Journal of Multivariate Analysis 170 (2019) 46–62

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Robust sieve estimators for functional canonical correlation analysis

Agustín Alvarez [a], Graciela Boente [a,b,*], Nadia Kudraszow [c,b]

[a] Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
[b] CONICET, Argentina
[c] Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina

ARTICLE INFO

Article history:
Received 14 September 2017
Available online 19 March 2018

AMS subject classifications:
62G35
62H20

Keywords:
Canonical correlation
Fisher-consistency
Functional data
Robust estimation
Sieves

ABSTRACT

In this paper, we propose robust estimators for the first canonical correlation and directions of random elements on Hilbert separable spaces by combining sieves and robust association measures, leading to Fisher-consistent estimators for appropriate choices of the association measure. Under regularity conditions, the resulting estimators are consistent. The robust procedure allows us to construct detection rules to identify possible influential observations. The finite sample performance is illustrated through a simulation study in which contaminated data is included. The benefits of considering robust estimators are also illustrated on a real data set where the detection methods reveal the presence of influential observations for the first canonical directions that would be missed otherwise.

1. Introduction

Due to the growing interest in studying complex data, functional data analysis has become a relevant subject. When dealing with functional data, each observation consists of one or several infinite-dimensional objects such as curves, surfaces or images rather than scalars or vectors. Functional data analysis has applications in a wide range of fields (archaeology, medical science, biometrics, econometrics, environmetrics, chemometrics, etc.). As mentioned in Ramsay and Silverman [40], in many areas of statistics the collected data are more naturally represented as functions rather than finite-dimensional numerical vectors. It has been extensively discussed that simplifying the functional model by discretizing the observations as sequences of numbers can often fail to capture some of its important characteristics, such as the smoothness and continuity of the underlying functions. Statistical methods to analyse such functional data may be found, for instance, in Ferraty and Romain [18], Ferraty and Vieu [19], Horváth and Kokoszka [28], Hsing and Eubank [29] and Ramsay and Silverman [40]. For a summary of recent advances in functional statistics see Aneiros et al. [4], Cuevas [14] and Goia and Vieu [22].

When the observed data are infinite-dimensional, dimension reduction is an important task. To tackle this problem and depending on the goal sought different procedures have been considered, including functional principal components, single-index functional data analysis and other effective reduction methods as discussed in Yao et al. [44]. Functional canonical correlation analysis provides a useful dimension reduction tool to quantify correlation or association between two functions recorded for a sample of subjects on the same population. For multivariate data, canonical correlation analysis is performed by obtaining linear combinations of each subset of variables that maximize their correlation with the restriction that their variances are equal to one. Under a Gaussian model, Leurgans et al. [35] showed that the natural extension of multivariate

* Correspondence to: Departamento de Matemáticas, FCEyN, UBA, Ciudad Universitaria, Pabellón 2, Buenos Aires, C1428EHA, Argentina.
E-mail addresses: agustinalvarez01@gmail.com (A. Alvarez), gboente@dm.uba.ar (G. Boente), nkudraszow@mate.unlp.edu.ar (N. Kudraszow).

https://doi.org/10.1016/j.jmva.2018.03.003
0047-259X/

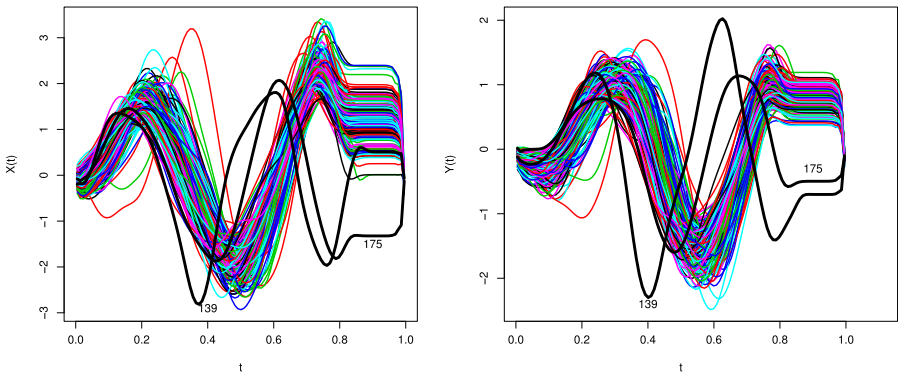

Fig. 1. Speed of the pen on the horizontal axis X (left panel) and on the vertical axis Y (right panel). The trajectories labelled 139 and 175 are shown in solid black lines.

estimators to the functional scenario fails since there is always a pair of directions with empirical canonical correlation equal to one. To solve this problem, they proposed a consistent estimator which penalizes the roughness of the canonical directions. On the other hand, He et al. [25] provided conditions ensuring the existence and proper definition of the canonical directions and correlations for processes admitting a Karhunen–Loève expansion, see also Cupidon et al. [16], while Cupidon et al. [17] derived the asymptotic distribution of regularized functional canonical correlation and variates.

All aforementioned authors studied the problem of maximizing the Pearson correlation, which is known to be sensitive to atypical observations (see Taskinen et al. [42]). In the multivariate scenario, there are several approaches which consider robust estimators for the canonical correlations and directions. Croux and Dehon [10] and Karnel [34] proposed a robust canonical correlation analysis replacing the sample covariance estimators by M-estimators or minimum covariance determinant estimators of multivariate scatter in the equations defining the classical canonical directions. The influence function of these robust plug-in canonical correlation estimators was discussed in Taskinen et al. [42]. Exploiting the relation between canonical correlation and regression estimation, Filzmoser et al. [20] proposed using robust alternating regression to obtain estimators of the first canonical variates. This proposal was generalized in Branco et al. [9] to estimate all the canonical variates. A different approach based on projection-pursuit uses a robust association measure instead of the Pearson correlation in the maximization procedure. This procedure has been considered in Alfons et al. [2] and Croux and Filzmoser [12] for multivariate data. Furthermore, Jin and Cui [33] studied the asymptotic distribution of the estimators obtained by projection-pursuit. As far as we know, there are no robust proposals in the functional setting.

As a motivation, consider the writing data set consisting of 2858 character samples that correspond to the speed profile of a pen writing on a tablet. The data is available at https://archive.ics.uci.edu/ml/datasets/Character+Trajectories and has been used by Williams et al. [43] to have a better understanding of the sub-blocks arising in biological movements and their timings. Among others, this data set has been considered in Hubert et al. [30] to illustrate their depth-based functional outlier detection method and in Hubert et al. [31] to perform robust supervised functional classification based on depth. We focus on the 186 data related to the speed of the pen on the horizontal and vertical axis when writing the letter "e", denoted $X(t)$ and $Y(t)$ respectively. We seek to explain how the movement variability on the horizontal axis is related to that of the vertical one. Fig. 1 displays the data trajectories where some trajectories, such as those labelled 139 and 175, clearly seem to deviate from the bulk of the data. For that reason, it is important to provide a procedure for estimating the canonical directions that will not be influenced by atypical trajectories, as well as to provide a rule allowing for the identification of these influential trajectories.

