\text{R,S},1}, x, P)\|^2$ for $\rho = 0.005$. No difference was obtained in the general shape of the plots except a change in the scale. Larger values of ρ imply lower value for the norm of the influence function, which as was noted before it is to be expected.

5. Diagnostic tools for the detection of influential observations

Even if the aim of the paper is to provide an expression for the influence functions of principal component projection-pursuit estimates in an infinite dimensional setting, as an application, in this section we provide a discussion on how the obtained influence functions may be used for diagnostic purposes, to detect observations with a significant impact on the principal direction estimators. As mentioned in the Introduction, besides being of theoretical interest, measuring the influence of an observation on the classical estimates can be used as a diagnostic tool to detect influential observations. In this sense, we are not interested in providing a rule to detect any kind of outliers in functional data, but only to identify observations which influence the principal direction estimation. It is worth noticing that, in general, an outlier may not be an influential observation for the estimation of the quantity of interest, but an influential observation is usually an outlier. As mentioned, for instance, in [27] and Boente et al. [6], an influential observation can be described as an observation with high influence on something, usually an estimate of the parameter of interest. In this sense, the influence function is a useful tool for their detection.

In the finite-dimensional case, several authors have considered this approach. For instance, Croux and Haesbroeck [10] discussed the use of the empirical influence functions in principal component analysis. As described in multivariate and regression settings, the influence function of the functional related to a robust estimator hardly changes when contaminated data points are included in the samples. On the other hand, if we consider the empirical influence of the classical estimators, that is, the influence function of the functional related to the classical estimators replacing now the unknown quantities by the classical estimators, a masking effect may appear and so outlying observations are not detected. A recommended approach is to consider a robustified empirical influence function for the classical estimators, that is, to consider the influence function for the classical estimators but plugging-in robust parameter estimates rather than classical ones, in order to avoid masking and to detect influential observations. This procedure is analogous to the use of the robustified version of Mahalanobis distance introduced by Rousseeuw and van Zomeren [30] and has been considered among others by Pison et al. [26], Boente et al. [5] and Pison and van Aelst [27] to identify the data points that do not obey the model assumptions.

The problem of identifying general atypical observations for functional data is much more complex mainly due to the difficulties of extending to the functional setting the robust Mahalanobis distance introduced in [30]. Besides, in a functional setting outliers may occur in several different ways. As mentioned by Locantore et al. [24] and Hyndman and Shang [22], outlying curves may correspond to atypical trajectories lying outside the range of the vast majority of the data, that is, with extreme values for the L^2 norm, also to isolated points within otherwise typical trajectories (corresponding to a single extreme measurement) or they can be related to an extreme on some principal components, that is, they may be within the range of the data but they may have a different shape from other curves. The latter situation is the most difficult to detect and phenomena showing a combination of these features may also arise.

To identify outlying functional data, several approaches have been considered so far. Some are based on depth measures while others investigate the data performance on some finite-dimensional principal component space. Febrero et al. [13,14] provide a revision on different depth measures and propose a procedure to detect atypical curves. Based on the band depth defined in [25], Sun and Genton [34] developed a functional boxplot and its generalization, the enhanced functional boxplot, which is an extension of the univariate boxplot and which provides a visualization tool for functional data, as well as, a detection rule for potential outliers (see also Sun and Genton [35]). Recently, to identify outlying observations, Gervini [20] considered an interdistance procedure based on the boxplot of the radius of the smallest ball centred at X_i that contains $100\alpha\%$ of the observations. As mentioned by Gervini [20], for outlier-screening purposes it may be useful to consider different values of α.

Also for functional data, Hyndman and Shang [22] propose two graphical procedures to detect atypical observations: the functional bagplot and the highest density region boxplot. Both of them are based on the scores of robust projection pursuit functional principal component estimators and outliers in the functional data are identified as outliers in the bivariate score space. In this sense, the detection rules defined in [22] are related to measures based on the influence function of the principal values considered in Section 3.

As mentioned above, we focus our attention on providing methods to identify observations which may be influential when estimating the principal directions and their size. As discussed above, to measure the influence of an observation on the analysis, the influence function of the classical estimators given in (13) and (14) plugging-in robust estimators of the unknown quantities may be helpful. Given i.i.d. observations $X_1, \dots, X_n$, let $\widehat{\mu}$ be a location estimator, such as the functional spatial median defined in [19]. Furthermore, let $\widehat{\phi}_{\mathrm{R},j}$ stand for the robust estimator of the jth principal direction, such as the unsmoothed projection-pursuit estimator (that is, taking $\rho = 0$ in (7)) and denote as $\widehat{\lambda}_{\mathrm{R},j} = s_n^2(\widehat{\phi}_{\mathrm{R},j})$ the robust estimator of its size.

Note that, when plugging-in $\widehat{\phi}_{\mathrm{R},j}$ and $\widehat{\lambda}_{\mathrm{R},j}$ into (13), only a finite sum can be evaluated, since we are only able to estimate a finite number q of directions. Hence, the quantities

$$\widehat{\mathrm{IF}}_{k,\phi}(x) = \langle x - \widehat{\mu}, \widehat{\phi}_{\mathrm{R},k}\rangle \sum_{1 \le j \ne k \le q} \frac{1}{\widehat{\lambda}_{\mathrm{R},k} - \widehat{\lambda}_{\mathrm{R},j}} \langle x - \widehat{\mu}, \widehat{\phi}_{\mathrm{R},j}\rangle \, \widehat{\phi}_{\mathrm{R},j}$$

$$\widehat{\mathrm{IF}}_{k,\lambda}(x) = \langle x - \widehat{\mu}, \widehat{\phi}_{\mathrm{R},k}\rangle^2 - \widehat{\lambda}_{\mathrm{R},k},$$

will allow the practitioner to evaluate if a given observation may be atypical or influential. More precisely, to measure the influence of X_i on the k-th principal direction or its size, one may plot the values of $\|\widehat{\mathrm{IF}}_{k,\phi}(X_i)\|$ or those of $\widehat{\mathrm{IF}}_{k,\lambda}(X_i)$, respectively, against the index of the observation. Also, a detection rule can be obtained by providing a functional boxplot of the new curves $Y_{i,k} = \widehat{\mathrm{IF}}_{k,\phi}(X_i)$, $1 \le i \le n$, for each value $1 \le k \le q$ of interest, identifying as atypical the trajectories which are labelled as outliers in the boxplot.

Since for Gaussian processes, $\mathrm{var}\left(\mathrm{IF}(X; \lambda_{\mathrm{SD},k}, P)\right) = 2\lambda_k^2$, to investigate the influence of a data X_i on the eigenvalue estimators, one may give a boxplot of the standardized values $\widehat{\mathrm{IF}}_{k,\lambda}(X_i)/(\sqrt{2}\widehat{\lambda}_k)$ or even to consider the aggregate measure for the first q principal values defined in [5] as

$$\mathrm{IM}^2_{\lambda}(x)=\sum_{k=1}^{q}\frac{\widehat{\mathrm{IF}}^2_{k,\lambda}(x)}{2\widehat{\lambda}_k^2}=\sum_{k=1}^{q}\frac{\left\{\langle x-\widehat{\mu},\widehat{\phi}_{\mathrm{R},k}\rangle^2-\widehat{\lambda}_{\mathrm{R},k}\right\}^2}{2\widehat{\lambda}_k^2}.$$

However, in many situations the researcher is interested in detecting curves with high impact on the estimation of some of the principal directions. Assume that we are interested in the first q principal directions. In this case, we also have $\mathrm{var}\left(\mathrm{IF}(X;\phi_{\mathrm{SD},k},P)\right)=\lambda_k\sum_{j\neq k}\left(\lambda_j/(\lambda_k-\lambda_j)^2\right)\phi_j\otimes\phi_j$, where we have assumed that $\lambda_1>\lambda_2>\cdots>\lambda_q$. However, we cannot proceed as in [5] unless we truncate the above operator. Using again that we can only estimate q principal directions, it is sensible to consider the truncated influence function of the classical principal directions

$$\mathrm{IF}_{\mathrm{TR}}(X;\phi_{\mathrm{SD},k},P)=\langle x-\mu,\phi_k\rangle\sum_{1\le j\neq k\le q}\frac{1}{\lambda_k-\lambda_j}\langle x-\mu,\phi_j\rangle\,\phi_j\quad 1\le k\le q.$$

It is worth noting that $\widehat{\mathrm{IF}}_{k,\phi}(x)$ correspond to plugging into $\mathrm{IF}_{\mathrm{TR}}(x;\phi_{\mathrm{SD},k},P)$ the unknown quantities by robust estimators $\widehat{\mu}$, $\widehat{\phi}_{\mathrm{R},j}$ and $\widehat{\lambda}_{\mathrm{R},j}$. As above, for Gaussian processes and, for any $k\le q$, $\mathrm{IF}_{\mathrm{TR}}(X;\phi_{\mathrm{SD},k},P)$ has a covariance operator given by $\lambda_k\sum_{1\le j\neq k\le q}\left(\lambda_j/(\lambda_k-\lambda_j)^2\right)\phi_j\otimes\phi_j$. So, as in [5], one may consider the aggregated diagnostic measure

$$\mathrm{IM}^2_{\phi}(x)=\sum_{k=1}^{q}\sum_{1\le j\neq k\le q}\frac{\langle x-\widehat{\mu},\widehat{\phi}_{\mathrm{R},k}\rangle^2\langle x-\widehat{\mu},\widehat{\phi}_{\mathrm{R},j}\rangle^2}{\widehat{\lambda}_{\mathrm{R},k}\widehat{\lambda}_{\mathrm{R},j}}.$$

Asymptotic cut-off values for $\mathrm{IM}^2_{\lambda}(x)$ and $\mathrm{IM}^2_{\phi}(x)$ are given in [5] for some values of q.

A different approach may be to consider an aggregate measure, similar to that defined in [27], that is

$$\widetilde{\mathrm{IM}}^2_{\phi}(x)=\sum_{k=1}^{q}\|\widehat{\mathrm{IF}}_{k,\phi}(x)\|^2=\sum_{k=1}^{q}\langle x-\widehat{\mu},\widehat{\phi}_{\mathrm{R},k}\rangle^2\sum_{j\neq k}\frac{1}{(\widehat{\lambda}_{\mathrm{R},k}-\widehat{\lambda}_{\mathrm{R},j})^2}\langle x-\widehat{\mu},\widehat{\phi}_{\mathrm{R},j}\rangle^2.$$

Note that $\widetilde{\mathrm{IM}}^2_{\phi}(x)$ corresponds to plugging into $\mathrm{IF}^2_{\mathrm{TR}}(x)=\sum_{k=1}^{q}\|\mathrm{IF}_{\mathrm{TR}}(x;\phi_{\mathrm{SD},k},P)\|^2$ the unknown quantities by robust estimators. The disadvantage of this last method is that, to detect influential points, the cut-off value for the overall influence needs to be computed by Monte Carlo simulation. One possibility is to use a procedure analogous to that considered in [27], to compute the cut-off points. A faster procedure can be implemented by noting that for Gaussian processes $\mathrm{IF}^2_{\mathrm{TR}}(X)$ has the same distribution as $\sum_{1\le j\neq k\le q}\left\{\lambda_j\lambda_k/(\lambda_k-\lambda_j)^2\right\}Z_j^2Z_k^2$, where Z_j are i.i.d. $Z_j\sim N(0,1)$. Hence, one may proceed by generating M times q random variables $N(0,1)$. For each of the datasets the measure $\mathrm{IF}^2_{\mathrm{TR}}(X)$ is computed for all data points replacing λ_j in the above expression by the estimators $\widehat{\lambda}_{\mathrm{R},k}$ obtained from the original sample. This replacement is reasonable because the robust estimators provide a good approximation for the true principal values, if the sample is large enough. The cut-off value is then the $(1-\alpha)$ quantile of the overall influences $\widetilde{\mathrm{IM}}^2_{\phi}(X_i)$. The procedure can be repeated nr times to obtain more stable estimators of the cut-off values. In our simulations we choose, $M=10\,000$, $nr=100$ and we selected the median of the obtained quantiles. In this way, we derive a critical value for the overall influence under the null hypothesis that there are no influential points in the dataset.

In order to evaluate the detection measures introduced, we performed a small simulation study in which we evaluate the capability of $\mathrm{IM}^2_{\lambda}(x)$, $\mathrm{IM}^2_{\phi}(x)$ and $\widetilde{\mathrm{IM}}^2_{\phi}(x)$ to detect the generated atypical observations. In all cases, we choose the sample size $n=100$, $\alpha=0.05$ and $\alpha=0.01$ and we selected $q=4$ principal directions.

A complete comparison with all the procedures discussed in the literature is beyond the scope of the paper. In this study, we only consider as potential competitors, the outlier detection rules provided by: (i) the functional boxplot of the data X_i introduced by Sun and Genton [34] and (ii) the bagplot of the first two scores $\mathbf{V}_i=(\langle X_i-\widehat{\mu},\widehat{\phi}_{\mathrm{R},1}\rangle,\langle x-\widehat{\mu},\widehat{\phi}_{\mathrm{R},2}\rangle)^{\mathrm{T}}$, considered in [22], denoted as FBox and FBag in the Tables, respectively.

We consider the following situations, which correspond to contaminated Gaussian processes with an infinite-dimensional and finite range, respectively:

(a) The uncontaminated or original observations $X_i^{(u)}$ correspond to a Gaussian process with covariance kernel equal to $\gamma_X(s,t)=(1/2)(1/2)^{0.9|s-t|}$, $0<t<1$. The contamination introduced correspond to a peak contamination and has been considered by Sawant et al. [31]. The contaminated observations X_i are defined as

$$X_i(s)=X_i^{(u)}(s)+V_i\,D_i\,M\,\mathbb{I}_{\{T_i<s<T_i+\ell\}}$$

where $V_i\sim Bi(1,p)$, D_i is such that $\mathbb{P}(D_i=1)=\mathbb{P}(D_i=-1)=1/2$, $T_i\sim\mathcal{U}(0,1-\ell)$, $\ell<1/2$ and V_i, X_i, D_i and T_i are independent. We choose $\ell=1/15$ and $p=0.1$. Several values of M were tested $M=2$, $M=5$ and $M=15$, the last two correspond to mild and extreme outliers, respectively while in the first one the atypical data are more difficult to detect.

Table 1
Mean over replications of the proportion of outliers detected PO and of the proportion of good observations not detected as atypical PG when $\alpha = 0.01$.

Model	PO					PG				
	$\widetilde{\text{IM}}^2_\phi(x)$	$\text{IM}^2_\lambda(x)$	$\text{IM}^2_\phi(x)$	FBox	FBag	$\widetilde{\text{IM}}^2_\phi(x)$	$\text{IM}^2_\lambda(x)$	$\text{IM}^2_\phi(x)$	FBox	FBag
(a) $M = 2$	0.2171	0.0964	0.0992	0.0906	0.0282	0.9461	0.9827	0.9810	0.9985	0.9832
(a) $M = 5$	0.6914	0.5988	0.5983	0.5080	0.1131	0.9563	0.9871	0.9851	0.9985	0.9867
(a) $M = 15$	0.9917	0.9916	0.9939	0.5174	0.7127	0.9605	0.9908	0.9882	0.9985	0.9892
(b) $\nu = 4$	0.7419	0.5229	0.5673	0.0098	0.0235	0.9648	0.9852	0.9724	0.9993	0.9885
(b) $\nu = 8$	0.9991	0.9823	0.9960	0.9317	0.5826	0.9533	0.9887	0.9752	0.9994	0.9855

Table 2
Mean over replications of the proportion of outliers detected PO and of the proportion of good observations not detected as atypical PG when $\alpha = 0.05$.

Model	PO					PG				
	$\widetilde{\text{IM}}^2_\phi(x)$	$\text{IM}^2_\lambda(x)$	$\text{IM}^2_\phi(x)$	FBox	FBag	$\widetilde{\text{IM}}^2_\phi(x)$	$\text{IM}^2_\lambda(x)$	$\text{IM}^2_\phi(x)$	FBox	FBag
(a) $M = 2$	0.3686	0.2185	0.2341	0.2351	0.0923	0.8825	0.9432	0.9381	0.9812	0.9403
(a) $M = 5$	0.8309	0.7524	0.7709	0.5207	0.2496	0.9021	0.9551	0.9502	0.9812	0.9484
(a) $M = 15$	0.9970	0.9967	0.9981	0.5208	0.8252	0.9111	0.9661	0.9600	0.9812	0.9562
(b) $\nu = 4$	0.9802	0.7874	0.8698	0.1803	0.1156	0.9096	0.9514	0.9239	0.9893	0.9552
(b) $\nu = 8$	0.9999	0.9975	0.9996	0.9998	0.7554	0.8934	0.9607	0.9313	0.9899	0.9492

Table 3
Mean over replications of the proportion of outliers detected PO and of the proportion of good observations not detected as atypical PG when $\alpha = 0.05$.

Model	PO				PG			
	FB_1	FB_2	FB_3	FB_4	FB_1	FB_2	FB_3	FB_4
(a) $M = 2$	0.2284	0.2666	0.3850	0.4146	0.8848	0.8993	0.8992	0.8960
(a) $M = 5$	0.5026	0.7220	0.8342	0.8506	0.8936	0.9099	0.9095	0.9077
(a) $M = 15$	0.9422	0.9885	0.9890	0.9888	0.9019	0.9160	0.9109	0.9059
(b) $\nu = 4$	0.4712	0.9302	0.9841	0.6860	0.9067	0.9250	0.9224	0.9006
(b) $\nu = 8$	0.8904	0.9991	0.9967	0.9190	0.9051	0.9222	0.9224	0.9046

(b) The observations are such that $X_i(t) = Z_{i1}\phi_1(t) + Z_{i2}\phi_2(t) + Z_{i3}\phi_3(t)$ where $\phi_1(t) = \sin(4\pi t)$, $\phi_2(t) = \cos(7\pi t)$ and $\phi_3(t) = \cos(15\pi t)$ are elements of the Fourier basis on $L^2(-1, 1)$. The distributions of the scores is given by $Z_{i1} \sim N(0, \sigma_1^2)$, $(Z_{i2}, Z_{i3}) \sim (1-\epsilon)\, N\left((0, 0), \text{diag}\left(\sigma_2^2, \sigma_3^2\right)\right) + \epsilon\, N\left((\nu, \nu), \text{diag}\,(0.01, 0.01)\right)$, where $\sigma_1 = 4$, $\sigma_2 = 2$ and $\sigma_3 = 1$. The situation $\epsilon = 0$ corresponds to the uncontaminated data, while in the situation under study $\epsilon = 0.1$, so that the observations are contaminated in the diagonal between the second and third principal directions. Two values of ν are studied, $\nu = 4$ and $\nu = 8$ corresponding to influential points that can be seen as mild and extreme outliers.