In this paper, we introduce robust estimators for the first canonical correlation and directions adapting the robust projection pursuit approach considered in Alfons et al. [2] to the functional data setting by means of a sieve method. As is well known, the sieves method involves approximating an infinite-dimensional parameter space Θ by a sequence of finite-dimensional parameter spaces Θ_n, which depend on the sample size n, and then estimating the parameters on the spaces Θ_n rather than on Θ, allowing for the introduction of the dimension of Θ_n as a smoothing parameter.

The paper is organized as follows. In Section 2, we begin by introducing some notation and then review some basic concepts used throughout the paper. In Section 3, we define the robust estimators of the first canonical direction and canonical correlation based on general association measures. Consistency results are established in Section 4, while Fisher-consistency of the related functionals is presented in Section 4.1. Section 5 discusses the use of the given robust proposals to construct detection rules allowing for the detection of influential observations, which is an important step in any analysis. The results of a Monte Carlo study conducted to examine the robustness and finite sample performance of the proposed procedures are reported in Section 6. The writing data set is analysed in Section 7, where the advantage of the proposed estimators is illustrated. Furthermore, the robust estimators allow to identify observations with a different writing pattern. Finally, Section 8 contains some concluding remarks. Proofs are relegated to the technical supplementary material available on-line.

2. Notation and preliminaries

Let $\mathcal{H}_j, j = 1, 2$, be Hilbert separable spaces with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_j}$ and norm $\|\cdot\|_{\mathcal{H}_j}$. Denote $\mathcal{S}_j = \{u \in \mathcal{H}_j : \|u\|_{\mathcal{H}_j} = 1\}$, $j = 1, 2$, the unit sphere in $\mathcal{H}_j$, $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$ the product space and $\langle (u_1, v_1), (u_2, v_2) \rangle_{\mathcal{H}} = \langle u_1, u_2 \rangle_{\mathcal{H}_1} + \langle v_1, v_2 \rangle_{\mathcal{H}_2}$ the inner product in $\mathcal{H}$. Let $Z = (X, Y)^\top$ be a random element of the Hilbert space $\mathcal{H}$ defined in $(\Omega, \mathcal{A}, \Pr)$. When Z has finite second moment, i.e., $\mathrm{E}(\|Z\|^2_{\mathcal{H}}) < \infty$, we denote as $\Gamma_{XX} : \mathcal{H}_1 \to \mathcal{H}_1$, $\Gamma_{YY} : \mathcal{H}_2 \to \mathcal{H}_2$, $\Gamma_{XY} : \mathcal{H}_2 \to \mathcal{H}_1$ and $\Gamma_{YX} : \mathcal{H}_1 \to \mathcal{H}_2$, the covariance and cross-covariance operators, respectively. More precisely, for any $u_1, u_2 \in \mathcal{H}_1$, $v \in \mathcal{H}_2$, we have that $\mathrm{Cov}(\langle u_1, X \rangle_{\mathcal{H}_1}, \langle u_2, X \rangle_{\mathcal{H}_1}) = \langle u_1, \Gamma_{XX} u_2 \rangle_{\mathcal{H}_1}$, $\mathrm{Cov}(\langle u_1, X \rangle_{\mathcal{H}_1}, \langle v, Y \rangle_{\mathcal{H}_2}) = \langle u_1, \Gamma_{XY} v \rangle_{\mathcal{H}_1}$ and similarly for Γ_{YY} and Γ_{YX}. With this notation, the covariance operator $\Gamma_{ZZ} : \mathcal{H} \to \mathcal{H}$ can be decomposed as $\Gamma_{ZZ}(u, v) = (\Gamma_{XX} u + \Gamma_{XY} v, \Gamma_{YX} u + \Gamma_{YY} v)^\top$, allowing to write Γ_{ZZ} in the matrix form

$$\Gamma_{ZZ}(u, v) = \begin{pmatrix} \Gamma_{XX} & \Gamma_{XY} \\ \Gamma_{YX} & \Gamma_{YY} \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}.$$

Given a random sample $\mathbf{W}_i = (S_i, T_i)^\top$, $1 \le i \le n$, of a random vector $\mathbf{W} = (S, T)^\top : \Omega \to \mathbb{R}^2$, $P_n[S, T]$ will denote the empirical measure induced on $\mathbb{R}^2$. More precisely, for $A \in \mathcal{B}$, where $\mathcal{B}$ is the Borel σ-Algebra of $\mathbb{R}^2$, $P_n[S, T](A) = (1/n) \sum_{i=1}^n \mathbb{I}_A(\mathbf{W}_i)$, where $\mathbb{I}_A(\cdot)$ stands for the indicator function of the set A.

2.1. Association measures

As mentioned in Alfons et al. [2], robust estimators of the canonical correlation for multivariate data can be defined using the maximal value that a bivariate association measure can attain between any pair of one-dimensional projections. For that purpose, association measures giving an alternative to the classical Pearson correlation must be considered. In this section, we recall the definition of an association measure and we review some examples of such measures.

The association ρ between two univariate variables S and T, denoted $\rho(S, T)$, is a functional defined in the space of bivariate probability measures that verifies the following conditions

(i) $\rho(S, T) = \rho(T, S)$,
(ii) $\rho(aS + b, cT + d) = \mathrm{sign}(ac)\, \rho(S, T)$ for all $a, b, c, d \in \mathbb{R}$,
(iii) $\rho^2(S, T) \le 1$.

When $(S, T)^\top \sim P[S, T]$, we also denote $\rho(P[S, T]) = \rho(S, T)$.

The most well-known association measure is the *Pearson correlation*, denoted ρ_{CL}, which measures linear correlation. For a bivariate random vector $\mathbf{W} = (S, T)^\top \sim P[S, T]$ with covariance matrix $\mathbf{\Sigma} = \mathrm{Var}(\mathbf{W})$, we have

$$\rho_{\mathrm{CL}}(\mathbf{W}) = \rho_{\mathrm{CL}}(P[S, T]) = \frac{\mathbf{\Sigma}_{12}}{\sqrt{\mathbf{\Sigma}_{11} \mathbf{\Sigma}_{22}}}.$$

It is well known that ρ_{CL} is very sensitive to the presence of outliers. Alternatives to the Pearson correlation can be seen in Maronna et al. [38] or Shevlyakov and Vilchevski [41] among others. We will now briefly describe some of them.

The *Spearman's rank correlation* $\widetilde{\rho}_{\mathrm{SP}}$ and *Kendall's tau correlation* $\widetilde{\rho}_{\mathrm{K}}$ are well known association measures that have been used in the finite-dimensional setting by Alfons et al. [2]. They are defined as

$$\widetilde{\rho}_{\mathrm{SP}}(P[S, T]) = \rho_{\mathrm{CL}}\{F_S(S), F_T(T)\} \qquad \text{and} \qquad \widetilde{\rho}_{\mathrm{K}}(P[S, T]) = \mathrm{E}\left[\mathrm{sgn}\left\{(S_1 - S_2)(T_1 - T_2)\right\}\right] \tag{1}$$

respectively, where F_S and F_T are the cumulative distribution functions of S and T and $(S_1, T_1)^\top$ and $(S_2, T_2)^\top$ are two independent copies of $(S, T)^\top$. A discussion on the robustness properties of Spearman's and Kendall's correlations can be found in Croux and Dehon [11].