Tables 1 and 2 summarize the results obtained for all the models described above and for the methods based on $\text{IM}^2_\lambda(x)$, $\text{IM}^2_\phi(x)$ and $\widetilde{\text{IM}}^2_\phi(x)$, for $\alpha = 0.01$ and 0.05 respectively. The reported results correspond to the mean over replications of the proportion of atypical data PO detected and to the mean over replications of proportion of good observations not detected as atypical PG. These two measures can be seen as measures of the sensitivity and specificity of the detection rules, respectively.

For fair comparisons with the proposed detection rules, when considering the bagplot of $\mathbf{V}_i$, we choose as the factors to define the fence $\kappa = 2.58$ (see [22]), and also $\kappa = 2.079$. When the projected bivariate scores have a normal distribution, $\kappa = 2.58$ allows the fence to contain a $100 \times (1-\alpha)\%$ of the observations, with $\alpha = 0.01$, while $\kappa = 2.079$ corresponds to $\alpha = 0.05$. Similarly, for the functional boxplot, the fences were obtained with two factors $\kappa_{\text{FBOX}} = 1.5$ and $\kappa_{\text{FBOX}} = 1$ which correspond to inflate the envelope of the 50% central region κ times the range of the 50% central region. The results corresponding to $\kappa_{\text{FBOX}} = 1.5$ are reported in Table 1, while those related to $\kappa_{\text{FBOX}} = 1$ in Table 2, since for an univariate boxplot, they correspond to rejecting outliers with a probability of $\alpha = 0.01$ and 0.05, respectively, when the data are normal.

Table 3 report similar measures when considering the functional boxplot of the new curves $Y_{i,k} = \widehat{\text{IF}}_{k,\phi}(X_i)$, $1 \le i \le n$, which are denoted as FB_k for each $1 \le k \le 4$ taking $\kappa_{\text{FBOX}} = 1.5$. Hence, they should be compared with the results given in Table 1.

The obtained results show that $\widetilde{\text{IM}}^2_\phi(x)$, $\text{IM}^2_\lambda(x)$, $\text{IM}^2_\phi(x)$, the functional boxplot of the data and the functional bagplot all have a high specificity meaning that in most cases, the uncontaminated data are not declared as influential except for the percentage allowed by the selection of α or the cut-off value given in the functional boxplot and bagplot respectively. The lower values are obtained when considering $\widetilde{\text{IM}}^2_\phi(x)$ which tends to identify more outliers than actually exist. On the other hand, the specificity of the functional boxplots FB_k of the new curves $Y_{i,k}$ is much smaller, indicating that in many situations the uncontaminated trajectories are declared as influential, identifying in this way spurious outliers.

With respect to the proportion of detected outliers, $\widetilde{\text{IM}}^2_\phi(x)$ seems to have the best performance followed by $\text{IM}^2_\phi(x)$ that also gives accurate results for the considered models. The better behaviour of $\widetilde{\text{IM}}^2_\phi(x)$ is at the expense of a loss of

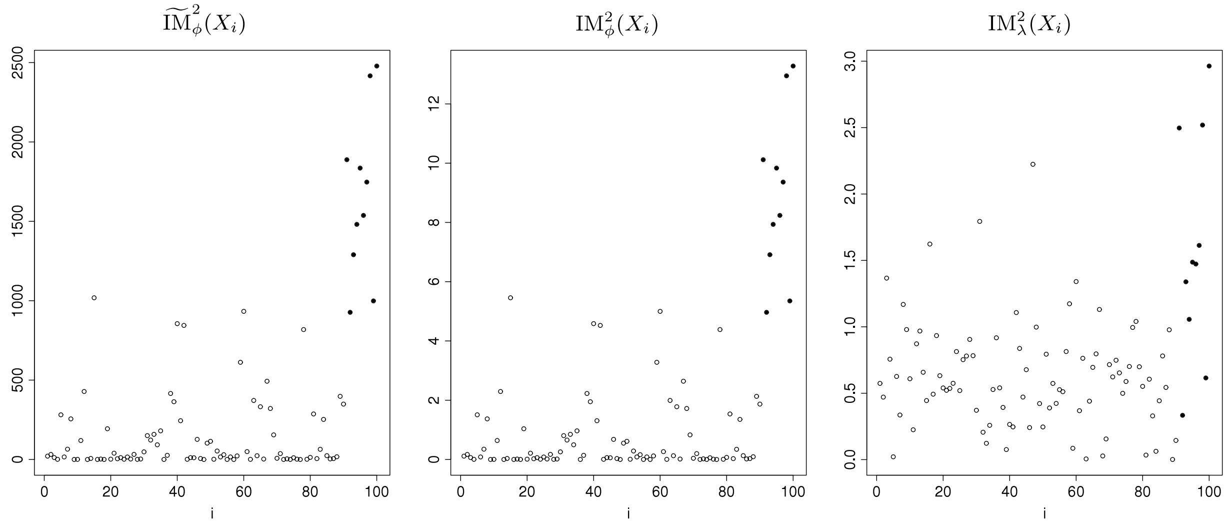

Fig. 4. Plots of $\widetilde{\mathrm{IM}}_{\phi}^{2}(X_i)$, $\mathrm{IM}_{\phi}^{2}(X_i)$ or $\mathrm{IM}_{\lambda}^{2}(X_i)$, against the index of the observation, when $n = 100$.

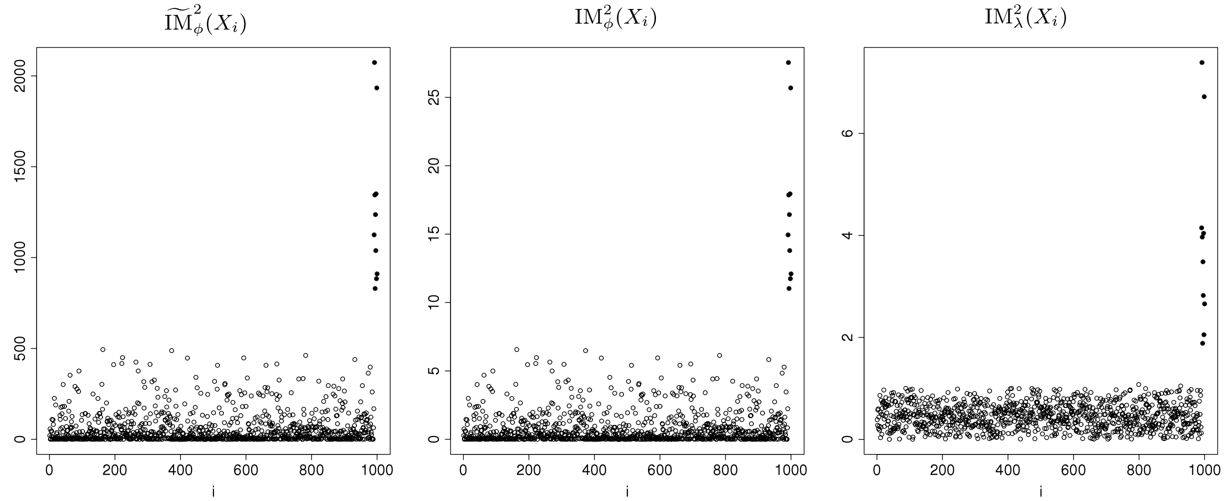

Fig. 5. Plots of $\widetilde{\mathrm{IM}}_{\phi}^{2}(X_i)$, $\mathrm{IM}_{\phi}^{2}(X_i)$ or $\mathrm{IM}_{\lambda}^{2}(X_i)$, against the index of the observation, when $n = 1000$.

specificity and also with a larger numerical cost. In most cases, the functional boxplot is not able to detect the contaminated observations. This is mainly due to the fact that the contaminations do not change the location (unless M or ν are large enough) but rather affects the principal directions. As mentioned by several authors, the functional depth approach does not take shape outliers into account, so that the functional boxplot fail to detect some outliers that are not far from the median curve. On the other hand, the functional bagplot looses its ability to detect the introduced atypical data when their effect on the first two scores is not large, as is the case when $\nu = 4$ and $M = 2$ or 5. In this sense, the functional boxplots FB_k of the new curves $Y_{i,k}$ provide an idea on which direction the contaminated data have their main influence.

Finally, we have also considered an example based on the model studied in [22], i.e., $X_i(t) = a_i \sin(2\pi t) + b_i \cos(2\pi t)$, with $0 < t < 1$, a_i and b_i i.i.d. such that $a_i \sim \mathcal{U}(0, 0.1)$ and $b_i \sim \mathcal{U}(0, 0.1)$. In their study, Hyndman and Shang [22] replaced 10 of the original observations by new observations such that $X_i^{(n)}(t) = a_i \sin(2\pi t) + b_i \cos(2\pi t)$, where a_i and b_i are i.i.d. but their distribution is now $\mathcal{U}(0.1, 0.12)$. In this situation, the two principal directions have the same size, and the process is not Gaussian. Hence, it is not appropriate to use cut-off values for $\mathrm{IM}_{\lambda}^{2}(x)$ and $\mathrm{IM}_{\phi}^{2}(x)$ given in [5], since they may lead to wrong conclusions. However, one may still plot $\widetilde{\mathrm{IM}}_{\phi}^{2}(X_i)$, $\mathrm{IM}_{\phi}^{2}(X_i)$ or $\mathrm{IM}_{\lambda}^{2}(X_i)$, against the index of the observation to identify the observations with larger values as candidates for atypical. Figs. 4 and 5 show the obtained plots when $n = 100$ and $n = 1000$, the 10 outliers correspond to the last ten generated data. We have chosen $q = 2$. It is worth noting that when $n = 100$, $\mathrm{IM}_{\lambda}^{2}(X_i)$ does not allow to clearly distinguish the outliers, moreover, some good observations may be considered as influential, leading to a high false positive rate. On the other hand, for this sample size, when considering $\widetilde{\mathrm{IM}}_{\phi}^{2}(X_i)$ or

$\mathrm{IM}^2_{\phi}(X_i)$, the generated shape outliers, attain the largest values. Finally, when $n = 1000$ which is the situation studied in [22], all measures are able to identify the anomalous observations. It is worth noting that when using the functional HDR boxplot, with $\alpha = 0.01$ then the ten outliers are identified when $n = 1000$, while the observation $i = 59$ is the only one classified as atypical when $n = 100$.

The above results show that, at least for the situations considered, the diagnostic measures based on the influence function have a better performance over other detection rules. Hence, they may be considered as helpful tools for detecting influential observations for the principal directions.

6. Conclusions

In this paper, we derive a general expression for the influence function of the functional related to the projection-pursuit estimators of the principal directions. As expected, the influence function of the classical estimators based on the standard deviation is not bounded. On the other hand, as in the multivariate case, using a robust scale as projection index, the eigenvalue influence function remains bounded. However, this property does not hold for the principal directions. Based on the influence function, we have introduced diagnostic measures which allow to detect trajectories suspicious of having an effect on the principal directions estimation. We have also obtained an implicit expression for the influence function of a smoothed functional related to the projection-pursuit estimators which penalize the scale. Though perhaps difficult to handle, nevertheless for some special cases we obtain an explicit formula which agrees with can be expected from the behaviour of the estimators: large values of ρ tend to produce over-smoothing. One of the consequences of over-smoothing is that the functional related to the principal directions will be more stable. However, this may entail losing Fisher-consistency, i.e., we may be estimating a direction different from the one of interest.

Acknowledgments

This research was partially supported by Grants PIP 112-200801-00216 from CONICET, PICT 2011-397 from ANPCYT and W276 from the Universidad de Buenos Aires at Buenos Aires, Argentina. The authors wish to thank the Associate Editor and two anonymous referees for valuable comments which led to an improved version of the original paper.

Appendix

A.1. Proof of Lemma 2.1

First note that since $X \sim \mathcal{E}(\mu, \mathbf{\Gamma}, \psi)$, we have that $X - \mu \sim \mathcal{E}(0, \mathbf{\Gamma}, \psi)$, so that $\langle \alpha, X - \mu \rangle$ has a symmetric distribution. Hence, we only have to prove that the distribution of $\langle \alpha, X - \mu \rangle / \sqrt{\langle \alpha, \mathbf{\Gamma}\alpha \rangle}$ will not depend on α, when $\langle \alpha, \mathbf{\Gamma}\alpha \rangle \neq 0$, which will follow from the fact that its characteristic function does not depend on α.

If we define $A : \mathcal{H} \to \mathbb{R}$ as $Ax = \langle \alpha, x \rangle$, we have that $X_\alpha = A(X - \mu) \sim \mathcal{E}(0, A\mathbf{\Gamma}A^*, \psi) = \mathcal{E}(0, \langle \alpha, \mathbf{\Gamma}\alpha \rangle, \psi)$. Therefore, the characteristic function of X_α is given by $\varphi_{X_\alpha}(y) = \psi(\langle \alpha, \mathbf{\Gamma}\alpha \rangle y^2)$. When $\langle \alpha, \mathbf{\Gamma}\alpha \rangle = 0$, we have that $\varphi_{X_\alpha}(y) = 1$ for any y, so that $\mathbb{P}(X_\alpha = 0) = 1$. On the other hand, if $\sigma_\alpha = \langle \alpha, \mathbf{\Gamma}\alpha \rangle \neq 0$, we get easily that $\varphi_{X_\alpha/\sigma_\alpha}(y) = \varphi_{X_\alpha}(y/\sigma_\alpha) = \psi(y^2)$ which does not depend on α, concluding the proof. $\square$

A.2. Proof of Theorem 3.1

We begin by fixing our notation and making some computations that will be used in the sequel. Denote by $I_{\mathcal{H}}$ the identity operator, i.e., $I_{\mathcal{H}}(\alpha) = \alpha$.

If V stands for the random variable independent of $X \sim P$ such that $\mathbb{P}(V = 1) = \epsilon$ and $\mathbb{P}(V = 0) = 1 - \epsilon$, we have that $(1 - V)X + V\delta_x, \sim P_{x,\epsilon}$.

For any $\epsilon < \epsilon_0,\ x \in \mathcal{H}$ denote as $\phi_{j,\epsilon} = \phi_{\mathrm{R},j}(P_{x,\epsilon})$ and $\lambda_{j,\epsilon} = \lambda_{\mathrm{R},j}(P_{x,\epsilon})$. Recall that $\|\phi_{j,\epsilon}\| = 1$ and $\langle \phi_{j,\epsilon}, \phi_{\ell,\epsilon} \rangle = 0$, for $\ell \neq j$.

Let P be as in Theorem 3.1, i.e., an elliptical probability measure $P = \mathcal{E}(\mu, \mathbf{\Gamma})$, where $\mu = 0$ and $\mathbf{\Gamma}$ is a self-adjoint, positive semidefinite and compact operator with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$ such that $\lambda_1 > \lambda_2 > \cdots > \lambda_q > \lambda_{q+1}$. Define $\Upsilon : [0, \epsilon_0] \to \mathbb{R}$ as $\Upsilon(\epsilon, \alpha) = \sigma^2_{\mathrm{R}}(P_{x,\epsilon}[\alpha])$.

Let $k \leq q$ and define the restrictions $\mathbf{\Psi} : \mathcal{H} \to \mathbb{R}^k$ as $\mathbf{\Psi}(\alpha) = (\Psi_0(\alpha), \ldots, \Psi_{k-1}(\alpha))$ with $\Psi_0(\alpha) = \|\alpha\|^2 - 1$ and $\Psi_j(\alpha) = \langle \alpha, \phi_{j,\epsilon} \rangle$, for $1 \leq j \leq k - 1$, where we understand that when $k = 1$, $\mathbf{\Psi}(\alpha) = \Psi_0(\alpha)$. From Remark 2.2, $\mathbf{\Psi}$ is $\mathcal{C}^1$-Fréchet differentiable. Moreover, $\mathbf{\Psi}'_{\phi_{k,\epsilon}}$ is onto $\mathbb{R}^k$ since $\mathbf{\Psi}'_{\phi_{k,\epsilon}} = (\Psi'_{0,\phi_{k,\epsilon}}, \ldots, \Psi'_{k-1,\phi_{k,\epsilon}})$ with

$$\Psi'_{0,\phi_{k,\epsilon}}(\alpha) = \langle 2\phi_{k,\epsilon}, \alpha \rangle \qquad \Psi'_{j,\phi_{k,\epsilon}}(\alpha) = \langle \phi_{j,\epsilon}, \alpha \rangle, \quad \text{for } 1 \leq j \leq k - 1.$$

Effectively, given $\mathbf{y} = (y_1, \ldots, y_k)^{\mathrm{T}} \in \mathbb{R}^k$, consider $\alpha = (y_1/2)\phi_{k,\epsilon} + \sum_{j=1}^{k-1} y_{j+1}\phi_{j,\epsilon}$, then $\mathbf{\Psi}'_{\phi_{k,\epsilon}}(\alpha) = \mathbf{y}$. Thus, assumption (iii) in Theorem 2.2 is satisfied.

From (3), we have that, for any fixed ϵ, $\phi_{k,\epsilon}$ maximizes $\Upsilon_{x,\epsilon}(\alpha) = \sigma^2_{\text{R}}(P_{x,\epsilon}[\alpha])$ over the surface $\{\alpha : \mathbf{\Psi}(\alpha) = 0\}$, that is, $\Upsilon_{x,\epsilon}(\alpha)$ has a local maximum at $\phi_{k,\epsilon}$ subject to the condition $\mathbf{\Psi}(\alpha) = 0$. Besides, $\Upsilon_{x,\epsilon} : \mathcal{H} \to \mathbb{R}$ is a Hadamard differentiable function.

Hence, keeping in mind that $\Upsilon_{x,\epsilon}$ depends on ϵ and denoting as $\Lambda(\alpha) = \Upsilon(\epsilon, \alpha) = \Upsilon_{x,\epsilon}(\alpha)$, we have that, Theorem 2.2 entails that there exist $\gamma_0, \gamma_1, \dots, \gamma_{k-1} \in \mathbb{R}$ (depending on ϵ) such that

$$\Lambda'_{\phi_{k,\epsilon}} = \sum_{j=0}^{k-1} \gamma_j \Psi'_{j,\phi_{k,\epsilon}}.$$

It is worth noting that $\Lambda'_{\phi_{k,\epsilon}} = D_{2,\theta_\epsilon}\Upsilon$ with $\theta_\epsilon = (\epsilon, \phi_{k,\epsilon}) \in [0, \epsilon_0] \times \mathcal{H}$ with $D_{2,(\epsilon,\alpha)}\Upsilon$ the first partial Hadamard derivative with respect to the second component of Υ at (ϵ, α).

The fact that $\Lambda'_{\phi_{k,\epsilon}}$ is a linear and continuous operator (see Remark 2.1) implies that there exists a unique $\nabla\Lambda_{\phi_{k,\epsilon}} \in \mathcal{H}$ such that $\Lambda'_{\phi_{k,\epsilon}}(\alpha) = \langle \nabla\Lambda_{\phi_{k,\epsilon}}, \alpha\rangle$ and similarly, $\Psi'_{j,\phi_{k,\epsilon}}(\alpha) = \langle \nabla\Psi_{j,\phi_{k,\epsilon}}, \alpha\rangle$. Hence, $\nabla\Lambda_{\phi_{k,\epsilon}} = \sum_{j=0}^{k-1}\gamma_j\nabla\Psi_{j,\phi_{k,\epsilon}}$. On the other hand, from Remark 2.2 we get that $\nabla\Psi_{0,\phi_{k,\epsilon}} = 2\phi_{k,\epsilon}$ while $\nabla\Psi_{j,\phi_{k,\epsilon}} = \phi_{j,\epsilon}$, for $1 \le j \le k-1$ which entails that

$$\nabla\Lambda_{\phi_{k,\epsilon}} = 2\gamma_0\phi_{k,\epsilon} + \sum_{j=1}^{k-1}\gamma_j\phi_{j,\epsilon}. \tag{A.1}$$

From (A.1), we obtain that $2\gamma_0 = \langle \nabla\Lambda_{\phi_{k,\epsilon}}, \phi_{k,\epsilon}\rangle$ while $\gamma_j = \langle \nabla\Lambda_{\phi_{k,\epsilon}}, \phi_{j,\epsilon}\rangle$. Hence, (A.1) can be written as

$$\nabla\Lambda_{\phi_{k,\epsilon}} = \sum_{j=1}^{k}\langle \nabla\Lambda_{\phi_{k,\epsilon}}, \phi_{j,\epsilon}\rangle\phi_{j,\epsilon}, \tag{A.2}$$

that is,

$$(D_{2,\theta_\epsilon}\Upsilon)\alpha = \Lambda'_{\phi_{k,\epsilon}}\alpha = \sum_{j=1}^{k}\langle \nabla\Lambda_{\phi_{k,\epsilon}}, \phi_{j,\epsilon}\rangle\langle\phi_{j,\epsilon}, \alpha\rangle. \tag{A.3}$$

We will use this expression several times in the sequel.

We will need the following lemma which is an extension of an analogous result obtained in the multivariate setting.

Lemma A.2.1. *Let P be an elliptical probability measure $P = \mathcal{E}(\mu, \mathbf{\Gamma})$, where $\mu = 0$ and $\mathbf{\Gamma}$ is a self-adjoint, positive semi-definite and compact operator with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots$ and let ϕ_j be the eigenfunction associated to λ_j. Assume that $\lambda_1 > \lambda_2 > \cdots > \lambda_q > \lambda_{q+1}$.*

Let F_0 be the univariate measure defined in Lemma 2.1. Suppose that the map $S : [0, 1] \times \mathbb{R} \to \mathbb{R}$ defined by $S(\epsilon, y) = \sigma_{\text{R}}((1-\epsilon)F_0 + \epsilon\delta_y)$ is twice continuously differentiable at any $(0, y)$. In particular, $\text{IF}(y; \sigma_{\text{R}}, F_0)$ needs to be differentiable with respect to y and its derivative will be denoted by $\text{DIF}(y; \sigma_{\text{R}}, F_0)$.

Then, for any $k \le q$, there exists a neighbourhood $\mathcal{U}_k$ of ϕ_k such that for any $\alpha \in \mathcal{U}_k$,

$$\text{IF}(\langle\alpha, x\rangle; \sigma^2_{\text{R}}, P[\alpha]) = \langle\alpha, \mathbf{\Gamma}\alpha\rangle\,\text{IF}\left(\frac{\langle\alpha, x\rangle}{\sqrt{\langle\alpha, \mathbf{\Gamma}\alpha\rangle}}; \sigma^2_{\text{R}}, F_0\right). \tag{A.4}$$

Moreover, if $\Lambda : \mathcal{H} \to \mathbb{R}$ stands for the map $\Lambda(\alpha) = \text{IF}(\langle\alpha, x\rangle; \sigma^2_{\text{R}}, P[\alpha])$, then Λ is Hadamard differentiable at ϕ_k and

$$\nabla\Lambda_{\phi_k} = 2\lambda_k\,\text{IF}\left(\frac{\langle\phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma^2_{\text{R}}, F_0\right)\phi_k + \lambda_k\text{DIF}\left(\frac{\langle\phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma^2_{\text{R}}, F_0\right)\left(\frac{x - \langle x, \phi_k\rangle\phi_k}{\sqrt{\lambda_k}}\right), \tag{A.5}$$

that is,

$$\Lambda'_{\phi_k}(\alpha) = 2\lambda_k\,\text{IF}\left(\frac{\langle\phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma^2_{\text{R}}, F_0\right)\langle\phi_k, \alpha\rangle + \lambda_k\text{DIF}\left(\frac{\langle\phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma^2_{\text{R}}, F_0\right)\left\langle\alpha, \frac{x - \langle x, \phi_k\rangle\phi_k}{\sqrt{\lambda_k}}\right\rangle.$$

Proof. Using that $\lambda_k > 0$, we get there exists a neighbourhood $\mathcal{U}_k$ of ϕ_k such that for any $\alpha \in \mathcal{U}_k$, $\alpha \notin \ker(\mathbf{\Gamma})$, hence Lemma 2.1 entails that $Z_\alpha = \langle\alpha, X\rangle/\sqrt{\langle\alpha, \mathbf{\Gamma}\alpha\rangle} \sim F_0$. Let us recall that

$$\text{IF}(\langle\alpha, x\rangle; \sigma^2_{\text{R}}, P[\alpha]) = \lim_{\epsilon\to 0}\frac{\sigma^2_{\text{R}}((1-\epsilon)P[\alpha] + \epsilon\delta_{\langle x,\alpha\rangle}) - \sigma^2_{\text{R}}(P[\alpha])}{\epsilon}. \tag{A.6}$$

Using that σ_{R} is a scale functional and the fact that $\langle\alpha, X\rangle \sim P[\alpha]$, we get that $\sigma_{\text{R}}(P[\alpha]) = \sqrt{\langle\alpha, \mathbf{\Gamma}\alpha\rangle}\sigma_{\text{R}}(F_0)$.

On the other hand, if V stands for a random variable independent of X such that $\mathbb{P}(V = 1) = \epsilon$ and $\mathbb{P}(V = 0) = 1 - \epsilon$, we have that $(1 - V)\langle X, \alpha\rangle + V\langle x, \alpha\rangle \sim P_{x,\epsilon}[\alpha] = (1-\epsilon)P[\alpha] + \epsilon\delta_{\langle x,\alpha\rangle}$. Noting that $(1 - V)\langle X, \alpha\rangle + V\langle x, \alpha\rangle =$

$\sqrt{\langle\alpha, \mathbf{\Gamma}\alpha\rangle}\,((1-V)Z_\alpha + Vz_\alpha)$ with $z_\alpha = \langle\alpha, x\rangle/\sqrt{\langle\alpha, \mathbf{\Gamma}\alpha\rangle}$ and that $(1-V)Z_\alpha + Vz_\alpha \sim (1-\epsilon)F_0 + \epsilon\delta_{z_\alpha}$, we also get that

$$\sigma_{\mathrm{R}}((1-\epsilon)P[\alpha] + \epsilon\delta_{\langle x,\alpha\rangle}) = \sqrt{\langle\alpha, \mathbf{\Gamma}\alpha\rangle}\,\sigma_{\mathrm{R}}((1-\epsilon)F_0 + \epsilon\delta_{z_\alpha}),$$

which using (A.6) entails (A.4).

It remains to show (A.5). Define the maps $\Upsilon : \mathcal{U}_k \to \mathbb{R}$ and $\Psi : \mathcal{U}_k \to \mathbb{R}$ as $\Upsilon(\alpha) = \langle\alpha, \mathbf{\Gamma}\alpha\rangle$ and

$$\Psi(\alpha) = \mathrm{IF}\left(\frac{\langle\alpha, x\rangle}{\sqrt{\langle\alpha, \mathbf{\Gamma}\alpha\rangle}}; \sigma_{\mathrm{R}}^2, F_0\right).$$

Then, $\Lambda(\alpha) = \mathrm{IF}(\langle\alpha, x\rangle; \sigma_{\mathrm{R}}^2, P[\alpha]) = \Upsilon(\alpha)\Psi(\alpha)$. Remark 2.2 entails that Υ is Hadamard differentiable with $\nabla\Upsilon_\alpha = 2\mathbf{\Gamma}\alpha$. Hence, if we show that Ψ is Hadamard differentiable at ϕ_k, we get that Λ is also Hadamard differentiable ϕ_k and

$$\nabla\Lambda_{\phi_k} = \Upsilon(\phi_k)\nabla\Psi_{\phi_k} + \Psi(\phi_k)\nabla\Upsilon_{\phi_k}. \tag{A.7}$$

Using that ϕ_k is the eigenfunction of $\mathbf{\Gamma}$ associated to λ_k, we get that $\Upsilon(\phi_k) = \lambda_k$, $\nabla\Upsilon_{\phi_k} = 2\mathbf{\Gamma}\phi_k = 2\lambda_k$ and

$$\Psi(\phi_k) = \mathrm{IF}\left(\frac{\langle\phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}^2, F_0\right),$$

which together with (A.7) entail that

$$\nabla\Lambda_{\phi_k} = \lambda_k\nabla\Psi_{\phi_k} + 2\lambda_k\mathrm{IF}\left(\frac{\langle\phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}^2, F_0\right)\phi_k.$$

Thus, to conclude the proof, it remains to show that Ψ is Hadamard differentiable at ϕ_k and

$$\nabla\Psi_{\phi_k} = \mathrm{DIF}\left(\frac{\langle\phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}^2, F_0\right)\left(\frac{x - \langle x, \phi_k\rangle\phi_k}{\sqrt{\lambda_k}}\right).$$

Define now $f : \mathbb{R} \to \mathbb{R}$ as $f(y) = \mathrm{IF}(y; \sigma_{\mathrm{R}}, F_0)$ and $\Phi : \mathcal{U}_k \to \mathbb{R}$ as $\Phi(\alpha) = \langle x, \alpha\rangle/\sqrt{\langle\alpha, \mathbf{\Gamma}\alpha\rangle} = \langle x, \alpha\rangle/\sqrt{\Upsilon(\alpha)}$. Then, the Hadamard differentiability of Ψ at ϕ_k follows easily from the chain rule and the fact that f is differentiable, Φ is Hadamard differentiable at ϕ_k, since $\Upsilon(\phi_k) = \lambda_k \neq 0$, and $\Psi = f \circ \Phi$. Moreover, the chain rule entails that $\Psi'_{\phi_k} = f'_{\Phi(\phi_k)} \circ \Phi'_{\phi_k}$, which together with the fact that $f'_{\Phi(\phi_k)}(v) = f'(\Phi(\phi_k))v$ and $\Phi(\phi_k) = \langle\phi_k, x\rangle/\sqrt{\lambda_k}$, $f(y) = \mathrm{DIF}(y; \sigma_{\mathrm{R}}, F_0)$ imply that

$$\begin{aligned}\Psi'_{\phi_k}(\alpha) &= f'_{\Phi(\phi_k)}\left(\Phi'_{\phi_k}(\alpha)\right) = f'(\Phi(\phi_k))\Phi'_{\phi_k}(\alpha)\\ &= \mathrm{DIF}\left(\frac{\langle\phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}, F_0\right)\Phi'_{\phi_k}(\alpha).\end{aligned}$$

Using again Remark 2.2 and the fact that $\mathbf{\Gamma}\phi_k = \lambda_k$, we get that

$$\Phi'_{\phi_k}(\alpha) = \frac{\langle x, \alpha\rangle\sqrt{\lambda_k} - \frac{\langle\alpha, \mathbf{\Gamma}\phi_k\rangle}{\sqrt{\lambda_k}}\langle x, \phi_k\rangle}{\lambda_k} = \frac{\langle x, \alpha\rangle - \langle\phi_k, \alpha\rangle\langle x, \phi_k\rangle}{\sqrt{\lambda_k}} = \left\langle\frac{x - \langle x, \phi_k\rangle\phi_k}{\sqrt{\lambda_k}}, \alpha\right\rangle,$$

concluding the proof. □

Proof of Theorem 3.1. Since, we have assumed that the scale functional σ_{R} is calibrated so that $c = 1$, i.e., $\sigma_{\mathrm{R}}(F_0) = 1$, we have that $\sigma(\alpha) = \langle\alpha, \mathbf{\Gamma}\alpha\rangle$.

We begin by proving the existence of the influence function. For that purpose, we will prove the existence of the influence function of the functional related to the first principal direction. With no loss of generality, we will assume that the scale functional is calibrated so that $\sigma_{\mathrm{R}}^2(P[\alpha]) = \langle\alpha, \mathbf{\Gamma}\alpha\rangle$.

Remind that $\Upsilon : [0, \epsilon_0] \times \mathcal{H} \to \mathbb{R}_+$ is given by $\Upsilon(\epsilon, \alpha) = \sigma_{\mathrm{R}}^2(P_{x,\epsilon}[\alpha])$. Then, we have $D_{2,(\epsilon,\phi_{1,\epsilon})}\Upsilon : \mathcal{H} \to \mathbb{R}$ defined through (A.3), is such that $(D_{2,(\epsilon,\phi_{1,\epsilon})}\Upsilon)(h) = \langle\nabla\Lambda_{\phi_{1,\epsilon}}, \phi_{1,\epsilon}\rangle\langle\phi_{1,\epsilon}, h\rangle$. Note that $(D_{2,(\epsilon,\phi_{1,\epsilon})}\Upsilon)(\phi_{1,\epsilon}) = \langle\nabla\Lambda_{\phi_{1,\epsilon}}, \phi_{1,\epsilon}\rangle \times \langle\phi_{1,\epsilon}, \phi_{1,\epsilon}\rangle$ so that $(D_{2,(\epsilon,\phi_{1,\epsilon})}\Upsilon)(h) = (D_{2,(\epsilon,\phi_{1,\epsilon})}\Upsilon)(\phi_{1,\epsilon})\langle\phi_{1,\epsilon}, h\rangle$.

To avoid burden notation, let $u(\epsilon, \alpha)(h) = (D_{2,(\epsilon,\alpha)}\Upsilon)(\alpha)\langle\alpha, h\rangle$. Define the map $\mathcal{L} : [0, \epsilon_0] \times \mathcal{H} \to B(\mathcal{H}, \mathbb{R})$ as $\mathcal{L}(\epsilon, \alpha) = D_{2,(\epsilon,\alpha)}\Upsilon - (D_{2,(\epsilon,\alpha)}\Upsilon)(\alpha)\langle\alpha, \cdot\rangle = D_{2,(\epsilon,\alpha)}\Upsilon - u(\epsilon, \alpha)$. Note that $\mathcal{L}(\epsilon, \phi_{1,\epsilon}) = 0$, in particular we have that $\mathcal{L}(0, \phi_1) = 0$. In order to apply the Implicit Function Theorem, we need to show that $D_{2,(0,\phi_1)}\mathcal{L} : \mathcal{H} \to B(B(\mathcal{H}, \mathbb{R}), \mathbb{R})$ is an isomorphism, since Υ is two times differentiable.