Another robust alternative to the Pearson correlation can be obtained by means of a bivariate robust scatter functional $\mathbf{V} = \mathbf{V}(S, T)$ instead of the classical covariance matrix $\mathbf{\Sigma}$. The association measure induced by a bivariate scatter matrix $\mathbf{V}$ is given by

$$\rho_{\mathbf{V}}(P[S, T]) = \frac{\mathbf{V}_{12}(S, T)}{\{\mathbf{V}_{11}(S, T) \mathbf{V}_{22}(S, T)\}^{\frac{1}{2}}}, \tag{2}$$

where $\mathbf{V}_{ij}(S, T)$ is the (i, j)th element of the scatter matrix $\mathbf{V}(S, T)$. One possible choice for $\mathbf{V}(S, T)$ is the M-scatter estimator defined by Maronna [37], since it provides an efficient estimator which is also highly robust in the bivariate case. Another possible choice is to consider the orthogonalized Gnanadesikan–Kettenring covariance proposed by Maronna and Zamar [39]. When using M-estimators or the orthogonalized Gnanadesikan–Kettenring covariance the corresponding association measures will be denoted by $\rho_{\mathbf{V}_{\mathrm{M}}}$ and ρ_{OGK} respectively.

A desirable property is that different measures of association determine the same value, which is the target one, at least for a given distribution family $\mathcal{P}$. If this occurs we will say that the corresponding measure is Fisher-consistent at $\mathcal{P}$. This property will allow us to guarantee the consistency of the canonical directions and canonical association estimators to the

population quantities of interest. In particular, if $\mathcal{P} = \{P_{\sigma_1,\sigma_2,\kappa}\}$ is the family of bivariate elliptical distributions (which contains the family of normal bivariate distributions) with location zero and scatter matrix

$$\mathbf{\Sigma}_\kappa = \begin{pmatrix} \sigma_1^2 & \kappa\,\sigma_1\sigma_2 \\ \kappa\,\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

we want to identify which association measures consistently estimate the quantity $-1 < \kappa < 1$. Without loss of generality we can assume $\sigma_1 = \sigma_2 = 1$ and denote $\mathcal{P} = \{P_\kappa\}$. Note that, if we define $T(\kappa) = \rho(P_\kappa)$, Fisher-consistency is achieved when $T(\kappa) = \kappa$. Maronna et al. [38] show that, if the robust scatter functional $\mathbf{V}$ is affine-equivariant, $\mathbf{V}(P_\kappa) = c\,\mathbf{\Sigma}_\kappa$ for some positive constant c. Hence, the association measure $\rho_{\mathbf{V}}$ is Fisher-consistent for κ at the family of elliptical distributions. In particular, we have $\rho_{\mathbf{V}_M}(P_\kappa) = \kappa$. Even if the orthogonalized Gnanadesikan–Kettenring covariance is not affine equivariant, the results in Section 6.12.10 in Maronna et al. [38] easily entail that $\rho_{\text{OGK}}(P_\kappa) = \kappa$.

Let us consider the family of normal distributions $\mathcal{P}_{\mathcal{N}} = \{\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma}_\kappa)\}$. One may transform the Spearman and Kendall correlations defined in (1) as

$$\rho_{\text{SP}}(P[S,T]) = 2\sin\left\{\frac{\pi}{6}\,\widetilde{\rho}_{\text{SP}}(P[S,T])\right\} \quad \text{and} \quad \rho_{\text{K}}(P[S,T]) = \sin\left\{\frac{\pi}{2}\,\widetilde{\rho}_{\text{K}}(P[S,T])\right\} \tag{3}$$

to ensure that ρ_{SP} and ρ_{K} are Fisher-consistent at the family of bivariate normal distributions. We will call ρ_{SP} and ρ_{K} the normalized (or transformed) Spearman and Kendall correlations respectively.

3. The estimators

3.1. Functional canonical correlation analysis

Given $(X,Y)^\top : \Omega \to \mathcal{H}_1 \times \mathcal{H}_2$ a random element of the Hilbert space $\mathcal{H}$ with probability measure P, we denote the probability measure of $\left(\langle u, X\rangle_{\mathcal{H}_1},\ \langle v, Y\rangle_{\mathcal{H}_2}\right)^\top \in \mathbb{R}^2$ induced by P as $P[\langle u, X\rangle_{\mathcal{H}_1},\ \langle v, Y\rangle_{\mathcal{H}_2}]$. The statistical functionals of the canonical directions (or canonical weights) and the maximum canonical association defined in the multivariate setting by Alfons et al. [2] can be extended to the functional setting, replacing the inner product in the Euclidean space by the corresponding inner product in the Hilbert space. Given ρ an association measure as defined in Section 2.1, let $\rho_{XY}(u,v) = \rho\left(P[\langle u, X\rangle_{\mathcal{H}_1},\ \langle v, Y\rangle_{\mathcal{H}_2}]\right)$, $\rho_{XX}(u_1,u_2) = \rho\left(P[\langle u_1, X\rangle_{\mathcal{H}_1},\ \langle u_2, X\rangle_{\mathcal{H}_1}]\right)$ and $\rho_{YY}(v_1,v_2) = \rho\left(P[\langle v_1, Y\rangle_{\mathcal{H}_2},\ \langle v_2, Y\rangle_{\mathcal{H}_2}]\right)$. The first canonical direction (or weight) functionals related to the association measure ρ are defined as

$$(\Phi_1(P), \Psi_1(P)) = (\Phi_1, \Psi_1) = \underset{\|u\|_{\mathcal{H}_1}=\|v\|_{\mathcal{H}_2}=1}{\operatorname{argmax}} \rho_{XY}(u,v). \tag{4}$$

As in the multivariate canonical correlation analysis, the first or maximum canonical association is given by

$$\rho_1(P) = \rho_1 = \rho_{XY}(\Phi_1, \Psi_1) = \rho(P[\langle \Phi_1, X\rangle_{\mathcal{H}_1}, \langle \Psi_1, Y\rangle_{\mathcal{H}_2}]), \tag{5}$$

where $\langle \Phi_1, X\rangle_{\mathcal{H}_1}$ and $\langle \Psi_1, Y\rangle_{\mathcal{H}_2}$ are the first canonical variates. As mentioned in Alfons et al. [2], condition (ii) in the definition of the association measure ρ entails that the maximum canonical association is indeed non-negative. As in He et al. [24], higher order canonical variates may also be defined, see Section 8 for a discussion on this topic.

When the association measure is the Pearson correlation and the random element $(X,Y)^\top$ has finite second moments, the first canonical weights given in (4) will be denoted as $(\Phi_{\text{CL},1}, \Psi_{\text{CL},1})$. Furthermore, in this case the function $\rho_{XY} : \mathcal{H} \to \mathbb{R}$ can be defined using the covariance operator Γ_{ZZ} as

$$\rho_{XY}(u,v) = \frac{\langle u, \Gamma_{XY} v\rangle_{\mathcal{H}_1}}{\sqrt{\langle u, \Gamma_{XX} u\rangle_{\mathcal{H}_1}\,\langle v, \Gamma_{YY} v\rangle_{\mathcal{H}_2}}},$$

where the ratio $\langle u, \Gamma_{XY} v\rangle^2_{\mathcal{H}_1} / \left(\langle u, \Gamma_{XX} u\rangle_{\mathcal{H}_1}\,\langle v, \Gamma_{YY} v\rangle_{\mathcal{H}_2}\right)$ is equal to 0 when $\langle u, \Gamma_{XX} u\rangle = 0$ or $\langle v, \Gamma_{YY} v\rangle = 0$.