Note that $D_{2,(0,\phi_1)}\mathcal{L} = D^2_{22,(0,\phi_1)}\Upsilon - D_{2,(0,\phi_1)}u$ and $\Upsilon(0, \alpha) = \langle\alpha, \mathbf{\Gamma}\alpha\rangle$, which implies that $D_{2,(0,\alpha)}\Upsilon = 2\langle\mathbf{\Gamma}\alpha, \cdot\rangle$. Hence, $\left(D^2_{22,(0,\phi_1)}\Upsilon(\alpha)\right)(\beta) = 2\langle\mathbf{\Gamma}\alpha, \beta\rangle$ and $u(0, \alpha) = (D_{2,(0,\alpha)}\Upsilon)(\alpha)\langle\alpha, \cdot\rangle = 2\langle\alpha, \mathbf{\Gamma}\alpha\rangle\langle\alpha, \cdot\rangle$. Therefore, $u(0, \alpha) = 2\Upsilon(0, \alpha)\langle\alpha, \cdot\rangle = \langle 2\Upsilon(0, \alpha)\alpha, \cdot\rangle = \langle\widetilde{u}(\alpha), \cdot\rangle$, entailing that $(D_{2,(0,\alpha)}u)(h) = 4\langle\mathbf{\Gamma}\alpha, h\rangle\alpha + 2\langle\alpha, \mathbf{\Gamma}\alpha\rangle h$. Hence, we get that

$(D_{2,(0,\phi_1)}u)(h) = 4\lambda_1\langle\phi_1, h\rangle\phi_1 + 2\lambda_1 h$, so that,

$$D_{2,(0,\phi_1)}\mathcal{L}(h) = 2\boldsymbol{\Gamma}h - 4\lambda_1\langle\phi_1, h\rangle\phi_1 - 2\lambda_1 h$$
$$= 2\left(\sum_{j\geq 1}\lambda_j\phi_j\otimes\phi_j - 2\lambda_1\phi_1\otimes\phi_1 - \lambda_1 I_{\mathcal{H}}\right)h.$$

Using that the identity operator $I_{\mathcal{H}}$ equals to $\sum_{j\geq 1}\phi_j\otimes\phi_j$, we obtain that

$$T = \frac{1}{2}D_{2,(0,\phi_1)}\mathcal{L} = -2\lambda_1(\phi_1\otimes\phi_1) - \sum_{j\geq 2}(\lambda_1-\lambda_j)(\phi_j\otimes\phi_j).$$

The fact that $\lambda_1 > 0$ and $\lambda_1 - \lambda_j > 0$ for $j \geq 2$ entails that T is a monomorphism. It remains to see that T is an epimorphism, that is, we have to prove that for any $y \in \mathcal{H}$ there exists $x \in \mathcal{H}$ such that $Tx = y$. We begin by proving that $\sum_{j\geq 2}(\lambda_1-\lambda_j)^{-2}\langle y, \phi_j\rangle^2 < \infty$. Indeed, $\lambda_1 - \lambda_j \geq \lambda_1 - \lambda_2 > 0$ for $j \geq 2$, then

$$\sum_{j\geq 2}\frac{1}{(\lambda_1-\lambda_j)^2}\langle y, \phi_j\rangle^2 \leq \frac{1}{(\lambda_1-\lambda_2)^2}\|y\|^2 < \infty.$$

Define $x = -\{\langle y, \phi_1\rangle/(2\lambda_1)\}\phi_1 - \sum_{j\geq 2}(\lambda_1-\lambda_j)^{-1}\langle y, \phi_j\rangle\phi_j$. It is easy to see that $x \in \mathcal{H}$ (that is, it has finite norm) and that $Tx = y$. Therefore, T is an isomorphism and so we can apply the Implicit Function Theorem to the equation $\mathcal{L}(0, \phi_1) = 0$ to ensure that $\partial\phi_{1,\epsilon}/\partial\epsilon|_{\epsilon=0}$ exists as desired.

We will now show that the influence function of the functional $\phi_{\mathrm{R},k}$ exists by using an induction argument. Assume that $\phi_{1,\epsilon}, \ldots, \phi_{k-1,\epsilon}$ are differentiable with respect to ϵ at $\epsilon = 0$. We want to see that $\phi_{k,\epsilon}$ is also differentiable with respect to ϵ at $\epsilon = 0$. Using the orthogonality of the directions $\phi_{j,\epsilon}$, we have

$$D_{2,(\epsilon,\phi_{k,\epsilon})}\Upsilon(h) = \left(D_{2,(\epsilon,\phi_{k,\epsilon})}\Upsilon\right)(\phi_{k,\epsilon})\langle\phi_{k,\epsilon}, h\rangle + \sum_{j=1}^{k-1}\left(D_{2,(\epsilon,\phi_{k,\epsilon})}\Upsilon\right)(\phi_{j,\epsilon})\langle\phi_{j,\epsilon}, h\rangle.$$

Let $\mathcal{L} : [0, \epsilon_0] \times \mathcal{H} \to \mathcal{B}(\mathcal{H}, \mathbb{R})$ defined as $\mathcal{L}(\epsilon, \alpha) = D_{2,(\epsilon,\alpha)}\Upsilon - u_0(\epsilon, \alpha) - \sum_{j=1}^{k-1}u_j(\epsilon, \alpha)$, where

$$u_0(\epsilon, \alpha) = \left(D_{2,(\epsilon,\alpha)}\Upsilon\right)(\alpha)\langle\alpha, \cdot\rangle$$
$$u_j(\epsilon, \alpha) = \left(D_{2,(\epsilon,\alpha)}\Upsilon\right)(\phi_{j,\epsilon})\langle\phi_{j,\epsilon}, \cdot\rangle, \quad \text{for } 1 \leq j \leq k-1.$$

Then, $\mathcal{L}(\epsilon, \phi_{k,\epsilon}) = 0$, in particular, $\mathcal{L}(0, \phi_k) = 0$ and $\mathcal{L}$ is differentiable. Again, to apply the Implicit Function Theorem, we have to show that $D_{2,(0,\phi_k)}\mathcal{L} : \mathcal{H} \to \mathcal{B}(\mathcal{B}(\mathcal{H}, \mathbb{R}), \mathbb{R})$ is an isomorphism.

Recall that $D_{2,(0,\phi_k)}\mathcal{L} = D^2_{22,(0,\phi_k)}\Upsilon - D_{2,(0,\phi_k)}u_0 - \sum_{j=1}^{k-1}D_{2,(0,\phi_k)}u_j$, where again, $D^2_{22,(0,\phi_k)}\Upsilon(h) = 2\langle\boldsymbol{\Gamma}h, \cdot\rangle$, $D_{2,(0,\phi_k)}u_0(h) = 4\lambda_k\langle\phi_k, h\rangle\langle\phi_k, \cdot\rangle + 2\lambda_k\langle h, \cdot\rangle$ and $D_{2,(0,\phi_k)}u_0 = \left(D^2_{22,(0,\phi_k)}\Upsilon\right)(\phi_j)\langle\phi_j, \cdot\rangle = 2\boldsymbol{\Gamma}\phi_j\langle\phi_j, \cdot\rangle = 2\lambda_j\phi_j\langle\phi_j, \cdot\rangle$. Therefore, $(D_{2,(0,\phi_k)}\mathcal{L})(h) = \langle 2\boldsymbol{\Gamma}h - (4\lambda_k\langle\phi_k, h\rangle\phi_k + 2\lambda_k h) - 2\sum_{j=1}^{k-1}\lambda_j\phi_j\langle\phi_j, h\rangle, \cdot\rangle$, that is, $D_{2,(0,\phi_k)}\mathcal{L} = 2\boldsymbol{\Gamma} - 4\lambda_k\phi_k\otimes\phi_k - 2\lambda_k I_{\mathcal{H}} - 2\sum_{j=1}^{k-1}\lambda_j\phi_j\otimes\phi_j$. Define

$$T = \frac{1}{2}D_{2,(0,\phi_k)}\mathcal{L} = \sum_{s\geq 1}\lambda_s\phi_s\otimes\phi_s - 2\lambda_k\phi_k\otimes\phi_k - \lambda_k I_{\mathcal{H}} - \sum_{j=1}^{k-1}\lambda_j\phi_j\otimes\phi_j.$$

Using again that $I_{\mathcal{H}} = \sum_{s\geq 1}\phi_s\otimes\phi_s$, we get that $T = -\sum_{s>k}(\lambda_k-\lambda_s)\phi_s\otimes\phi_s - 2\lambda_k(\phi_k\otimes\phi_k) - \lambda_k\sum_{j=1}^{k-1}\phi_s\otimes\phi_s$. Therefore, arguing as above, we obtain that T is an isomorphism since $\lambda_k - \lambda_s > \lambda_k - \lambda_{k+1} > 0$, which concludes the proof of the existence of the influence function.

We will now derive (5). Recall that $(D_{2,\theta_\epsilon}\Upsilon)\alpha = \Lambda'_{\phi_{k,\epsilon}}\alpha = \sum_{j=1}^{k}\langle\nabla\Lambda_{\phi_{k,\epsilon}}, \phi_{j,\epsilon}\rangle\langle\phi_{j,\epsilon}, \alpha\rangle$. Note that the fact that the map $\Upsilon(\epsilon, \alpha) = \sigma^2_{\mathrm{R}}(P_{x,\epsilon}[\alpha])$ is twice continuously Hadamard differentiable at $(0, \phi_k)$ implies that $\nabla\Upsilon_{\epsilon,\phi_{k,\epsilon}}$ is differentiable with respect to ϵ at $\epsilon = 0$, where $\nabla\Upsilon_{\epsilon,\phi_{k,\epsilon}} \in \mathcal{H}$ is such that $\Upsilon'_{\epsilon,\phi_{k,\epsilon}}(\alpha) = \langle\nabla\Upsilon_{\epsilon,\phi_{k,\epsilon}}, \alpha\rangle$ with $\Upsilon'_{\epsilon,\alpha} = D_{2,(\epsilon,\alpha)}\Upsilon$.

Define now the function $f : [0, \epsilon_0] \to \mathcal{H}$ as $f(\epsilon) = \nabla\Lambda_{\phi_{k,\epsilon}} = \nabla\Upsilon_{\epsilon,\phi_{k,\epsilon}}$. The above arguments entail that f is differentiable with respect to ϵ at $\epsilon = 0$. For the sake of simplicity denote $f'(0) = f'_0 \in \mathcal{H}$ its derivative at 0. Using that $h_j(\epsilon) = \phi_{j,\epsilon}$ and $g_j(\epsilon) = \langle\nabla\Lambda_{\phi_{k,\epsilon}}, \phi_{j,\epsilon}\rangle = \langle f(\epsilon), h_j(\epsilon)\rangle$ are also differentiable with respect to ϵ at $\epsilon = 0$ and (A.2), we get that $f'(0) = \sum_{j=1}^{k}g'_j(0)h_j(0) + g_j(0)h'_j(0)$. Using Remarks 2.2 and 2.3, we have that

$$h'_j(0) = \mathrm{IF}(x; \phi_{\mathrm{R},j}, P)$$
$$g'_j(0) = \langle f'(0), h_j(0)\rangle + \langle f(0), h'_j(0)\rangle = \langle f'(0), \phi_j\rangle + \langle f(0), \mathrm{IF}(x; \phi_{\mathrm{R},j}, P)\rangle.$$

Noting that $g_j(0) = 2\langle \phi_k, \phi_j\rangle$ and $f(0) = \nabla \Upsilon_{0,\phi_k} = 2\lambda_k \phi_k$, since $\Upsilon(0,\alpha) = \sigma^2_{\text{R}}(P[\alpha]) = \langle \alpha, \mathbf{\Gamma}\alpha\rangle$, we get

$$f'(0) = \sum_{j=1}^{k} g_j'(0)h_j(0) + g_j(0)h_j'(0)$$
$$= \sum_{j=1}^{k} \left[\langle f'(0), \phi_j\rangle + 2\lambda_k \langle \phi_k, \text{IF}(x;\phi_{\text{R},j}, P)\rangle\right]\phi_j + 2\lambda_k\langle \phi_k, \phi_j\rangle \text{IF}(x;\phi_{\text{R},j}, P).$$

Since $\langle \phi_k, \phi_j\rangle = 0$ for $j \neq k$, we obtain that $f'(0) = \sum_{j=1}^{k}\left[\langle f'(0), \phi_j\rangle + 2\lambda_k\langle \phi_k, \text{IF}(x;\phi_{\text{R},j}, P)\rangle\right]\phi_j + 2\lambda_k \text{IF}(x;\phi_{\text{R},k}, P)$. Denote by $\mathcal{L}_k$ the linear space spanned by $\phi_1 \ldots \phi_k$ and by π_k the orthogonal projection over $\mathcal{L}_k^{\perp}$, that is, $\pi_k = I_{\mathcal{H}} \sum_{j=1}^{k} \phi_j \otimes \phi_j$. Then,

$$\pi_k f'(0) = 2\lambda_k \text{IF}(x;\phi_{\text{R},k}, P) + \sum_{j=1}^{k} 2\lambda_k \langle \phi_k, \text{IF}(x;\phi_{\text{R},j}, P)\rangle \phi_j. \tag{A.8}$$

In order to compute $f'(0)$, let us consider $\epsilon_1 < \epsilon_0$ and $\mathcal{U}_k$ a neighbourhood of ϕ_k such that $\Upsilon(\epsilon,\alpha)$ is continuously Hadamard differentiable at $(\epsilon,\alpha) \in [0,\epsilon_1]\times\mathcal{U}_k$. Define $\Phi : [0,\epsilon_1]\times \mathcal{U}_k \to \mathcal{H}^\star$ as $\Phi(\epsilon,\alpha) = D_{2,\theta}\Upsilon = \Upsilon'_{\epsilon,\alpha}$ with $\theta = (\epsilon,\alpha)$ and $g : [0,\epsilon_1] \to \mathcal{H}^\star$ as $g(\epsilon) = \Phi(\epsilon, \phi_{k,\epsilon})$. Note that $g(\epsilon)$ can be identified with $\nabla \Upsilon_{\epsilon,\phi_{k,\epsilon}} = f(\epsilon)$. Therefore, using Remark 2.3(d), we get that $g_0' = D_{1,(0,\phi_k)}\Phi + (D_{2,(0,\phi_k)}\Phi)\text{IF}(x;\phi_{\text{R},k}, P)$. Using again Remark 2.2, we get that $D_{2,(0,\phi_k)}\Phi = D^2_{22,(0,\phi_k)}\Upsilon$ so that $(D_{2,(0,\phi_k)}\Phi(\alpha))(\beta) = 2\langle \Gamma\alpha, \beta\rangle$ since $\Upsilon(0,\alpha) = \sigma^2_{\text{R}}(P[\alpha]) = \langle\alpha, \mathbf{\Gamma}\alpha\rangle$. On the other hand, for any $(\epsilon,\alpha) \in \mathbb{R}\times\mathcal{H}$, $\left(D_{1,(0,\phi_k)}\Phi\right)(\epsilon,\alpha) = D^2_{12,(0,\phi_k)}\Upsilon(\epsilon,\alpha)$. Using that Υ is twice continuously differentiable at $(0,\phi_k)$ we obtain easily that $D^2_{12,(0,\phi_k)}\Upsilon(\epsilon,\alpha) = D^2_{21,(0,\phi_k)}\Upsilon(\alpha,\epsilon) = (D_{2,(0,\phi_k)}\xi)(\epsilon,\alpha)$ with

$$\xi(\epsilon,\alpha) = D_{1,(\epsilon,\alpha)}\Upsilon. \tag{A.9}$$

The fact that $\xi(0,\alpha) = D_{1,(0,\alpha)}\Upsilon = \text{IF}(\langle \alpha, x\rangle; \sigma^2_{\text{R}}, P)$ and Lemma A.2.1 entail that

$$D^2_{12,(0,\phi_k)}\Upsilon(\epsilon,\alpha) = \epsilon\left[2\lambda_k \text{ IF}\left(\frac{\langle \phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma^2_{\text{R}}, F_0\right)\langle \phi_k,\alpha\rangle + \lambda_k \text{DIF}\left(\frac{\langle \phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma^2_{\text{R}}, F_0\right)\langle \alpha, \frac{x - \langle x,\phi_k\rangle\phi_k}{\sqrt{\lambda_k}}\rangle\right],$$

which leads us to

$$(g_0'(\epsilon))(\alpha) = \left(D_{1,(0,\phi_k)}\Phi\right)(\epsilon,\alpha) + \epsilon\langle (D_{2,(0,\phi_k)}\Phi)\text{IF}(x;\phi_{\text{R},k}, P), \alpha\rangle = \epsilon\langle \nabla_k, \alpha\rangle + \epsilon\langle 2\,\mathbf{\Gamma}\,\text{IF}(x;\phi_{\text{R},k}, P), \alpha\rangle$$
$$= \epsilon\langle \nabla_k + 2\,\mathbf{\Gamma}\,\text{IF}(x;\phi_{\text{R},k}, P), \alpha\rangle,$$

where

$$\nabla_k = 2\lambda_k \text{ IF}\left(\frac{\langle \phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma^2_{\text{R}}, F_0\right)\phi_k + \lambda_k \text{DIF}\left(\frac{\langle \phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma^2_{\text{R}}, F_0\right)\left(\frac{x - \langle x,\phi_k\rangle\phi_k}{\sqrt{\lambda_k}}\right).$$

Since $(g_0'(\epsilon))(\alpha) = \epsilon\langle f'(0),\alpha\rangle$, we obtain that $f'(0) = 2\,\mathbf{\Gamma}\,\text{IF}(x;\phi_{\text{R},k}, P) + \nabla_k$, which replacing in (A.8) implies that $\pi_k\left(2\,\mathbf{\Gamma}\,\text{IF}(x;\phi_{\text{R},k}, P) + \nabla_k\right) = 2\lambda_k\text{IF}(x;\phi_{\text{R},k}, P) + \sum_{j=1}^{k} 2\lambda_k\langle \phi_k, \text{IF}(x;\phi_{\text{R},j}, P)\rangle\phi_j$. Using that $\pi_k\phi_k = 0$, we obtain that $\pi_k\nabla_k = \sqrt{\lambda_k}\text{DIF}\left(\langle \phi_k, x\rangle/\sqrt{\lambda_k}; \sigma^2_{\text{R}}, F_0\right)\pi_k x$. Thus,

$$\pi_k\mathbf{\Gamma}\,\text{IF}(x;\phi_{\text{R},k}, P) - \lambda_k\text{IF}(x;\phi_{\text{R},k}, P) = -\frac{\sqrt{\lambda_k}}{2}\text{DIF}\left(\frac{\langle \phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma^2_{\text{R}}, F_0\right)\pi_k x + \sum_{j=1}^{k}\lambda_k\langle \phi_k, \text{IF}(x;\phi_{\text{R},j}, P)\rangle\phi_j,$$

so that

$$(\pi_k\mathbf{\Gamma} - \lambda_k I_{\mathcal{H}})\text{IF}(x;\phi_{\text{R},k}, P) = -\frac{\sqrt{\lambda_k}}{2}\text{DIF}\left(\frac{\langle \phi_k, x\rangle}{\sqrt{\lambda_k}}; \sigma^2_{\text{R}}, F_0\right)\pi_k x + \sum_{j=1}^{k}\lambda_k\langle \phi_k, \text{IF}(x;\phi_{\text{R},j}, P)\rangle\phi_j. \tag{A.10}$$

Note that for any $\alpha \in \mathcal{H}$, $\pi_k\mathbf{\Gamma}\alpha = \mathbf{\Gamma}\alpha - \sum_{j=1}^{k}\langle \mathbf{\Gamma}\alpha, \phi_j\rangle\phi_j = \mathbf{\Gamma}\alpha - \sum_{j=1}^{k}\lambda_j\langle\alpha,\phi_j\rangle\phi_j = \sum_{j=k+1}^{\infty}\lambda_j\langle\alpha,\phi_j\rangle\phi_j$, hence

$$\pi_k\mathbf{\Gamma} - \lambda_k I_{\mathcal{H}} = \sum_{j=k+1}^{\infty}(\lambda_j - \lambda_k)\phi_j\otimes\phi_j - \lambda_k\sum_{j=1}^{k}\phi_j\otimes\phi_j.$$

Define $\tau_k : \mathcal{H} \to \mathcal{H}$ as

$$\tau_k = \sum_{j\geq k+1}\frac{1}{\lambda_j - \lambda_k}\phi_j\otimes\phi_j - \frac{1}{\lambda_k}\sum_{j=1}^{k}\phi_j\otimes\phi_j = \widetilde{\tau}_k - \frac{1}{\lambda_k}\sum_{j=1}^{k}\phi_j\otimes\phi_j.$$

We have to check that the map is well defined, i.e., that $\|\tau_k(\alpha)\| < \infty$ for any $\alpha \in \mathcal{H}$. Clearly since $\sum_{j=1}^k \phi_j \otimes \phi_j$ has a finite range, we only need to show that $\|\widetilde{\tau}_k(\alpha)\| < \infty$. First note that, since $k \le q$, $\lambda_k > \lambda_{k+1}$ so that, $\lambda_k - \lambda_j \ge \lambda_k - \lambda_{k+1} = M_k > 0$, for $j \ge k+1$, thus, $1/(\lambda_k - \lambda_j) \le 1/M_k$. Then, for any $N > k+1$ we have that

$$\left\| \sum_{j=k+1}^N \frac{1}{\lambda_k - \lambda_j} \langle \phi_j, \alpha \rangle \phi_j \right\|^2 = \sum_{j=k+1}^N \frac{1}{(\lambda_k - \lambda_j)^2} \langle \phi_j, \alpha \rangle^2 \le \frac{1}{M_k^2} \sum_{j=k+1}^N \langle \phi_j, \alpha \rangle^2 \le \frac{1}{M_k^2} \|\alpha\|^2.$$

Hence, $\widetilde{\tau}_k(\alpha) \in \mathcal{H}$. It remains to show that $\tau_k \circ (\pi_k \boldsymbol{\Gamma} - \lambda_k I_{\mathcal{H}})(\alpha) = \alpha$. Effectively,

$$\begin{aligned} \tau_k \circ (\pi_k \boldsymbol{\Gamma} - \lambda_k I_{\mathcal{H}})(\alpha) &= \tau_k \left(\sum_{j=k+1}^\infty (\lambda_j - \lambda_k) \langle \phi_j, \alpha \rangle \phi_j - \lambda_k \sum_{j=1}^k \langle \phi_j, \alpha \rangle \phi_j \right) \\ &= \sum_{j=k+1}^\infty (\lambda_j - \lambda_k) \langle \phi_j, \alpha \rangle \tau_k \phi_j - \lambda_k \sum_{j=1}^k \langle \phi_j, \alpha \rangle \tau_k \phi_j. \end{aligned}$$

Noticing that $\tau_k \phi_j = \phi_j / (\lambda_j - \lambda_k)$ for $j \ge k+1$ and $\tau_k \phi_j = -\phi_j / \lambda_k$ when $j \le k$, we get that

$$\tau_k \circ (\pi_k \boldsymbol{\Gamma} - \lambda_k I_{\mathcal{H}})(\alpha) = \sum_{j=k+1}^\infty (\lambda_j - \lambda_k) \langle \phi_j, \alpha \rangle \frac{1}{\lambda_j - \lambda_k} \phi_j - \lambda_k \sum_{j=1}^k \langle \phi_j, \alpha \rangle \left(-\frac{1}{\lambda_k} \right) \phi_j = \sum_{j \ge 1} \langle \phi_j, \alpha \rangle = \alpha.$$

Hence, applying τ_k in both sides of (A.10), we get

$$\mathrm{IF}(x; \phi_{\mathrm{R},k}, P) = \tau_k \left(-\frac{\sqrt{\lambda_k}}{2} \mathrm{DIF}\left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}^2, F_0 \right) \pi_k x + \sum_{j=1}^k \lambda_k \langle \phi_k, \mathrm{IF}(x; \phi_{\mathrm{R},j}, P) \rangle \phi_j \right). \tag{A.11}$$

Using that $1 = \|\phi_{1,\epsilon}\|^2 = \langle \phi_{k,\epsilon}, \phi_{k,\epsilon} \rangle$ and the chain rule, we get that $\langle \mathrm{IF}(x; \phi_{\mathrm{R},k}, P), \phi_k \rangle = 0$. Therefore, using $\tau_k \phi_j = -\phi_j / \lambda_k$ when $j \le k$ and that $\tau_k \circ \pi_k = \widetilde{\pi}_k$, (A.11) can be re-written as

$$\begin{aligned} \mathrm{IF}(x; \phi_{\mathrm{R},k}, P) &= \tau_k \left(-\frac{\sqrt{\lambda_k}}{2} \mathrm{DIF}\left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}^2, F_0 \right) \pi_k x + \sum_{j=1}^{k-1} \lambda_k \langle \phi_k, \mathrm{IF}(x; \phi_{\mathrm{R},j}, P) \rangle \phi_j \right) \\ &= -\frac{\sqrt{\lambda_k}}{2} \mathrm{DIF}\left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}^2, F_0 \right) \widetilde{\tau}_k x - \sum_{j=1}^{k-1} \langle \phi_k, \mathrm{IF}(x; \phi_{\mathrm{R},j}, P) \rangle \phi_j \\ &= \frac{\sqrt{\lambda_k}}{2} \mathrm{DIF}\left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}^2, F_0 \right) \sum_{j=k+1}^\infty \frac{\langle \phi_j, x \rangle}{\lambda_k - \lambda_j} \phi_j - \sum_{j=1}^{k-1} \langle \phi_k, \mathrm{IF}(x; \phi_{\mathrm{R},j}, P) \rangle \phi_j. \end{aligned} \tag{A.12}$$

Therefore, using that $\mathrm{IF}(x; \phi_{\mathrm{R},k}, P) = \sum_{j \ge 1} \langle \mathrm{IF}(x; \phi_{\mathrm{R},k}, P), \phi_j \rangle \phi_j$, $\langle \mathrm{IF}(x; \phi_{\mathrm{R},k}, P), \phi_k \rangle = 0$ and that (A.12) entails that for $j \ge k+1$

$$\langle \mathrm{IF}(x; \phi_{\mathrm{R},k}, P), \phi_j \rangle = \frac{\sqrt{\lambda_k}}{2} \mathrm{DIF}\left(\frac{\langle \phi_k, x \rangle}{\sqrt{\lambda_k}}; \sigma_{\mathrm{R}}^2, F_0 \right) \frac{\langle \phi_j, x \rangle}{\lambda_k - \lambda_j}, \tag{A.13}$$

while for $j < k$ we get

$$\begin{aligned} \langle \mathrm{IF}(x; \phi_{\mathrm{R},k}, P), \phi_j \rangle &= -\langle \phi_k, \mathrm{IF}(x; \phi_{\mathrm{R},j}, P) \rangle = -\frac{\sqrt{\lambda_j}}{2} \mathrm{DIF}\left(\frac{\langle \phi_j, x \rangle}{\sqrt{\lambda_j}}; \sigma_{\mathrm{R}}^2, F_0 \right) \frac{\langle \phi_k, x \rangle}{\lambda_j - \lambda_k} \\ &= \frac{\sqrt{\lambda_j}}{2} \mathrm{DIF}\left(\frac{\langle \phi_j, x \rangle}{\sqrt{\lambda_j}}; \sigma_{\mathrm{R}}^2, F_0 \right) \frac{\langle \phi_k, x \rangle}{\lambda_k - \lambda_j}, \end{aligned}$$

where we have used (A.13). The proof of (5) is now concluded from the fact that $\mathrm{DIF}(a; \sigma_{\mathrm{R}}^2, F_0) = 2\, \mathrm{DIF}(a; \sigma_{\mathrm{R}}, F_0)$.

It remains to show (6). Denote $\lambda_{k,\epsilon} = \lambda_{\mathrm{R},k}(P_{x,\epsilon})$. Then, $h(\epsilon) = \lambda_{k,\epsilon} = \Upsilon(\epsilon, \phi_{k,\epsilon})$, so that the chain rule entails that $h'(0) = \mathrm{IF}(x; \lambda_{\mathrm{R},k}, P) = D_{1,(0,\phi_k)} \Upsilon + (D_{2,(0,\phi_k)} \Upsilon) \mathrm{IF}(x; \phi_{\mathrm{R},k}, P)$. Recall that $\Lambda'_{\phi_{k,\epsilon}} = D_{2,(\epsilon,\phi_{k,\epsilon})} \Upsilon$ and $\Lambda'_{\phi_{k,\epsilon}}(\alpha) = \langle \nabla \Lambda_{\phi_{k,\epsilon}}, \alpha \rangle$ where $f(\epsilon) = \nabla \Lambda_{\phi_{k,\epsilon}}$ is such that $f(0) = \nabla \Lambda_{\phi_k} = 2\lambda_k \phi_k$. Thus, the fact that $\langle \mathrm{IF}(x; \phi_{\mathrm{R},k}, P), \phi_k \rangle = 0$ implies that $(D_{2,(0,\phi_k)} \Upsilon) \mathrm{IF}(x; \phi_{\mathrm{R},k}, P) = 2\lambda_k \langle \mathrm{IF}(x; \phi_{\mathrm{R},k}, P), \phi_k \rangle = 0$.

On the other hand, (A.9) entails that $D_{1,(0,\phi_k)} \Upsilon = \xi(0, \phi_k)$. We have already shown that $\xi(0, \alpha) = \mathrm{IF}(\langle \alpha, x \rangle; \sigma_{\mathrm{R}}^2, P)$ hence, $\mathrm{IF}(x; \lambda_{\mathrm{R},k}, P) = \mathrm{IF}(\langle x, \phi_k \rangle; \sigma_{\mathrm{R}}^2, P)$. The proof is now concluded using Lemma A.2.1 and the fact that $\mathrm{IF}(a; \sigma_{\mathrm{R}}^2, F_0) = 2\, \mathrm{IF}(a; \sigma_{\mathrm{R}}, F_0)$. □

A.3. Proof of Theorem 3.2

In the proofs of this section, we will use the fact that if $\|\alpha\|_\rho \to 0$, then $\|\alpha\| \to 0$ and that if $\mathcal{K}$ is compact in the topology induced by the norm $\|\cdot\|_\rho$, it is compact in the topology induced by $\|\cdot\|$. These facts entail that if an application $\Lambda : \mathcal{H} \to \mathbb{R}$ is Hadamard differentiable with respect to $\|\cdot\|$, it is Hadamard differentiable with respect to $\|\cdot\|_\rho$.

First, we will re-phrase Lemma A.2.1 to Lemma A.3.1 which does not require the Fisher consistency of $\phi_{\text{R},\text{S},k}$ and that uses another type of differentiability. Lemma A.3.1 will be helpful to derive Theorem 3.2.

Lemma A.3.1. *Let P be an elliptical probability measure $P = \mathcal{E}(\mu, \mathbf{\Gamma})$, where $\mu = 0$ and $\mathbf{\Gamma}$ is a self-adjoint, positive semidefinite and compact operator with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$ and let ϕ_j be the eigenfunction associated to λ_j. Denote $\widetilde{\lambda}_j = \sigma_{\text{R}}(\phi_{\text{R},\text{S},j})$. Assume that $\widetilde{\lambda}_1 > \widetilde{\lambda}_2 > \cdots > \widetilde{\lambda}_q > \widetilde{\lambda}_{q+1}$.*

Let F_0 be the univariate measure defined in Lemma 2.1 and assume that $\sigma_{\text{R}}(F_0) = 1$, so that $\widetilde{\lambda}_j = \langle \phi_{\text{R},\text{S},j}, \mathbf{\Gamma}\phi_{\text{R},\text{S},j}\rangle$. Moreover, assume that the map $S : [0, 1] \times \mathbb{R} \to \mathbb{R}$ defined by $S(\epsilon, y) = \sigma_{\text{R}}((1-\epsilon)F_0 + \epsilon\delta_y)$ is twice continuously differentiable at any $(0, y)$. In particular, $\text{IF}(y; \sigma_{\text{R}}, F_0)$ is differentiable with respect to y and its derivative will be denoted by $\text{DIF}(y; \sigma_{\text{R}}, F_0)$.

Then, for any $k \leq q$, there exists a neighbourhood $\mathcal{U}_k$ of $\phi_{\text{R},k}$ such that for any $\alpha \in \mathcal{U}_k$,

$$\text{IF}(\langle \alpha, x\rangle; \sigma^2_{\text{R}}, P[\alpha]) = \langle \alpha, \mathbf{\Gamma}\alpha\rangle \text{IF}\left(\frac{\langle \alpha, x\rangle}{\sqrt{\langle \alpha, \mathbf{\Gamma}\alpha\rangle}}; \sigma^2_{\text{R}}, F_0\right). \tag{A.14}$$

Moreover, if $\Lambda : \mathcal{H} \to \mathbb{R}$ stands for the map $\Lambda(\alpha) = \text{IF}(\langle \alpha, x\rangle; \sigma^2_{\text{R}}, P[\alpha])$, then Λ is Hadamard differentiable with respect to the norm $\|\cdot\|_\rho$ at $\phi_{\text{R},\text{S},k}$ and

$$\nabla_k = \nabla\Lambda_{\phi_{\text{R},\text{S},k}} = 2\,\text{IF}\left(\frac{\langle \phi_{\text{R},\text{S},k}, x\rangle}{\sqrt{\widetilde{\lambda}_k}}; \sigma^2_{\text{R}}, F_0\right)\mathbf{\Gamma}\phi_{\text{R},\text{S},k} + \widetilde{\lambda}_k \text{DIF}\left(\frac{\langle \phi_{\text{R},\text{S},k}, x\rangle}{\sqrt{\widetilde{\lambda}_k}}; \sigma^2_{\text{R}}, F_0\right)\left(\frac{x - \langle x, \phi_{\text{R},\text{S},k}\rangle \mathbf{\Gamma}\phi_{\text{R},\text{S},k}/\widetilde{\lambda}_k}{\sqrt{\widetilde{\lambda}_k}}\right). \tag{A.15}$$

Proof. Using that $\widetilde{\lambda}_k > 0$, we get there exists a neighbourhood $\mathcal{U}_k$ of ϕ_k such that for any $\alpha \in \mathcal{U}_k$, $\alpha \notin \ker(\mathbf{\Gamma})$, hence Lemma 2.1 entails that $Z_\alpha = \langle \alpha, X\rangle/\sqrt{\langle \alpha, \mathbf{\Gamma}\alpha\rangle} \sim F_0$. Let us recall that

$$\text{IF}(\langle \alpha, x\rangle; \sigma^2_{\text{R}}, P[\alpha]) = \lim_{\epsilon\to 0} \frac{\sigma^2_{\text{R}}((1-\epsilon)P[\alpha] + \epsilon\delta_{\langle x,\alpha\rangle}) - \sigma^2_{\text{R}}(P[\alpha])}{\epsilon}. \tag{A.16}$$

Using that σ_{R} is a scale functional and the fact that $\langle \alpha, X\rangle \sim P[\alpha]$, we get that $\sigma_{\text{R}}(P[\alpha]) = \sqrt{\langle \alpha, \mathbf{\Gamma}\alpha\rangle}$. As in the proof of Lemma A.2.1, we have that $\sigma_{\text{R}}((1-\epsilon)P[\alpha] + \epsilon\delta_{\langle x,\alpha\rangle}) = \sqrt{\langle \alpha, \mathbf{\Gamma}\alpha\rangle}\,\sigma_{\text{R}}((1-\epsilon)F_0 + \epsilon\delta_{z_\alpha})$, which using (A.16) entails (A.14).