More generally, given a linear self-adjoint, positive semi-definite and compact operator $\Gamma : \mathcal{H} \to \mathcal{H}$ such that

$$\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}, \tag{6}$$

Γ induces the functionals $\rho_\Gamma : \mathcal{H} \to \mathbb{R}$, $\rho_{\Gamma_{11}} : \mathcal{H}_1 \times \mathcal{H}_1 \to \mathbb{R}$ and $\rho_{\Gamma_{22}} : \mathcal{H}_2 \times \mathcal{H}_2 \to \mathbb{R}$ as

$$\rho_\Gamma(u,v) = \frac{\langle u, \Gamma_{12} v\rangle_{\mathcal{H}_1}}{\sqrt{\langle u, \Gamma_{11} u\rangle_{\mathcal{H}_1}\langle v, \Gamma_{22} v\rangle_{\mathcal{H}_2}}} \tag{7}$$

$$\rho_{\Gamma_{11}}(u_1,u_2) = \frac{\langle u_1, \Gamma_{11} u_2\rangle_{\mathcal{H}_1}}{\sqrt{\langle u_1, \Gamma_{11} u_1\rangle_{\mathcal{H}_1}\langle u_2, \Gamma_{11} u_2\rangle_{\mathcal{H}_1}}} \qquad \rho_{\Gamma_{22}}(v_1,v_2) = \frac{\langle v_1, \Gamma_{22} v_2\rangle_{\mathcal{H}_2}}{\sqrt{\langle v_1, \Gamma_{22} v_1\rangle_{\mathcal{H}_2}\langle v_2, \Gamma_{11} v_2\rangle_{\mathcal{H}_2}}}, \tag{8}$$

where $u, u_1, u_2 \in \mathcal{H}_1$, $v, v_1, v_2 \in \mathcal{H}_2$ and $\rho_\Gamma(u,v) = 0$ if either $\langle u, \Gamma_{11} u\rangle_{\mathcal{H}_1} = 0$ or $\langle v, \Gamma_{22} v\rangle_{\mathcal{H}_2} = 0$. Moreover, $\rho_{\Gamma_{11}}(u_1,u_2) = 0$ when either $\langle u_1, \Gamma_{11} u_1\rangle_{\mathcal{H}_1} = 0$ or $\langle u_2, \Gamma_{11} u_2\rangle_{\mathcal{H}_1} = 0$ and similarly for $\rho_{\Gamma_{22}}$.

Using ρ_Γ, the canonical directions $(\Phi_{\Gamma,1}, \Psi_{\Gamma,1})$ associated to Γ can be defined as in (4), replacing ρ by ρ_Γ. The first or maximum canonical association associated to Γ, $\rho_{\Gamma,1}$, satisfies the equation

$$\rho_{\Gamma,1} = \rho_\Gamma(P[\langle \Phi_{\Gamma,1}, X\rangle_{\mathcal{H}_1}, \langle \Psi_{\Gamma,1}, Y\rangle_{\mathcal{H}_2}]) . \tag{9}$$

When ρ is the Pearson correlation and $\mathcal{H} = L^2(0,1) \times L^2(0,1)$, Leurgans et al. [35] showed that the estimation method used in the Euclidean case breaks down in the infinite-dimensional setting. More precisely, let $(X_1, Y_1)^\top, \ldots, (X_n, Y_n)^\top$ be a random sample of $(X, Y)^\top$ and denote as P_n its empirical distribution. When ρ is the Pearson correlation and $(X, Y)^\top$ is a Gaussian process, replacing in (4) and (5) the probability measure P by the empirical distribution P_n does not lead to reliable estimators since directions can be found with empirical canonical correlation equal to one. This property shows that in the classical setting it is necessary to use a method involving some smoothing. Proposition 3.1 shows that the same result holds when a general association measure is considered.

Given $(u, v) \in \mathcal{H}_1 \times \mathcal{H}_2$, denote as ρ_n the empirical version of ρ_{XY}, that is,

$$\rho_n(u, v) = \rho(P_n[\langle u, X\rangle_{\mathcal{H}_1}, \langle v, Y\rangle_{\mathcal{H}_2}]) . \tag{10}$$

We will need the following assumptions

- **A1** $\Pr(X \in \mathcal{L}) = 0$ for any finite-dimensional proper linear space $\mathcal{L}$ of $\mathcal{H}_1$.
- **A2** Given any bivariate probability measure Q supported on the line $\mathcal{R} = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = x_1\}$ such that $Q(\{x_1 = c\}) = Q(\{x_2 = c\}) < 1$ for all $c \in \mathbb{R}$, the bivariate association measure ρ satisfies that $\rho(Q) = 1$.

Remark 3.1. Straightforward arguments allow us to see that **A1** holds if for any orthonormal basis $\{\delta_j\}_{j\geq 1}$ of $\mathcal{H}_1$ and any $k \in \mathbb{N}$ such that $k \leq \dim(\mathcal{H}_1)$, the k-dimensional random vector $(\langle X, \delta_1\rangle_{\mathcal{H}_1}, \ldots, \langle X, \delta_k\rangle_{\mathcal{H}_1})^\top$ has a density. In particular, this last statement holds if X has an elliptical distribution, as defined in Bali and Boente [6], with a dispersion operator Γ_{11} with kernel reducing to $\{0\}$. Effectively, if $X \sim \mathcal{E}(\mu_1, \Gamma_{11}, \varphi)$ and $\ker(\Gamma_{11}) = \{0\}$, then X can be written as $X = \mu_1 + S\widetilde{X}$, where $\Pr(S > 0) = 1$ and the scalar random variable S and the Gaussian random element $\widetilde{X} \in \mathcal{H}_1$ are independent (see Boente et al. [8]). Moreover, the covariance operator of $\widetilde{X}$ is proportional to Γ_{11}. Hence, $(\langle X - \mu_1, \delta_1\rangle_{\mathcal{H}_1}, \ldots, \langle X - \mu_1, \delta_k\rangle_{\mathcal{H}_1})^\top = S\,(\langle \widetilde{X}, \delta_1\rangle_{\mathcal{H}_1}, \ldots, \langle \widetilde{X}, \delta_k\rangle_{\mathcal{H}_1})^\top$ and the result follows from the fact that $\mathbf{w} = (\langle \widetilde{X}, \delta_1\rangle_{\mathcal{H}_1}, \ldots, \langle \widetilde{X}, \delta_k\rangle_{\mathcal{H}_1})^\top$ has a kth dimensional non-singular multivariate normal distribution and is independent of S.