It remains to show (A.15). Define the maps $\Upsilon : \mathcal{U}_k \to \mathbb{R}$ and $\Psi : \mathcal{U}_k \to \mathbb{R}$ as $\Upsilon(\alpha) = \langle \alpha, \mathbf{\Gamma}\alpha\rangle$ and $\Psi(\alpha) = \text{IF}\left(\langle \alpha, x\rangle/\sqrt{\langle \alpha, \mathbf{\Gamma}\alpha\rangle}; \sigma^2_{\text{R}}, F_0\right)$. Then, $\Lambda(\alpha) = \text{IF}(\langle \alpha, x\rangle; \sigma^2_{\text{R}}, P[\alpha]) = \Upsilon(\alpha)\Psi(\alpha)$. Note that Υ is Hadamard differentiable with respect to the $\|\cdot\|_\rho$ norm with $\nabla\Upsilon_\alpha = 2\mathbf{\Gamma}\alpha$. Hence, if we show that Ψ is Hadamard differentiable at $\phi_{\text{R},\text{S},k}$, we will get that Λ is also Hadamard differentiable $\phi_{\text{R},\text{S},k}$ and

$$\nabla\Lambda_{\phi_{\text{R},\text{S},k}} = \Upsilon(\phi_{\text{R},\text{S},k})\nabla\Psi_{\phi_{\text{R},\text{S},k}} + \Psi(\phi_{\text{R},\text{S},k})\nabla\Upsilon_{\phi_{\text{R},\text{S},k}}. \tag{A.17}$$

Note that $\Upsilon(\phi_{\text{R},\text{S},k}) = \langle \phi_{\text{R},\text{S},k}, \mathbf{\Gamma}\phi_{\text{R},\text{S},k}\rangle = \widetilde{\lambda}_k$, $\nabla\Upsilon_{\phi_{\text{R},\text{S},k}} = 2\mathbf{\Gamma}\phi_{\text{R},\text{S},k}$ and $\Psi(\phi_{\text{R},\text{S},k}) = \text{IF}\left(\langle \phi_{\text{R},\text{S},k}, x\rangle/\sqrt{\widetilde{\lambda}_k}; \sigma^2_{\text{R}}, F_0\right)$, which together with (A.17) entail that $\nabla\Lambda_{\phi_{\text{R},\text{S},k}} = \widetilde{\lambda}_k\nabla\Psi_{\phi_{\text{R},\text{S},k}} + 2\,\text{IF}\left(\langle \phi_{\text{R},\text{S},k}, x\rangle/\sqrt{\widetilde{\lambda}_k}; \sigma^2_{\text{R}}, F_0\right)\mathbf{\Gamma}\phi_{\text{R},\text{S},k}$. Thus, to conclude the proof, it remains to show that Ψ is Hadamard differentiable at $\phi_{\text{R},\text{S},k}$ with respect to $\|\cdot\|$ and

$$\nabla\Psi_{\phi_{\text{R},\text{S},k}} = \text{DIF}\left(\frac{\langle \phi_{\text{R},\text{S},k}, x\rangle}{\sqrt{\widetilde{\lambda}_k}}; \sigma^2_{\text{R}}, F_0\right)\left(\frac{x - \langle x, \phi_{\text{R},\text{S},k}\rangle \frac{\mathbf{\Gamma}\phi_{\text{R},\text{S},k}}{\widetilde{\lambda}_k}}{\sqrt{\widetilde{\lambda}_k}}\right).$$

Define now $f : \mathbb{R} \to \mathbb{R}$ as $f(y) = \text{IF}(y; \sigma_{\text{R}}, F_0)$ and $\Phi : \mathcal{U}_k \to \mathbb{R}$ as $\Phi(\alpha) = \langle x, \alpha\rangle/\sqrt{\langle \alpha, \mathbf{\Gamma}\alpha\rangle} = \langle x, \alpha\rangle/\sqrt{\Upsilon(\alpha)}$. Then, the Hadamard differentiability of Ψ at $\phi_{\text{R},\text{S},k}$ follows easily from the chain rule and the fact that f is differentiable, Φ is Hadamard differentiable at $\phi_{\text{R},\text{S},k}$ with respect to $\|\cdot\|_\rho$, since $\Upsilon(\phi_{\text{R},\text{S},k}) = \widetilde{\lambda}_k \neq 0$, and $\Psi = f \circ \Phi$. Moreover, the chain rule entails that $\Psi'_{\phi_{\text{R},\text{S},k}} = f'_{\Phi(\phi_{\text{R},\text{S},k})} \circ \Phi'_{\phi_{\text{R},\text{S},k}}$, which together with the fact that $f'_{\Phi(\phi_{\text{R},\text{S},k})}(v) = f'(\Phi(\phi_{\text{R},\text{S},k}))v$ and $\Phi(\phi_{\text{R},\text{S},k}) = \langle \phi_{\text{R},\text{S},k}, x\rangle/\sqrt{\widetilde{\lambda}_k}$, $f'(y) = \text{DIF}(y; \sigma_{\text{R}}, F_0)$ imply that

$$\Psi'_{\phi_{\text{R},\text{S},k}}(\alpha) = f'_{\Phi(\phi_{\text{R},\text{S},k})}\left(\Phi'_{\phi_{\text{R},\text{S},k}}(\alpha)\right) = f'(\Phi(\phi_{\text{R},\text{S},k}))\Phi'_{\phi_{\text{R},\text{S},k}}(\alpha) = \text{DIF}\left(\frac{\langle \phi_{\text{R},\text{S},k}, x\rangle}{\sqrt{\widetilde{\lambda}_k}}; \sigma_{\text{R}}, F_0\right)\Phi'_{\phi_{\text{R},\text{S},k}}(\alpha).$$

Using again Remark 2.2, we get that

$$\Phi'_{\phi_{R,S,k}}(\alpha) = \frac{\langle x, \alpha\rangle \sqrt{\widetilde{\lambda}_k} - \frac{\langle \alpha, \boldsymbol{\Gamma}\phi_{R,S,k}\rangle}{\sqrt{\widetilde{\lambda}_k}} \langle x, \phi_{R,S,k}\rangle}{\widetilde{\lambda}_k} = \frac{\left\langle \alpha, x - \langle x, \phi_{R,k}\rangle \frac{\boldsymbol{\Gamma}\phi_{R,S,k}}{\widetilde{\lambda}_k}\right\rangle}{\sqrt{\widetilde{\lambda}_k}}$$

concluding the proof. □

As in Appendix A.2, for any $\epsilon < \epsilon_0$, $x \in \mathcal{H}$, denote as $\phi_{j,\epsilon} = \phi_{R,S,j}(P_{x,\epsilon})$ and $\lambda_{j,\epsilon} = \lambda_{R,S,j}(P_{x,\epsilon})$. Recall that $\|\phi_{j,\epsilon}\| = 1$ and that $\langle \phi_{j,\epsilon}, \phi_{\ell,\epsilon}\rangle = 0$ for $\ell \neq j$. Moreover, since x is fixed and to avoid burden notation, let $\widetilde{\Upsilon}(\epsilon, \alpha) = \Upsilon(\epsilon, \alpha) - \rho L(\alpha)$.

Let $k \leq q$ and define the restrictions $\boldsymbol{\Psi} : \mathcal{H} \to \mathbb{R}^k$ as $\boldsymbol{\Psi}(\alpha) = (\Psi_0(\alpha), \ldots, \Psi_{k-1}(\alpha))$ with $\Psi_0(\alpha) = \|\alpha\|^2 - 1$ and $\boldsymbol{\Psi}_j(\alpha) = \langle \alpha, \phi_{j,\epsilon}\rangle$, for $1 \leq j \leq k-1$, where we understand that when $k = 1$, $\boldsymbol{\Psi}(\alpha) = \Psi_0(\alpha)$. As Appendix A.2, it is easy to show that $\boldsymbol{\Psi}$ is $\mathcal{C}^1$-Fréchet differentiable with respect to the $\|\cdot\|_\rho$ norm. Moreover, $\boldsymbol{\Psi}'_{\phi_{k,\epsilon}}$ is onto $\mathbb{R}^k$ since $\boldsymbol{\Psi}'_{\phi_{k,\epsilon}} = (\Psi'_{0,\phi_{k,\epsilon}}, \ldots \Psi'_{k-1,\phi_{k,\epsilon}})$ with $\Psi'_{0,\phi_{k,\epsilon}}(\alpha) = \langle 2\phi_{k,\epsilon}, \alpha\rangle$, $\Psi'_{j,\phi_{k,\epsilon}} = \langle \phi_{j,\epsilon}, \alpha\rangle$, for $1 \leq j \leq k-1$. We have that for any fixed ϵ, $\phi_{k,\epsilon}$ maximizes $\widetilde{\Upsilon}(\epsilon, \alpha) = \sigma_R^2(P_{x,\epsilon}[\alpha]) - \rho L(\alpha)$ over $\{\alpha, \boldsymbol{\Psi}(\alpha) = 0\}$, that is, $\widetilde{\Upsilon}_{x,\epsilon} = \widetilde{\Upsilon}(\epsilon, \cdot)$ has a local maximum at $\phi_{k,\epsilon}$ subject to the condition $\boldsymbol{\Psi}(\alpha) = 0$. Besides, $\widetilde{\Upsilon}_{x,\epsilon} : \mathcal{H} \to \mathbb{R}$ is a Hadamard differentiable function with respect to $\|\cdot\|_\rho$, since both Υ and L are Hadamard differentiable with respect to $\|\cdot\|_\rho$.

Define $\Lambda : \mathcal{H} \to \mathbb{R}$ as $\Lambda(\alpha) = \Upsilon(\epsilon, \alpha)$, where we temporarily omit the dependence on ϵ and $\widetilde{\Lambda}(\alpha) = \Lambda(\alpha) - \rho L(\alpha)$. By the Lagrange multipliers theorem for Hadamard differentiable functions, we get that there will exist $\gamma_0, \ldots, \gamma_{k-1} \in \mathbb{R}$ (depending on ϵ) such that $\widetilde{\Lambda}'_{\phi_{k,\epsilon}} = \Lambda'_{\phi_{k,\epsilon}} - \rho L'_{\phi_{k,\epsilon}} = \sum_{j=0}^{k-1} \gamma_j \Psi'_{j,\phi_{k,\epsilon}}$. It is worth noting that $\Lambda'_{\phi_{k,\epsilon}} = D_{2,\theta_\epsilon}\Upsilon$ with $\theta_\epsilon = (\epsilon, \phi_{k,\epsilon}) \in [0, \epsilon_0] \times \mathcal{H}$. Recall that $\Lambda'_{\phi_{k,\epsilon}} : \mathcal{H} \to \mathbb{R}$ is a linear and continuous operator, that is, an element of $\mathcal{H}^\star$. Hence, there exists a gradient, i.e., an element $\nabla\Lambda_{\phi_{k,\epsilon}}$ such that $\Lambda'_{\phi_{k,\epsilon}}(\alpha) = \langle \nabla\Lambda_{\phi_{k,\epsilon}}, \alpha\rangle$.

Define $f : [0, \epsilon_0] \to \mathcal{H}$ as $f(\epsilon) = \nabla\Lambda_{\phi_{k,\epsilon}}$. The derivative of this element at 0 can be computed as in the proof of Theorem 3.1 obtaining $f'_0 = \nabla f_0 = 2\,\boldsymbol{\Gamma}\mathrm{IF}(x; \phi_{R,S,k}, P) + \nabla_k$, with $\phi_{k,0} = \phi_{R,S,k}$ and

$$\nabla_k = 2\ \mathrm{IF}\left(\frac{\langle \phi_{k,0}, x\rangle}{\sqrt{\widetilde{\lambda}_k}}; \sigma_R^2, F_0\right) \boldsymbol{\Gamma}\,\phi_{k,0} + \widetilde{\lambda}_k \mathrm{DIF}\left(\frac{\langle \phi_{k,0}, x\rangle}{\sqrt{\widetilde{\lambda}_k}}; \sigma_R^2, F_0\right) \left(\frac{x - \langle x, \phi_{k,0}\rangle \frac{\boldsymbol{\Gamma}\phi_{k,0}}{\widetilde{\lambda}_k}}{\sqrt{\widetilde{\lambda}_k}}\right).$$

Thus, $f'_0(t) = t(2\,\boldsymbol{\Gamma}\mathrm{IF}(x; \phi_{R,S,k}, P) + \nabla_k)$.

Define $\Phi : [0, \epsilon_0] \to \mathcal{H}^\star$ as $\Phi(\epsilon) = \Lambda'_{\phi_{k,\epsilon}}$, that is, we are considering the derivative instead of the gradient. Using the previous computations for f, we get that $\Phi'_0 : \mathbb{R} \to \mathcal{H}^\star$ equals $\left(\Phi'_0(t)\right)(\alpha) = t\langle 2\,\boldsymbol{\Gamma}\mathrm{IF}(x; \phi_{R,S,k}, P) + \nabla_k, \alpha\rangle$.

Define $\Xi : [0, \epsilon_0] \to \mathcal{H}^\star$ as $\Xi(t) = \rho L'_{\phi_{k,\epsilon}}$. Recall that $L'_{\phi_{k,\epsilon}}(\alpha) = 2\lceil \phi_{k,\epsilon}, \alpha\rceil$. A further derivation of this with respect to ϵ will yield, using the chain rule, that $\Xi'_0(t)(\alpha) = 2\rho t\lceil \mathrm{IF}(x; \phi_{R,S,k}, P), \alpha\rceil$. If we define $\widetilde{\Phi} = \Phi - \rho\,\Xi$, we get that $\widetilde{\Phi}(\epsilon) = \widetilde{\Lambda}'_{\phi_{k,\epsilon}}$, so that

$$\begin{aligned}\widetilde{\Phi}'_0(t)(\alpha) &= \Phi'_0(t)(\alpha) - \Xi'_0(t)(\alpha) = t\langle 2\,\boldsymbol{\Gamma}\mathrm{IF}(x; \phi_{R,S,k}, P) + \nabla_k, \alpha\rangle - 2\rho t\lceil \mathrm{IF}(x; \phi_{R,S,k}, P), \alpha\rceil \\ &= t\left(\langle 2\,\boldsymbol{\Gamma}\mathrm{IF}(x; \phi_{R,S,k}, P) + \nabla_k, \alpha\rangle - 2\rho\lceil \mathrm{IF}(x; \phi_{R,S,k}, P), \alpha\rceil\right).\end{aligned}$$

In particular, we have that

$$\widetilde{\Phi}'_0(1)(\alpha) = \langle 2\ \boldsymbol{\Gamma}\ \mathrm{IF}(x; \phi_{R,S,k}, P) + \nabla_k, \alpha\rangle - 2\rho\lceil \mathrm{IF}(x; \phi_{R,S,k}, P), \alpha\rceil. \tag{A.18}$$

Let us denote by $\xi_j = \widetilde{\Phi}'_0(1)(\phi_{j,0}) \in \mathbb{R}$, using (A.18), we get that for any $j \geq 1$

$$\xi_j = \widetilde{\Phi}'_0(1)(\phi_{j,0}) = \langle 2\ \boldsymbol{\Gamma}\ \mathrm{IF}(x; \phi_{R,S,k}, P) + \nabla_k, \phi_{j,0}\rangle - 2\rho\lceil \mathrm{IF}(x; \phi_{R,S,k}, P), \phi_{j,0}\rceil. \tag{A.19}$$

As in the previous section, will proceed now with an alternative way of computing $\widetilde{\Phi}'_0(1)(\alpha)$. Define $H : \mathbb{R} \to \mathcal{H}^\star$ as $H(\epsilon) = \widetilde{\Lambda}'_{\phi_{k,\epsilon}} - \rho L'_{\phi_{k,\epsilon}} - \sum_{j=0}^{k-1} \gamma_j \Psi'_{j,\phi_{k,\epsilon}}$. We have that $H(\epsilon) = 0$ for all $\epsilon < \epsilon_0$. Using that $\Psi'_{0,\phi_{k,\epsilon}}(\alpha) = \langle 2\phi_{k,\epsilon}, \alpha\rangle$ and $\Psi'_{j,\phi_{k,\epsilon}} = \langle \phi_{j,\epsilon}, \alpha\rangle$, for $1 \leq j \leq k-1$, we get that $H(\epsilon)(\alpha) = \widetilde{\Lambda}'_{\phi_{k,\epsilon}}(\alpha) - 2\gamma_0\langle \alpha, \phi_{k,\epsilon}\rangle + \sum_{j=1}^{k-1} \gamma_j\langle \alpha, \phi_{j,\epsilon}\rangle$. This entails that, $\widetilde{\Lambda}'_{\phi_{k,\epsilon}}(\phi_{k,\epsilon}) = 2\gamma_0$, while $\widetilde{\Lambda}'_{\phi_{k,\epsilon}}(\phi_{j,\epsilon}) = \gamma_j$ for $1 \leq j \leq k-1$. This provides us the following alternative way of writing $\widetilde{\Lambda}'_{\phi_{k,\epsilon}}$ as $\widetilde{\Lambda}'_{\phi_{k,\epsilon}}(\alpha) = \sum_{j=1}^{k} \widetilde{\Lambda}'_{\phi_{k,\epsilon}}(\phi_{j,\epsilon})\langle \alpha, \phi_{j,\epsilon}\rangle$. As in Appendix A.2, we want to link this expression to that for $\widetilde{\Phi}'_0$ given in (A.18). Therefore, we need to differentiate $\widetilde{\Lambda}'_{\phi_{k,\epsilon}}$ with respect to ϵ at $\epsilon = 0$. Write $\widetilde{\Lambda}'_{\phi_{k,\epsilon}}(\alpha) = \sum_{j=1}^{k} H_1(\epsilon, h_j(\epsilon))H_2(\alpha, h_j(\epsilon))$, with $h_j(\epsilon) = \phi_{j,\epsilon}$, $H_1(\epsilon, \alpha) = \widetilde{\Lambda}'_{\phi_{k,\epsilon}}(\alpha)$ and $H_2(\alpha, \beta) = \langle \alpha, \beta\rangle$. Recall that $\widetilde{\Phi}(\epsilon) = \widetilde{\Lambda}'_{\phi_{k,\epsilon}}$. When differentiating with respect to $\epsilon = 0$, note that $(\nabla h_j)_0 = \mathrm{IF}(x; \phi_{j,0}, P)$, $\widetilde{\Phi}'_0(t) = t\widetilde{\Phi}'_0(1)$, $H_2(\alpha, h_j(0)) = \langle \alpha, \phi_{j,0}\rangle$, $H_1(0, h_j(0)) = \widetilde{\Lambda}'_{\phi_{k,0}}(\phi_{j,0}) = \left(\widetilde{\Phi}(0)\right)(\phi_{j,0})$, so that we get easily that

$$\widetilde{\Phi}'_0(1)(\alpha) = \sum_{j=1}^{k} \left\langle \left(\widetilde{\Phi}'_0(1)(\phi_{j,0}) + \left(\widetilde{\Phi}(0)\right)(\mathrm{IF}(x; \phi_{R,S,k}, P))\right)\phi_{j,0} + \widetilde{\Phi}(0)(\phi_{j,0})\mathrm{IF}(x; \phi_{R,S,j}, P), \alpha\right\rangle. \tag{A.20}$$

Using that $\Upsilon(0,\alpha)=\sigma^2_{\text{R}}(P[\alpha])=\langle\alpha,\mathbf{\Gamma}\alpha\rangle$, we obtain that $\Lambda'_{\phi_{k,0}}(\alpha)=D_{2,(0,\phi_{k,0})}\Upsilon(\alpha)=2\langle\alpha,\mathbf{\Gamma}\phi_{k,0}\rangle$, so

$$\widetilde{\Phi}(0)(\alpha)=\widetilde{\Lambda}'_{\phi_{k,0}}(\alpha)=\Lambda'_{\phi_{k,0}}(\alpha)-2\rho\lceil\phi_{k,0},\alpha\rceil=2\langle\mathbf{\Gamma}\phi_{k,0},\alpha\rangle-2\rho\lceil\phi_{k,0},\alpha\rceil,$$

which entails that $\widetilde{\Phi}(0)(\phi_{j,0})=2\langle\mathbf{\Gamma}\phi_{k,0},\phi_{j,0}\rangle-2\rho\lceil\phi_{k,0},\phi_{j,0}\rceil$, for $j\le k$ and $\widetilde{\Phi}(0)(\text{IF}(x;\phi_{\text{R,s},j},P))=2\langle\mathbf{\Gamma}\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},j},P)\rangle-2\rho\lceil\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},j},P)\rceil$. Replacing in (A.20) and recalling that $\xi_j=\widetilde{\Phi}'_0(1)(\phi_{j,0})\in\mathbb{R}$, we get

$$\begin{aligned}\widetilde{\Phi}'_0(1)(\alpha)&=\sum_{\ell=1}^{k}\langle\alpha,\xi_\ell\phi_{\ell,0}+2\langle\mathbf{\Gamma}\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle\phi_{\ell,0}-2\rho\lceil\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rceil\phi_{\ell,0}\\&\quad+2\langle\mathbf{\Gamma}\phi_{k,0},\phi_{\ell,0}\rangle\text{IF}(x;\phi_{\text{R,s},\ell},P)-2\rho\lceil\phi_{k,0},\phi_{\ell,0}\rceil\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle.\end{aligned}\tag{A.21}$$

Using (A.21), the fact that $\langle\phi_{j,0},\phi_{\ell,0}\rangle=0$, for $\ell\neq j$ and the fact that $\|\phi_{j,\epsilon}\|^2=1$ and $\langle\phi_{j,\epsilon},\phi_{\ell,\epsilon}\rangle=0$, for $\ell\neq j$ entail that $\langle\text{IF}(x;\phi_{\text{R,s},j},P),\phi_{j,0}\rangle=0$ and $\langle\text{IF}(x;\phi_{\text{R,s},j},P),\phi_{\ell,0}\rangle=-\langle\text{IF}(x;\phi_{\text{R,s},\ell},P),\phi_{j,0}\rangle$, we get that for $j\ge 1$,

$$\begin{aligned}\xi_j&=\sum_{\ell=1}^{k}\langle\phi_{j,0},\xi_\ell\phi_{\ell,0}+2\langle\mathbf{\Gamma}\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle\phi_{\ell,0}-2\rho\lceil\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rceil\phi_{\ell,0}\\&\quad+2\langle\mathbf{\Gamma}\phi_{k,0},\phi_{\ell,0}\rangle\text{IF}(x;\phi_{\text{R,s},\ell},P)-2\rho\lceil\phi_{k,0},\phi_{\ell,0}\rceil\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle\end{aligned}$$

so that for $j>k$, we get

$$\xi_j=\sum_{\ell=1}^{k}\left\{2\langle\mathbf{\Gamma}\phi_{k,0},\phi_{\ell,0}\rangle-2\rho\lceil\phi_{k,0},\phi_{\ell,0}\rceil\right\}\langle\phi_{j,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle$$

which together with (A.19), implies that for $j>k$,

$$\begin{aligned}2\rho\lceil\text{IF}(x;\phi_{\text{R,s},k},P),\phi_{j,0}\rceil&=\langle 2\,\mathbf{\Gamma}\,\text{IF}(x;\phi_{\text{R,s},k},P)+\nabla_k,\phi_{j,0}\rangle\\&\quad-\sum_{\ell=1}^{k}\left\{2\langle\mathbf{\Gamma}\phi_{k,0},\phi_{\ell,0}\rangle-2\rho\lceil\phi_{k,0},\phi_{\ell,0}\rceil\right\}\langle\phi_{j,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle.\end{aligned}$$

On the other hand, for $1\le j\le k$

$$\begin{aligned}\xi_j&=\xi_j+2\langle\mathbf{\Gamma}\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},j},P)\rangle-2\rho\lceil\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},j},P)\rceil\\&\quad+\sum_{\ell=1,j\neq\ell}^{k}\left(2\langle\mathbf{\Gamma}\phi_{k,0},\phi_{\ell,0}\rangle-2\rho\lceil\phi_{k,0},\phi_{\ell,0}\rceil\right)\langle\phi_{j,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle\end{aligned}$$

which together with $\langle\text{IF}(x;\phi_{\text{R,s},j},P),\phi_{\ell,0}\rangle=-\langle\text{IF}(x;\phi_{\text{R,s},\ell},P),\phi_{j,0}\rangle$ implies that for any $j\le k$

$$\begin{aligned}\upsilon_{k,j}&=2\rho\lceil\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},j},P)\rceil=2\langle\mathbf{\Gamma}\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},j},P)\rangle\\&\quad-\sum_{\ell=1,j\neq\ell}^{k}\left(2\langle\mathbf{\Gamma}\phi_{k,0},\phi_{\ell,0}\rangle-2\rho\lceil\phi_{k,0},\phi_{\ell,0}\rceil\right)\langle\phi_{\ell,0},\text{IF}(x;\phi_{\text{R,s},j},P)\rangle.\end{aligned}$$

Using (A.21) we obtain the equalities

$$\begin{aligned}\widetilde{\Phi}'_0(1)(\alpha)&=\sum_{\ell=1}^{k}\widetilde{\Phi}'_0(1)(\phi_{\ell,0})\langle\alpha,\phi_{\ell,0}\rangle+2\langle\mathbf{\Gamma}\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle\langle\alpha,\phi_{\ell,0}\rangle-2\rho\lceil\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rceil\langle\alpha,\phi_{\ell,0}\rangle\\&\quad+2\langle\mathbf{\Gamma}\phi_{k,0},\phi_{\ell,0}\rangle\langle\alpha,\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle-2\rho\lceil\phi_{k,0},\phi_{\ell,0}\rceil\langle\alpha,\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle\\&=\left\langle\alpha,\sum_{\ell=1}^{k}\widetilde{\Phi}'_0(1)(\phi_{\ell,0})\phi_{\ell,0}\right\rangle+2\sum_{\ell=1}^{k}\langle\mathbf{\Gamma}\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle\langle\alpha,\phi_{\ell,0}\rangle\\&\quad-2\rho\sum_{\ell=1}^{k}\lceil\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rceil\langle\alpha,\phi_{\ell,0}\rangle\\&\quad+\sum_{\ell=1}^{k}2\langle\mathbf{\Gamma}\phi_{k,0},\phi_{\ell,0}\rangle\langle\alpha,\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle-2\sum_{\ell=1}^{k}\rho\lceil\phi_{k,0},\phi_{\ell,0}\rceil\langle\alpha,\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle.\end{aligned}$$

From this last equality and Eq. (A.18), we obtain that

$$\begin{aligned}&\langle 2\,\mathbf{\Gamma}\,\text{IF}(x;\phi_{\text{R,s},k},P)+\nabla_k,\alpha\rangle-2\rho\lceil\text{IF}(x;\phi_{\text{R,s},k},P),\alpha\rceil\\&=\left\langle\alpha,\sum_{\ell=1}^{k}\widetilde{\Phi}'_0(1)(\phi_{\ell,0})\phi_{\ell,0}\right\rangle+2\sum_{\ell=1}^{k}\langle\mathbf{\Gamma}\phi_{k,0},\text{IF}(x;\phi_{\text{R,s},\ell},P)\rangle\langle\alpha,\phi_{\ell,0}\rangle\end{aligned}$$

$$- \sum_{\ell=1}^{k} 2\rho\lceil \phi_{k,0}, \mathrm{IF}(x;\phi_{\mathrm{R,S},\ell},P)\rceil \langle \alpha, \phi_{\ell,0}\rangle + \sum_{\ell=1}^{k} 2\langle \boldsymbol{\Gamma}\phi_{k,0}, \phi_{\ell,0}\rangle \langle \alpha, \mathrm{IF}(x;\phi_{\mathrm{R,S},\ell},P)\rangle$$
$$-2\sum_{\ell=1}^{k} \rho\lceil \phi_{k,0}, \phi_{\ell,0}\rceil \langle \alpha, \mathrm{IF}(x;\phi_{\mathrm{R,S},\ell},P)\rangle. \tag{A.22}$$

Define $A_k = 2\boldsymbol{\Gamma}\,\mathrm{IF}(x;\phi_{\mathrm{R,S},k},P) + \nabla_k$ and $\mathrm{IF}_k = \mathrm{IF}(x,\phi_{\mathrm{R,S},k},P)$. Then, we can rewrite (A.19) as $\xi_\ell = \langle A_k, \phi_\ell\rangle - 2\rho\lceil \mathrm{IF}_k, \phi_\ell\rceil = \widetilde{\Phi}_0'(1)(\phi_{\ell,0})$ and also (A.22) as

$$\langle A_k, \alpha\rangle - 2\rho\lceil \mathrm{IF}_k, \alpha\rceil = \sum_{\ell=1}^{k} \xi_\ell \langle \phi_{\ell,0}, \alpha\rangle + 2\sum_{\ell=1}^{k} \langle \boldsymbol{\Gamma}\phi_{k,0}, \mathrm{IF}_\ell\rangle \langle \alpha, \phi_{\ell,0}\rangle - 2\sum_{\ell=1}^{k} \rho\lceil \phi_{k,0}, \mathrm{IF}_\ell\rceil \langle \alpha, \phi_{\ell,0}\rangle$$
$$+ \sum_{\ell=1}^{k} 2\langle \boldsymbol{\Gamma}\phi_{k,0}, \phi_{\ell,0}\rangle \langle \alpha, \mathrm{IF}_\ell\rangle - 2\sum_{\ell=1}^{k} \rho\lceil \phi_{k,0}, \phi_{\ell,0}\rceil \langle \alpha, \mathrm{IF}_\ell\rangle,$$

which leads to

$$\langle A_k, \alpha\rangle - 2\rho\lceil \mathrm{IF}_k, \alpha\rceil - \sum_{\ell=1}^{k} \xi_\ell \langle \phi_{\ell,0}, \alpha\rangle = 2\widetilde{\lambda}_k \langle \mathrm{IF}_k, \alpha\rangle + 2\sum_{l<k} \langle \boldsymbol{\Gamma}\phi_{k,0}, \phi_{\ell,0}\rangle \langle \mathrm{IF}_\ell, \alpha\rangle + 2\sum_{\ell=1}^{k} \langle \boldsymbol{\Gamma}\phi_{k,0}, \mathrm{IF}_\ell\rangle \langle \phi_{\ell,0}, \alpha\rangle$$
$$-2\rho\sum_{\ell=1}^{k} \lceil \mathrm{IF}_\ell, \phi_{k,0}\rceil \langle \phi_{\ell,0}, \alpha\rangle - 2\rho\sum_{\ell=1}^{k} \lceil \phi_{k,0}, \phi_{\ell,0}\rceil \langle \mathrm{IF}_\ell, \alpha\rangle. \tag{A.23}$$

On the other hand, we have that

$$\sum_{\ell=1}^{k} \xi_\ell \langle \phi_{\ell,0}, \alpha\rangle = \sum_{\ell=1}^{k} \langle A_k, \phi_{\ell,0}\rangle \langle \phi_{\ell,0}, \alpha\rangle - 2\rho\sum_{\ell=1}^{k} \lceil \phi_{k,0}, \phi_{\ell,0}\rceil \langle \mathrm{IF}_\ell, \alpha\rangle$$
$$= \sum_{\ell=1}^{k} \langle (\phi_{\ell,0}\otimes\phi_{\ell,0})A_k, \alpha\rangle - 2\rho\sum_{\ell=1}^{k} \lceil \phi_{k,0}, \phi_{\ell,0}\rceil \langle \mathrm{IF}_\ell, \alpha\rangle. \tag{A.24}$$

Define $\pi_k : \mathcal{H} \to \mathcal{H}$ as the projection operator over the linear space orthogonal to that spanned by $\{\phi_{1,0}, \dots, \phi_{k,0}\}$, that is, $\pi = I_{\mathcal{H}} - \sum_{i=1}^{k} (\phi_{i,0}\otimes\phi_{i,0})$. Using (A.24) and (A.23) we get that

$$\langle \pi_k A_k, \alpha\rangle - 2\rho\lceil \mathrm{IF}_k, \alpha\rceil + 2\rho\sum_{\ell=1}^{k} \lceil \mathrm{IF}_k, \phi_{\ell,0}\rceil \langle \phi_{\ell,0}, \alpha\rangle = 2\widetilde{\lambda}_k \langle \mathrm{IF}_k, \alpha\rangle + 2\sum_{\ell<k} \langle \boldsymbol{\Gamma}\phi_{k,0}, \phi_{\ell,0}\rangle \langle \mathrm{IF}_\ell, \alpha\rangle$$
$$+2\sum_{\ell=1}^{k} \langle \boldsymbol{\Gamma}\phi_{k,0}, \mathrm{IF}_\ell\rangle \langle \phi_{\ell,0}, \alpha\rangle - 2\rho\sum_{\ell=1}^{k} \lceil \mathrm{IF}_\ell, \phi_{k,0}\rceil \langle \phi_{\ell,0}, \alpha\rangle - 4\rho\sum_{\ell=1}^{k} \lceil \phi_{k,0}, \phi_{\ell,0}\rceil \langle \mathrm{IF}_\ell, \alpha\rangle,$$

concluding the proof. □

A.4. Proof of (10)

Since that $X \sim \mathcal{E}(0, \boldsymbol{\Gamma})$, without loss of generality, we can assume that $\sigma(\alpha) = \langle \alpha, \boldsymbol{\Gamma}\alpha\rangle$. Then, $\phi_{1,\mathrm{R,S}} = \mathrm{argmax}_{\|\alpha\|=1} \langle \alpha, \boldsymbol{\Gamma}\alpha\rangle - \rho\int_0^1 (\alpha''(t))^2 dt$. Denoting $\alpha_i = \langle \alpha, \phi_i\rangle$ we get that $\langle \alpha, \boldsymbol{\Gamma}\alpha\rangle = \sum_{i\geq 1} \alpha_i\lambda_i$, and $\rho\int_0^1 (\alpha''(t))^2 dt = \rho\sum_{i\geq 1} \alpha_i \int_0^1 (\phi_i''(t))^2 dt = \rho\sum_{i\geq 1} \alpha_i a_i^4 \pi^4$. Thus, we have that

$$\phi_{1,\mathrm{R,S}} = \underset{\|\alpha\|=1}{\mathrm{argmax}} \sum_{i\geq 1} \alpha_i\lambda_i - \rho\sum_{i\geq 1} \alpha_i a_i^4 \pi^4.$$

Since λ_i are decreasing and a_i are non-decreasing, it is easy to see that regardless of the value of ρ, the quantity $\sum_{i\geq 1} \alpha_i\lambda_i - \rho\sum_{i\geq 1} \alpha_i a_i^4\pi^4$ is maximized when $\alpha_1 = 1$ and $\alpha_i = 0$ for $i \geq 2$. Therefore, we obtain that $\phi_{1,\mathrm{R,S}} = \phi_1$. Similarly, we can obtain that $\phi_{k,\mathrm{R,S}} = \phi_k$, so that $\widetilde{\lambda}_k = \lambda_k$.

Using the closability of $\lceil \cdot, \cdot \rceil$, we can rewrite (9) as

$$\langle \pi_k A_k, \alpha\rangle - 2\rho\langle D^2\mathrm{IF}_k, \alpha\rangle + 2\rho\sum_{\ell=1}^{k} \langle D^2\mathrm{IF}_k, \phi_\ell\rangle \langle \phi_\ell, \alpha\rangle = 2\lambda_k \langle \mathrm{IF}_k, \alpha\rangle + 2\sum_{\ell<k} \langle \boldsymbol{\Gamma}\phi_k, \phi_\ell\rangle \langle \mathrm{IF}_\ell, \alpha\rangle$$
$$+2\sum_{\ell=1}^{k} \langle \boldsymbol{\Gamma}\phi_k, \mathrm{IF}_\ell\rangle \langle \phi_\ell, \alpha\rangle - 2\rho\sum_{\ell=1}^{k} \langle D^2\mathrm{IF}_\ell, \phi_k\rangle \langle \phi_\ell, \alpha\rangle - 4\rho\sum_{\ell=1}^{k} \langle D\phi_k, D\phi_\ell\rangle \langle \mathrm{IF}_\ell, \alpha\rangle.$$

Since this equality holds for all α, we obtain

$$\pi_k A_k - 2\rho D^2 \mathrm{IF}_k + 2\rho \sum_{\ell=1}^{k} \langle D^2 \mathrm{IF}_k, \phi_\ell \rangle \phi_\ell = 2\lambda_k \mathrm{IF}_k + 2 \sum_{\ell<k} \langle \mathbf{\Gamma}\phi_k, \phi_\ell \rangle \mathrm{IF}_\ell + 2 \sum_{\ell=1}^{k} \langle \mathbf{\Gamma}\phi_k, \mathrm{IF}_\ell \rangle \phi_\ell - 2\rho \sum_{\ell=1}^{k} \langle D^2 \mathrm{IF}_\ell, \phi_k \rangle \phi_\ell - 4\rho \sum_{\ell=1}^{k} \langle D\phi_k, D\phi_\ell \rangle \mathrm{IF}_\ell .$$

After some reordering, we have that

$$(\pi_k \mathbf{\Gamma} - \lambda_k I_{\mathcal{H}}) \mathrm{IF}_k = -\frac{1}{2}\pi_k \nabla_k + \rho D^2 \mathrm{IF}_k - \rho \sum_{\ell=1}^{k} \langle D^2 \mathrm{IF}_k, \phi_\ell \rangle \phi_\ell + \sum_{\ell<k} \langle \mathbf{\Gamma}\phi_k, \phi_\ell \rangle \mathrm{IF}_\ell - \rho \sum_{\ell=1}^{k} \langle D^2 \mathrm{IF}_\ell, \phi_k \rangle \phi_\ell - 2\rho \sum_{\ell=1}^{k} \langle D\phi_k, D\phi_\ell \rangle \mathrm{IF}_\ell$$

where π_k is the projection onto the linear space orthogonal to that spanned by $\{\phi_1, \ldots, \phi_k\}$.