On the other hand, **A2** states that given a bivariate random vector $(S, T) \in \mathbb{R}^2$ such that $S = T$ almost surely, then $\rho(S, T) = 1$, whenever S and T are not constant. The Pearson correlation as well as the association measure ρ_{OGK} and the transformed Spearman or Kendall correlations, ρ_{SP} and ρ_{K} satisfy **A2**. As noted in Maronna [37], the M-scatter functional is not defined when $S = T$. However, defining the Mahalanobis distance as in page 185 of Maronna et al. [38] we get that $\rho_{\mathbf{v}_{\text{M}}} = 1$.

Proposition 3.1. *Assume that **A1** holds and that for any $v \in \mathcal{H}_2$, $\Pr(Y = v) = 0$. Let ρ be a bivariate association measure satisfying **A2** and $n \in \mathbb{N}$, $2 \leq n \leq \dim(\mathcal{H}_1)$. Then, with probability one there exist $u_n \in \mathcal{H}_1$ and $v_n \in \mathcal{H}_2$ such that $\rho_n(u_n, v_n) = 1$.*

It is worth noticing that the conclusion in Proposition 3.1 still holds when the roles of X and Y are reversed.

A direct consequence of Proposition 3.1 is that, if $\dim(\mathcal{H}_1) = \infty$, some kind of smoothing is needed since $\sup_{u\in\mathcal{S}_1, v\in\mathcal{S}_2} \rho_n(u, v) = 1$. Section 3.2 describes our proposal which combines robust projection-pursuit with the method of sieves as a smoothing tool.

3.2. Sieve approach for robust functional canonical correlation analysis

Let $(X_1, Y_1)^\top, \ldots, (X_n, Y_n)^\top$ be a random sample of the random element $(X, Y)^\top : \Omega \to \mathcal{H}$. Let $\{\delta_i\}_{i\geq 1}$ and $\{\eta_j\}_{j\geq 1}$ be orthonormal bases for $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively. From now on, let $\mathcal{H}_{1,p}$ and $\mathcal{H}_{2,q}$ denote the subspaces of $\mathcal{H}_1$ and $\mathcal{H}_2$ spanned by $\{\delta_1, \ldots, \delta_p\}$ and $\{\eta_1, \ldots, \eta_q\}$ respectively. We denote by $\mathcal{S}_{1,p} = \mathcal{S}_1 \cap \mathcal{H}_{1,p}$ and $\mathcal{S}_{2,q} = \mathcal{S}_2 \cap \mathcal{H}_{2,q}$ the unit spheres of $\mathcal{H}_{1,p}$ and $\mathcal{H}_{2,q}$ respectively. The sieve estimators for the first canonical directions are defined as

$$(\hat{\Phi}_1, \hat{\Psi}_1) = \underset{u\in\mathcal{S}_{1,p_n}, v\in\mathcal{S}_{2,q_n}}{\operatorname{argmax}} \rho_n(u, v) \tag{11}$$

where ρ_n is defined in (10), while the first or maximum canonical association estimator $\hat{\rho}_1 \geq 0$ is such that

$$\hat{\rho}_1 = \rho\left(P_n[\langle \hat{\Phi}_1, X\rangle_{\mathcal{H}_1}, \langle \hat{\Psi}_1, Y\rangle_{\mathcal{H}_2}]\right) = \rho_n\left(\hat{\Phi}_1, \hat{\Psi}_1\right) . \tag{12}$$

These estimators depend on the chosen bases of $\mathcal{H}_1$ and $\mathcal{H}_2$, on the sequences $(p_n)_{n\in\mathbb{N}}$ and $(q_n)_{n\in\mathbb{N}}$ and on the association measure ρ used. Some of the frequently used bases for functional data are the Fourier, polynomial, splines and wavelet bases; see, for instance, Ramsay and Silverman [40].

3.3. Selection of the smoothing parameters

Once the association measure ρ and the bases $\{\delta_i\}_{i\geq 1}$ and $\{\eta_j\}_{j\geq 1}$ are chosen, the selection of the approximating linear space dimensions p_n and q_n that play the role of the smoothing parameters is an important practical issue. The most popular general approach to address such a selection problem is to use cross-validation methods.

Denote by $\mathbf{r}_n = (p_n, q_n)$ the parameter to be determined. Let $(\hat{\Phi}^{(-i)}_{\mathbf{r}_n,1}, \hat{\Psi}^{(-i)}_{\mathbf{r}_n,1})$ be the first canonical direction estimators computed without the ith observation and when the approximating subspaces have dimensions p_n and q_n, that is,

$$(\hat{\Phi}^{(-i)}_{\mathbf{r}_n,1}, \hat{\Psi}^{(-i)}_{\mathbf{r}_n,1}) = \underset{u\in\mathcal{S}_{1,p_n}, v\in\mathcal{S}_{2,q_n}}{\operatorname{argmax}} \rho\left(P_n^{(-i)}[\langle u, X\rangle_{\mathcal{H}_1}, \langle v, Y\rangle_{\mathcal{H}_2}]\right),$$

where $P_n^{(-i)}(A) = (1/n)\sum_{j\neq i}\mathbb{I}_A(X_j, Y_j)$ for a Borel set A of $\mathcal{H}$.

Let $U^{(i)}_{\mathbf{r}_n,1} = \langle\hat{\Phi}^{(-i)}_{\mathbf{r}_n,1}, X_i\rangle_{\mathcal{H}_1}$ and $V^{(i)}_{\mathbf{r}_n,1} = \langle\hat{\Psi}^{(-i)}_{\mathbf{r}_n,1}, Y_i\rangle_{\mathcal{H}_2}$ be the canonical variates of the ith subject. He et al. [26] proposed a selection method based on maximizing the sample correlation of the canonical variates. However, as in other settings, using a non-robust criterion to select the smoothing parameters even if combined with a robust estimation procedure may lead to non-resistant estimators. For that reason, we propose using the same association measure ρ considered in the estimation step to select the smoothing parameters. Let

$$RCV_{\mathbf{r}_n} = \rho^2\left\{\frac{1}{n}\sum_{i=1}^{n}\Delta_{(U^{(i)}_{\mathbf{r}_n,1}, V^{(i)}_{\mathbf{r}_n,1})}\right\},$$

where $\Delta_{(a,b)}$ denotes the bivariate probability measure giving all its mass to the point (a, b). Given $\mathcal{R}$ a set of possible values for the parameter $\mathbf{r}_n$, the cross-validation parameter equals $\widehat{\mathbf{r}}$ where

$$\widehat{\mathbf{r}} = \underset{\mathbf{r}_n\in\mathcal{R}}{\operatorname{argmax}} RCV_{\mathbf{r}_n}. \tag{13}$$

As in He et al. [26], once the dimension is chosen, one may also choose the value $\widetilde{\rho}_1 = \sqrt{RCV_{\widehat{\mathbf{r}}}}$ as the estimator of the first canonical association.

3.4. Numerical implementation of the estimators

Except for the Pearson correlation ρ_{CL}, the maximizers of (11) cannot be computed exactly. For that reason, algorithms to obtain approximate solutions are needed.