Let us compute IF_1. We have

$$(\pi_1 \mathbf{\Gamma} - \lambda_1 I_{\mathcal{H}}) \mathrm{IF}_1 = -\frac{1}{2}\pi_1 \nabla_1 + \rho D^2 \mathrm{IF}_1 - 2\rho \langle D^2 \mathrm{IF}_1, \phi_1 \rangle \phi_1 - 2\rho \langle D\phi_1, D\phi_1 \rangle \mathrm{IF}_1 . \tag{A.25}$$

Using that $\phi_i(t) = \cos(a_i \pi t)$ or $\sin(a_i \pi t)$, where $a_1 = 0$ if $\phi_1 = 1$, we get $D\phi_1 = -a_1^2 \pi^2 \phi_1$ and $D^2 \phi_1 = a_1^4 \pi^4 \phi_1$. Thus, we can rewrite (A.25) as $(\pi_1 \mathbf{\Gamma} - \lambda_1 I_{\mathcal{H}}) \mathrm{IF}_1 - \rho D^2 \mathrm{IF}_1 + 2\rho a_1^4 \pi^4 (\phi_1 \otimes \phi_1) \mathrm{IF}_1 + 2\rho a_1^4 \pi^4 \mathrm{IF}_1 = -\pi_1 \nabla_1 / 2$, so that

$$\{\pi_1 \mathbf{\Gamma} + 2\rho a_1^4 \pi^4 \phi_1 \otimes \phi_1 - (\lambda_1 - 2\rho a_1^4 \pi^4) I_{\mathcal{H}}\} \mathrm{IF}_1 = -\frac{1}{2}\pi_1 \nabla_1 + \rho D^2 \mathrm{IF}_1 . \tag{A.26}$$

Let $T = \pi_1 \mathbf{\Gamma} + 2\rho a_1^4 \pi^4 \phi_1 \otimes \phi_1 - (\lambda_1 - 2\rho a_1^4 \pi^4) I_{\mathcal{H}}$. We will strive to found a left-inverse of T. Note that $\pi_1 \mathbf{\Gamma} = \sum_{i\geq 2} \lambda_i \phi_i \otimes \phi_i$. Therefore, $T = (-\lambda_1 + 2\rho a_1^4 \pi^4)(\phi_1 \otimes \phi_1) + \sum_{i\geq 2} (\lambda_i - \lambda_1 + 2\rho a_1^4 \pi^4) \phi_i \otimes \phi_i$, so that

$$T^{-1} = \frac{1}{4\rho a_1^4 \pi^4 - \lambda_1} (\phi_1 \otimes \phi_1) + \sum_{i\geq 2} \frac{1}{\lambda_i - \lambda_1 + 2\rho a_1^4 \pi^4} \phi_i \otimes \phi_i = \sum_{i\geq 1} \upsilon_i \phi_i \otimes \phi_i .$$

Let us show that T^{-1} is well defined, that is, that $T^{-1}x \in \mathcal{H}$. We only have to see that $\|T^{-1}x\|^2 < \infty$ which is equivalent to show that

$$\sum_{i\geq 2} \left(\frac{1}{\lambda_i - \lambda_1 + 2\rho a_1^4 \pi^4} \right)^2 \langle \phi_i, x \rangle^2 < \infty .$$

Notice that $|\lambda_i - \lambda_1 + 2\rho a_1^4 \pi^4| = |\lambda_1 - \lambda_i - 2\rho a_1^4 \pi^4|$, $\lambda_1 - \lambda_i - 2\rho a_1^4 \pi^4 > \lambda_1 - \lambda_2 - 2\rho a_1^4 \pi^4 \neq 0$ for all i and also that $\lambda_1 - 2\rho a_1^4 \pi^4 = M \neq 0$.

Therefore, there exists i_0 such that for $i \geq i_0$ we have that $|\lambda_1 - \lambda_i - 2\rho a_1^4 \pi^4| \leq |M|/2$. Thus,

$$\sum_{i\geq i_0} \left(\frac{1}{\lambda_i - \lambda_1 + 2\rho a_1^4 \pi^4} \right)^2 \langle \phi_i, x \rangle^2 \leq \frac{4}{M^2} \sum_{i\geq i_0} \langle \phi_i, x \rangle^2 \leq \frac{4}{M^2} \|x\|^2 < \infty$$

which entails that T^{-1} is well-defined.

Applying this inverse T^{-1} to (A.26) we obtain that

$$\mathrm{IF}_1 = T^{-1} \left(-\frac{1}{2}\pi_1 \nabla_1 + \rho D^2 \mathrm{IF}_1 \right) = -\frac{1}{2} \sum_{i\geq 2} \upsilon_i (\phi_i \otimes \phi_i) \pi_1 \nabla_1 + \rho \sum_{i\geq 1} \upsilon_i \langle \mathrm{IF}_1, a_i^4 \pi^4 \phi_i \rangle \phi_i ,$$

where we have used that $\langle D^2 \mathrm{IF}_1, \phi_i \rangle = \langle \mathrm{IF}_1, D^2 \phi_i \rangle = \langle \mathrm{IF}_1, a_i^4 \pi^4 \phi_i \rangle$.

For $i \geq 2$, we have that $\phi_i \otimes \phi_i \pi_1 = \phi_i \otimes \phi_i$, hence, we get

$$\mathrm{IF}_1 = -\frac{1}{2} \sum_{i\geq 2} \upsilon_i \langle \phi_i, \nabla_1 \rangle \phi_i + \rho \sum_{i\geq 1} \upsilon_i \langle \mathrm{IF}_1, a_i^4 \pi^4 \phi_i \rangle \phi_i . \tag{A.27}$$

Writing, $\mathrm{IF}_1 = \sum_{\ell\geq 1} \langle \phi_\ell, \mathrm{IF}_1 \rangle \phi_\ell = \sum_{\ell\geq 1} \nu_\ell \phi_\ell$. From (A.27) we get that if $\ell = 1$, $\nu_1 = \langle \phi_1, \mathrm{IF}_1 \rangle = \rho \upsilon_1 a_1^4 \pi^4 \nu_1$. Since all υ_1 and ρ are not zero, we conclude that $\nu_1 = 0$.

Let us compute ν_ℓ for $\ell > 1$. We have that $\nu_\ell = \langle \phi_\ell, \text{IF}_1 \rangle = -\upsilon_\ell \langle \phi_\ell, \nabla_1 \rangle / 2 + \rho \upsilon_\ell a_\ell^4 \pi^4 \nu_\ell$, which entails that $\nu_\ell = -\upsilon_\ell \langle \phi_\ell, \nabla_1 \rangle / \{2(1 - \rho \upsilon_\ell a_\ell^4 \pi^4)\}$, so that

$$\text{IF}_1 = -\frac{1}{2} \sum_{\ell \geq 2} \frac{\upsilon_\ell \langle \phi_\ell, \nabla_1 \rangle}{(1 - \rho \upsilon_\ell a_\ell^4 \pi^4)} \phi_\ell. \tag{A.28}$$

Using that $a_\ell \to \infty$, we obtain that $\upsilon_\ell \to 0$. Let us check that IF_1 is well defined, that is, that $\|\text{IF}_1\|^2 < \infty$. Since,

$$\begin{aligned} \|\text{IF}_1\|^2 &= \frac{1}{4} \sum_{\ell \geq 2} \frac{\upsilon_\ell^2 \langle \phi_\ell, \nabla_1 \rangle^2}{(1 - \rho \upsilon_\ell a_\ell^4 \pi^4)^2} = \frac{1}{4} \sum_{\ell \geq 2} \frac{\left(\frac{1}{\lambda_\ell - \lambda_1 + 2\rho a_1^4 \pi^4} \right)^2}{\left(1 - \frac{\rho a_\ell^4 \pi^4}{\lambda_\ell - \lambda_1 + 2\rho a_1^4 \pi^4} \right)^2} \langle \phi_\ell, \nabla_1 \rangle^2 \\ &= \frac{1}{4} \sum_{\ell \geq 2} \left(\frac{1}{\lambda_\ell - \lambda_1 + 2\rho a_1^4 \pi^4 - \rho a_\ell^4 \pi^4} \right)^2 \langle \phi_\ell, \nabla_1 \rangle^2 \leq \frac{M}{4} \sum_{\ell \geq 2} \langle \phi_\ell, \nabla_1 \rangle^2 \leq \frac{M}{4} \|\nabla_1\|^2 < \infty, \end{aligned}$$

where we have used again that since $\lambda_\ell \to 0$ and $a_\ell \to \infty$, the terms $|\lambda_\ell - \lambda_1 + 2\rho a_1^4 \pi^4 - \rho a_\ell^4 \pi^4| > \lambda_1 - \lambda_2 + 2\rho a_2^4 \pi^4 - \rho a_1^4 \pi^4 = M^{-1}$. Thus, IF_1 is well defined.

Recall that

$$\nabla_1 = 2 \, \text{IF}\left(\frac{\langle \phi_1, x \rangle}{\sqrt{\lambda_1}}; \sigma_{\text{R}}^2, F_0 \right) \lambda_1 \phi_1 + \sqrt{\lambda_1} \text{DIF}\left(\frac{\langle \phi_1, x \rangle}{\sqrt{\lambda_1}}; \sigma_{\text{R}}^2, F_0 \right) (x - \langle x, \phi_1 \rangle \phi_1).$$

Thus, (A.28) entails that

$$\text{IF}_1 = -\frac{1}{2} \sum_{\ell \geq 2} \frac{\upsilon_\ell \left\langle \phi_\ell, \sqrt{\lambda_1} \text{DIF}\left(\frac{\langle \phi_1, x \rangle}{\sqrt{\lambda_1}}; \sigma_{\text{R}}^2, F_0 \right) x \right\rangle}{(1 - \rho \upsilon_\ell a_\ell^4 \pi^4)} \phi_\ell = -\frac{1}{2} \sum_{\ell \geq 2} \sqrt{\lambda_1} \text{DIF}\left(\frac{\langle \phi_1, x \rangle}{\sqrt{\lambda_1}}; \sigma_{\text{R}}^2, F_0 \right) \frac{\upsilon_\ell \langle \phi_\ell, x \rangle}{(1 - \rho \upsilon_\ell a_\ell^4 \pi^4)} \phi_\ell.$$

Therefore,

$$\text{IF}_1 = -\frac{1}{2} \sqrt{\lambda_1} \text{DIF}\left(\frac{\langle \phi_1, x \rangle}{\sqrt{\lambda_1}}; \sigma_{\text{R}}^2, F_0 \right) \sum_{\ell \geq 2} \frac{\upsilon_\ell \langle \phi_\ell, x \rangle}{(1 - \rho \upsilon_\ell a_\ell^4 \pi^4)} \phi_\ell.$$

Using the definition of υ_ℓ, and the fact that $2\text{DIF}(x, \sigma_{\text{R}}, F_0) = 2\text{DIF}(x, \sigma_{\text{R}}^2, F_0)$, we get that

$$\text{IF}_1 = -\sqrt{\lambda_1} \text{DIF}\left(\frac{\langle \phi_1, x \rangle}{\sqrt{\lambda_1}}; \sigma_{\text{R}}, F_0 \right) \sum_{\ell \geq 2} \frac{1}{\lambda_\ell - \lambda_1 + 2\rho a_1^4 \pi^4} \frac{\langle \phi_\ell, x \rangle}{(1 - \rho a_\ell^4 \pi^4 / (\lambda_\ell - 2\lambda_1 + \rho a_1^4 \pi^4))} \phi_\ell,$$

which implies that

$$\text{IF}_1 = -\sqrt{\lambda_1} \text{DIF}\left(\frac{\langle \phi_1, x \rangle}{\sqrt{\lambda_1}}; \sigma_{\text{R}}, F_0 \right) \sum_{\ell \geq 2} \frac{\langle \phi_\ell, x \rangle}{\lambda_\ell - \lambda_1 - \rho(a_\ell^4 - 2a_1^4)\pi^4} \phi_\ell,$$

as was desired.

References

[1] Y. Aït Sahalia, The delta method for nonlinear kernel functionals (Ph.D. dissertation), University of Chicago, 1995.

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

[3] J. Bali, G. Boente, D.E. Tyler, J.L. Wang, Robust functional principal components: a projection-pursuit approach, Ann. Statist. 39 (2011) 2852–2882.

[4] A.E. Beaton, J.W. Tukey, The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data, Technometrics 16 (2) (1974) 147–185.

[5] G. Boente, A.M. Pires, I.M. Rodrigues, Influence functions and outlier detection under the common principal components model: a robust approach, Biometrika 89 (2002) 861–875.

[6] G. Boente, A. Pires, I. Rodrigues, Detecting influential observations in principal components and common principal components, Comput. Statist. Data Anal. 54 (2010) 2967–2975.

[7] G. Boente, M. Salibian-Barrera, D. Tyler, A characterization of elliptical distributions and some optimality properties of principal components for functional data, J. Multivariate Anal. 131 (2014) 254–264.

[8] F. Critchley, Influence in principal components analysis, Biometrika 72 (1985) 627–636.

[9] C. Croux, Efficient high-breakdown *M*-estimators of scale, Statist. Probab. Lett. 19 (1994) 371–379.

[10] C. Croux, G. Haesbroeck, Empirical influence functions for robust principal component analysis, in: Proceedings of the Statistical Computing Section of the American Statistical Association, Am. Statist. Assoc., Alexandria, VA, 1999, pp. 201–206.

[11] C. Croux, A. Ruiz-Gazen, High breakdown estimators for principal components: the projection-pursuit approach revisited, J. Multivariate Anal. 95 (2005) 206–226.

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

[13] M. Febrero, P. Galeano, W. González-Manteiga, Functional analysis of NO_x levels: location and scale estimation and outlier detection, Comput. Statist. 22 (2007) 411–427.

[14] M. Febrero, P. Galeano, W. González-Manteiga, Outlier detection in functional data by depth measures, with application to identify abnormal NO_x levels, Environmetrics 19 (2008) 331–345.
[15] L.T. Fernholz, von Mises Calculus for Statistical Functionals, in: Lecture Notes in Statistics, vol. 19, Springer Verlag, New York, 1983.
[16] L.T. Fernholz, Smoothed versions of statistical functionals, in: S. Morgenthaler, E. Ronchetti, W. Stahel (Eds.), New Directions in Statistical Data Analysis and Robustness, Springer-Verlag, Basel, CH, 1993.
[17] L.T. Fernholz, Robustness issues regarding content corrected tolerance limits, Metrika 55 (2002) 53–66.
[18] T. Flett, Differential Analysis, Cambridge University Press, Cambridge, 1980.
[19] D. Gervini, Robust functional estimation using the spatial median and spherical principal components, Biometrika 95 (2008) 587–600.
[20] D. Gervini, Outlier detection and trimmed estimation for general functional data, Statist. Sinica 22 (2012) 1639–1660.
[21] P.J. Huber, Robust Statistics, Wiley, 1981.
[22] R.J. Hyndman, H.L. Shang, Rainbow plots, bagplots, and boxplots for functional data, J. Comput. Graph. Statist. 19 (2010) 29–45.
[23] G. Li, Z. Chen, Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo, J. Amer. Statist. Assoc. 80 (1985) 759–766.
[24] N. Locantore, J.S. Marron, D.G. Simpson, N. Tripoli, J.T. Zhang, K.L. Cohen, Robust principal components for functional data (with discussion), TEST 8 (1999) 1–73.
[25] S. López-Pintado, J. Romo, On the concept of depth for functional data, J. Amer. Statist. Assoc. 104 (2009) 718–734.
[26] G. Pison, P.J. Rousseeuw, P. Filzmoser, C. Croux, A robust version of principal factor analysis, in: J. Bethlehem, P. van der Heijden (Eds.), COMPSTAT: Proceedings in Computational Statistics, Physica-Verlag, Heidelberg, 2000, pp. 385–390.
[27] G. Pison, S. van Aelst, Diagnostic plots for robust multivariate methods, J. Comput. Graph. Statist. 13 (2004) 1–20.
[28] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer-Verlag, New York, 2005.
[29] J.A. Rice, B.W. Silverman, Estimating the mean and covariance structure nonparametrically when the data are curves, J. R. Stat. Soc. Ser. B 53 (1991) 233–243.
[30] P.J. Rousseeuw, B.C. van Zomeren, Unmasking multivariate outliers and leverage points, J. Amer. Statist. Assoc. 85 (1990) 633–639.
[31] P. Sawant, N. Billor, H. Shin, Functional outlier detection with robust functional principal component analysis, Comput. Statist. 27 (2011) 83–102.
[32] L. Shi, Local influence in principal components analysis, Biometrika 84 (1997) 175–186.
[33] B.W. Silverman, Smoothed functional principal components analysis by choice of norm, Ann. Statist. 24 (1996) 1–24.
[34] Y. Sun, M. Genton, Functional Boxplots, J. Comput. Graph. Statist. 20 (2011) 316–334.
[35] Y. Sun, M. Genton, Adjusted functional boxplots for spatio-temporal data visualization and outlier detection, Environmetrics 23 (2012) 54–64.
[36] J. Tamine, Smoothed influence function: another view at robust nonparametric regression, in: Discussion Paper 62, Sonderforschungsbereich 373, Humboldt-Universität zu Berlin, 2002.
[37] J. Tukey, Exploratory Data Analysis, Addison-Wesley, Reading, MA, 1977.