Given $\{\delta_i\}_{i\geq 1}$ and $\{\eta_j\}_{j\geq 1}$ orthonormal basis of $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, the estimators proposed in (11) are obtained searching directions $u = \sum_{i=1}^{p_n} a_i\delta_i \in \mathcal{H}_1$ and $v = \sum_{i=1}^{q_n} b_i\eta_i \in \mathcal{H}_2$, $\sum_{i=1}^{p_n} a_i^2 = \sum_{i=1}^{q_n} b_i^2 = 1$, that lead to the maximum value of $\rho(P_n[\langle u, X\rangle_{\mathcal{H}_1}, \langle v, Y\rangle_{\mathcal{H}_2}])$. Denote $\mathbf{a} = (a_1, \ldots, a_{p_n})^\top$ and $\mathbf{b} = (b_1, \ldots, b_{q_n})^\top$ the coefficients of u and v in the considered basis and let $\mathbf{x} = \left(\langle X, \delta_1\rangle_{\mathcal{H}_1}, \ldots, \langle X, \delta_{p_n}\rangle_{\mathcal{H}_1}\right)^\top$ and $\mathbf{y} = (\langle Y, \eta_1\rangle_{\mathcal{H}_2}, \ldots, \langle Y, \eta_{q_n}\rangle_{\mathcal{H}_2})^\top$. Noting that $\langle u, X\rangle_{\mathcal{H}_1} = \mathbf{a}^\top\mathbf{x}$ and $\langle v, Y\rangle_{\mathcal{H}_2} = \mathbf{b}^\top\mathbf{y}$, the estimators given in (11) can be obtained using any multivariate algorithm allowing to find the vectors $\hat{\mathbf{a}}_1 = (\hat{a}_{11}, \ldots, \hat{a}_{1p_n})^\top$ and $\hat{\mathbf{b}}_1 = (\hat{b}_{11}, \ldots, \hat{b}_{1q_n})^\top$ with norm 1 which maximize $\rho(P_n[\mathbf{a}^\top\mathbf{x}, \mathbf{b}^\top\mathbf{y}])$, i.e.,

$$(\hat{\mathbf{a}}_1, \hat{\mathbf{b}}_1) = \underset{\|\mathbf{a}\|_{\mathbb{R}^p}=1, \|\mathbf{b}\|_{\mathbb{R}^q}=1}{\operatorname{argmin}} \rho(P_n[\mathbf{a}^\top\mathbf{x}, \mathbf{b}^\top\mathbf{y}]). \tag{14}$$

The alternate GRID algorithm described in Alfons et al. [1,2] is a well known algorithm allowing to perform multivariate canonical analysis for association measures such as the Spearman and Kendall correlation and the association measure based on a bivariate M-scatter matrix. It provides an accurate method to approximate the weights $\hat{\mathbf{a}}_1$ and $\hat{\mathbf{b}}_1$ maximizing $\rho(P_n[\mathbf{a}^\top\mathbf{x}, \mathbf{b}^\top\mathbf{y}])$, using optimization in two-dimensional spaces. The algorithm is implemented through the function `ccaGrid` in the `R` package `ccaPP`. Once the multivariate weights $\hat{\mathbf{a}}_1$ and $\hat{\mathbf{b}}_1$ are obtained, the canonical direction estimators in $\mathcal{H}_j$ can be reconstructed as

$$\hat{\Phi}_1 = \sum_{j=1}^{p_n}\hat{a}_{1j}\delta_j \quad \text{and} \quad \hat{\Psi}_1 = \sum_{j=1}^{q_n}\hat{b}_{1j}\eta_j. \tag{15}$$

4. Consistency

As in Cupidon et al. [17] and Leurgans et al. [35], in this section we show that under mild conditions the estimators of the first canonical weights and the related first canonical association defined in Section 3 are consistent to the functionals given in (4) and (5). It is worth noticing that our results include the multivariate setting by taking $p_n = \dim(\mathcal{H}_1)$ and $q_n = \dim(\mathcal{H}_2)$. In this sense, we extend the results given in Jin and Cui [33] from the Euclidean to the infinite-dimensional case in a more general setting, since weaker assumptions are required. Finally, it should be noticed that, in Theorem 4.1 when

$\dim(\mathcal{H}_1) < \infty$, we understand that the requirement $p_n \to \infty$ is replaced by $p_n = \dim(\mathcal{H}_1)$. Similarly, when $\dim(\mathcal{H}_2) < \infty$, $q_n = \dim(\mathcal{H}_2)$.

To derive the consistency of the estimators we will need the following assumptions. To avoid burden notation, from now on, let $(\Phi_1, \Psi_1) = (\Phi_1(P), \Psi_1(P))$ the solution of (4), which we assume to exist, and $\rho_1 = \rho_1(P)$ the functional defined in (5).

C1 $\sup_{u\in\mathcal{S}_{1,p_n}, v\in\mathcal{S}_{2,q_n}} |\rho^2_{XY}(u, v) - \rho^2_n(u, v)| \xrightarrow{a.s.} 0$.

C2 $\rho^2_{XY} : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{R}$ is continuous at the first canonical directions (Φ_1, Ψ_1).

C3 There exists a compact, self-adjoint and positive definite operator $\Gamma : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathcal{H}_1 \times \mathcal{H}_2$ such that $\rho^2_{XX}(u_1, u_2) = h_X\left\{\rho^2_{\Gamma_{11}}(u_1, u_2)\right\}$, $\rho^2_{YY}(v_1, v_2) = h_Y\left\{\rho^2_{\Gamma_{22}}(v_1, v_2)\right\}$ and

$$\rho^2_{XY}(u, v) = h\left\{\rho^2_{\Gamma}(u, v)\right\}, \tag{16}$$

where ρ_Γ, $\rho_{\Gamma_{11}}$ and $\rho_{\Gamma_{22}}$ are defined in (7) and (8). Furthermore, if $\widetilde{h} : [0, 1] \to [0, 1]$ stands for any of the functions h, h_X or h_Y, then $\widetilde{h}$ is a strictly increasing function such that $\widetilde{h}(0) = 0$ and $\lim_{x\to 1^-} \widetilde{h}(x) = 1$.

C4 (Φ_1, Ψ_1) exists and is unique up to change of sign. Moreover, there exists ρ_2 with $0 \le \rho_2 < \rho_1$ such that if $\rho_{XX}(u, \Phi_1) = \rho_{YY}(v, \Psi_1) = 0$ then $\rho^2_{XY}(u, v) \le \rho^2_2$.

The notion of convergence of the first canonical directions estimators to the first population canonical weights will be the convergence with respect to the measures of association that is analogous to the Γ-norm convergence defined in Leurgans et al. [35]. More precisely, given sequences $(u_n)_{n\in\mathbb{N}} \subseteq \mathcal{H}_1$, $(v_n)_{n\in\mathbb{N}} \subseteq \mathcal{H}_2$, we say that (u_n, v_n) converges to $(u, v) \in \mathcal{H}$ in association if $\rho^2_{XX}(u_n, u) \to 1$ and $\rho^2_{YY}(v_n, v) \to 1$. As noted in Leurgans et al. [35], this convergence means that the canonical variates obtained from (u_n, v_n) for a given random element (X, Y) behave as those obtained from (u, v). This kind of consistency is a desirable property to hold for the estimated canonical directions.

The following theorem shows the strong consistency of our estimators.

Theorem 4.1. *Let $(\hat{\Phi}_1, \hat{\Psi}_1)$ and $\hat{\rho}_1$ be the estimators defined in (11) and (12). Assume that the sequences $(p_n)_{n\in\mathbb{N}}$ and $(q_n)_{n\in\mathbb{N}}$ are such that $p_n \to \infty$ and $q_n \to \infty$. Then, under **C1** and **C2**, we have*

(i) $\hat{\rho}^2_1 \xrightarrow{a.s.} \rho^2_1$;

(ii) $\rho^2_{XY}(\hat{\Phi}_1, \hat{\Psi}_1) \xrightarrow{a.s.} \rho^2_1$;

(iii) *If in addition **C3** and **C4** hold, then $\rho^2_{XX}(\Phi_1, \hat{\Phi}_1) \xrightarrow{a.s.} 1$ and $\rho^2_{YY}(\Psi_1, \hat{\Psi}_1) \xrightarrow{a.s.} 1$.*

We include below some comments regarding assumptions **C2** to **C4**. Conditions under which **C1** holds and further comments on **C2** are relegated to the supplementary file available on-line.

Remark 4.1. Assume that the association measure satisfies (16) for a continuous increasing function h such that $h(0) = 0$ and that $\rho_1 > 0$, then **C2** holds. Effectively, from $\rho_1 > 0$ and the fact that h is increasing, we get that the ratio

$$\frac{\langle \Phi_1, \Gamma_{12}\Psi_1\rangle_{\mathcal{H}_1}}{\sqrt{\langle \Phi_1, \Gamma_{11}\Phi_1\rangle_{\mathcal{H}_1}\langle \Psi_1, \Gamma_{22}\Psi_1\rangle_{\mathcal{H}_2}}}$$

is positive, so $\langle \Phi_1, \Gamma_{11}\Phi_1\rangle_{\mathcal{H}_1} \neq 0$ and $\langle \Psi_1, \Gamma_{22}\Psi_1\rangle_{\mathcal{H}_2} \neq 0$. On the other hand, the compactness of Γ entails that the functions $g_1(u) = \langle u, \Gamma_{11}u\rangle_{\mathcal{H}_1}$, $g_2(u) = \langle v, \Gamma_{22}v\rangle_{\mathcal{H}_2}$ and $g_{12}(u, v) = \langle u, \Gamma_{12}v\rangle_{\mathcal{H}_1}$ are continuous functions which together with the fact that $g_1(\Phi_1) \neq 0$ and $g_2(\Psi_1) \neq 0$ entail that **C2** holds.

Note that when the kernel of Γ_{jj} is equal to $\{0\}$ for $j = 1, 2$, we have $g_1(\Phi_1) \neq 0$ and $g_2(\Psi_1) \neq 0$. Hence, if $h(t) = t$, **C3** implies **C2**. On the other hand, Lemma S.3.2 in the Supplementary file implies that **C2** holds if the association measure ρ is continuous with respect to the Prohorov distance at the bivariate distribution $P[\langle \Phi_1, X\rangle_{\mathcal{H}_1}, \langle \Psi_1, Y\rangle_{\mathcal{H}_2}]$.

It is worth mentioning that (16) in **C3** is analogous to assumption (iv) of Alfons et al. [2].

Remark 4.2. When we use Pearson's correlation $\rho = \rho_{\mathrm{CL}}$ in (4), $\rho_{XY} = \rho_{\Gamma_Z}$, with Γ_Z the covariance operator of Z, hence **C3** holds. Notice that, in this case, a necessary condition for a good definition of the canonical weights is that both random elements X and Y have finite second moments. This condition may be relaxed when the random element Z is elliptic and the association measure corresponds to one of those described in Section 2.1.

Let $Z = (X, Y)^\top \sim \mathcal{E}(\mu, \Gamma, \varphi)$ be a random element in $\mathcal{H}$ where $\mu \in \mathcal{H}$ and Γ is as in (6). Given $w = (w_1, w_2) \in \mathcal{H}_1 \times \mathcal{H}_2$, define $A_w : \mathcal{H} \to \mathbb{R}^2$ as $A_w(x, y) = (\langle w_1, x\rangle_{\mathcal{H}_1}, \langle w_2, y\rangle_{\mathcal{H}_2})^\top$. Using that $Z \sim \mathcal{E}(\mu, \Gamma, \varphi)$, we get that $\mathbf{z}_w = A_w(X, Y)$ has an elliptical distribution, $\mathbf{z}_w \sim \mathcal{E}_2(A_w\mu, A_w\Gamma A^*_w)$, where the (i, j)th element of $A_w\Gamma A^*_w$ equals $\langle w_i, \Gamma_{ij}w_j\rangle_{\mathcal{H}_i}$, for $i, j \in \{1, 2\}$. Hence, given an association measure ρ Fisher-consistent at the elliptical family, from the discussion given in Section 2.1 we conclude that

$$\rho^2_{XY}(u, v) = \frac{\langle u, \Gamma_{12}v\rangle^2_{\mathcal{H}_1}}{\langle u, \Gamma_{11}u\rangle_{\mathcal{H}_1}\langle v, \Gamma_{22}v\rangle_{\mathcal{H}_2}} = \rho^2_\Gamma(u, v). \tag{17}$$

Therefore, ρ satisfies (16) with $h(t) = t$. Moreover, we also have $\rho^2_{XX}(u_1, u_2) = \rho^2_{\Gamma_{11}}(u_1, u_2)$ and $\rho^2_{YY}(v_1, v_2) = \rho^2_{\Gamma_{22}}(v_1, v_2)$, that is, Fisher-consistent association measures for bivariate elliptic families satisfy **C3**.

Remark 4.3. Assumption **C4** is similar to assumption 3 in Leurgans et al. [35]. Consider the special situation of an elliptical random element $Z = (X, Y)^\top \sim \mathcal{E}(\mu, \Gamma, \varphi)$ where Γ is given by (6) and let ρ be an association measure Fisher-consistent for elliptical families. Assume that $\mathrm{E}\,\|Z\|^2_{\mathcal{H}} < \infty$, then, without loss of generality, we may assume that Γ is the covariance operator of Z. Define $R = \Gamma_{11}^{-1/2}\Gamma_{12}\Gamma_{22}^{-1/2}$, where $\Gamma_{11}^{-1/2}$ and $\Gamma_{22}^{-1/2}$ are the generalized inverses of the roots of Γ_{11} and Γ_{22} respectively. Assume that there exists only one orthonormal eigenfunction v_1 associated to the first eigenvalue ζ_1 of R^*R and denote as $u_1 = Rv_1/\sqrt{\lambda_1}$. When $Z \in L^2(0,1) \times L^2(0,1)$ and ρ is the Pearson correlation, He et al. [25] provide conditions ensuring the existence and proper definition of the canonical directions and correlations. Under these conditions, Theorem 4.8 of He et al. [25] entails that $(\Phi_1, \Psi_1) = (\Gamma_{11}^{-1/2}u_1/\|\Gamma_{11}^{-1/2}u_1\|_{\mathcal{H}_1}, \Gamma_{22}^{-1/2}v_1/\|\Gamma_{22}^{-1/2}v_1\|_{\mathcal{H}_2})$ is the unique maximizer in (4) except for sign change. Furthermore, if $\zeta_2 < \zeta_1$ where ζ_2 is the second eigenvalue of R^*R, then, given $u \in \mathcal{H}_1$ and $v \in \mathcal{H}_2$ such that $\rho_{XX}(u, \Phi_1) = \rho_{YY}(v, \Psi_1) = 0$, we have $\rho_{XY}(u, v) \le \rho_2 = \sqrt{\zeta_2}$ which shows that **C4** holds.

As mentioned in Remark 4.2, if $Z = (X, Y)^\top \sim \mathcal{E}(\mu, \Gamma, \varphi)$ is an elliptical random element, Fisher-consistent association measures for bivariate elliptic families, such as the coefficients $\rho_{\mathbf{v}_M}$ or ρ_{OGK} defined through (2), satisfy (16). Hence, we have $\rho_1(P) = h(\rho_{\Gamma,1})$, $\Phi_1(P) = \Phi_{\Gamma,1}$ and $\Psi_1(P) = \Psi_{\Gamma,1}$. Therefore, the above discussion implies that **C4** holds for these association measures, under the assumptions in Theorem 4.8 of He et al. [25].

On the other hand, for Gaussian processes, **C4** holds when considering the transformed Spearman or Kendall correlations, ρ_{SP} and ρ_{K}, under the assumptions of Theorem 4.8 of He et al. [25].

4.1. Fisher-consistency

Theorem 4.1 shows that the estimators defined in (11) and (12) are consistent for the population first canonical directions and the maximum canonical association given in (4) and (5). It is important to highlight that the quantities $\rho_1 = \rho_1(P)$ and $(\Phi_1, \Psi_1) = (\Phi_1(P), \Psi_1(P))$ depend on the chosen association measure ρ and we need to clarify what they represent. This section focusses on showing that the functionals $\rho_1(P)$ and $(\Phi_1(P), \Psi_1(P))$ have a simple interpretation, at least in some situations. In particular, our results hold for the elliptical families, even though they are not restricted to them.

When the measure of association satisfies **C3**, we have $\rho_1(P) = h(\rho_{\Gamma,1})$, $\Phi_1(P) = \Phi_{\Gamma,1}$ and $\Psi_1(P) = \Psi_{\Gamma,1}$. Hence, in this case, the estimated canonical directions defined in (11) are consistent for the first canonical weights associated to Γ. Furthermore, if $h(t) = t$, $\hat{\rho}_1$ provides a consistent estimator of the first canonical association associated to Γ. As mentioned above, if $Z \sim \mathcal{E}(\mu, \Gamma, \varphi)$ and has finite second moment, there exists a constant $c > 0$ such that $\Gamma = c\,\Gamma_{ZZ}$, where Γ_{ZZ} is the covariance operator of Z. Therefore, the canonical analysis done using the dispersion operator Γ or the covariance operator Γ_{ZZ} are identical.

We will see that for elliptic families, the functionals $\rho_1(P)$, $\Phi_1(P)$ and $\Psi_1(P)$ have a simple interpretation for some of the association measures described in Section 2.1. Let $Z = (X, Y)^\top \sim \mathcal{E}(\mu, \Gamma, \varphi)$ be a random element in $\mathcal{H}$ where $\mu \in \mathcal{H}$ and Γ is as in (6). From Remark 4.2, we have that **C3** holds with the scatter operator Γ, see (17), implying that $\rho_1(P)$, $\Phi_1(P)$ and $\Psi_1(P)$ are the first canonical association and directions associated to Γ respectively. Hence, Fisher-consistent association measures for bivariate elliptic families provide estimators of the quantities of interest. Examples of association measures Fisher-consistent at the elliptical distributions are, for instance, those defined from scatter matrices, i.e., the measures $\rho_{\mathbf{v}_M}$ or ρ_{OGK} defined through (2).

Finally, consider the family of Gaussian distributions and any association measure Fisher-consistent for normal bivariate vectors, such as the normalized Spearman or Kendall correlations ρ_{SP} and ρ_{K} given in (3). Then, using (17) we get that (16) is satisfied with $h(t) = t$ which implies that $\rho_1(P)$, $\Phi_1(P)$ and $\Psi_1(P)$ are the first canonical association and directions associated to Γ respectively.

5. Detection methods to identify influential observations

An important use of robust estimators is the detection of potential outliers. In this section, we describe two criteria to detect observations with a significant impact on the first canonical weight estimators. More precisely, 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 first canonical directions estimators. The first method is based on prediction, while the second is based on cross-prediction, both related to the functional canonical analysis described above.

To describe the first detection rule, let $(X, Y)^\top$ be a centred random element of $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$. Our detection method considers the orthogonal projections on the first canonical weights $\langle X, \Phi_1\rangle_{\mathcal{H}_1}\Phi_1$ and $\langle Y, \Psi_1\rangle_{\mathcal{H}_2}\Psi_1$ as predictors of X and Y respectively. More precisely, given a sample $\{(X_i, Y_i)^\top\}_{i=1}^n \subset \mathcal{H}$, let $\hat{\mu} = (\hat{\mu}_X, \hat{\mu}_Y)$ be robust location estimates computed from this sample, such as the spatial median, that is, $\hat{\mu}_X = \mathrm{argmin}_{\theta\in\mathcal{H}_1}\sum_{i=1}^n \left(\|X_i - \theta\|_{\mathcal{H}_1} - \|X_i\|_{\mathcal{H}_1}\right)/n$ and $\hat{\mu}_Y = \mathrm{argmin}_{\theta\in\mathcal{H}_2}\sum_{i=1}^n \left(\|Y_i - \theta\|_{\mathcal{H}_2} - \|Y_i\|_{\mathcal{H}_2}\right)/n$. Let $(\hat{\Phi}_1, \hat{\Psi}_1)$ be the robust estimates of the first canonical weights defined in Section 3.2 and computed as described in Section 3.4. Robust estimators are needed since, as in other settings, the detection methods based on the Pearson correlation may produce a masking effect that will not allow us to properly identify the influential observations.

Denote as $X_i^{(c)} = X_i - \hat{\mu}_X$ and $Y_i^{(c)} = Y_i - \hat{\mu}_Y$ the centred observations and let $\hat{X}_i^{(c)} = \langle X_i^{(c)}, \hat{\Phi}_1\rangle_{\mathcal{H}_1}\hat{\Phi}_1$ and $\hat{Y}_i^{(c)} = \langle Y_i^{(c)}, \hat{\Psi}_1\rangle_{\mathcal{H}_2}\hat{\Psi}_1$ be their predictors respectively. We expect that an influential or atypical observation will be poorly fitted leading to large values of at least one of the two squared residuals norms $R^2_{X,i} = \|X_i^{(c)} - \hat{X}_i^{(c)}\|^2_{\mathcal{H}_1}$ or $R^2_{Y,i} = \|Y_i^{(c)} - \hat{Y}_i^{(c)}\|^2_{\mathcal{H}_2}$.

As noticed in Boente and Salibián–Barrera [7], exploring the residuals norms may allow to detect abnormal points in the data. Taking into account that the distribution of the residuals squared norm is typically skewed to the right, we propose to flag an observation as atypical if its squared residual norm exceeds the upper whisker of a skewed-adjusted boxplot (see Hubert and Vandervieren [32]). More precisely, denote as $\mathcal{G}_X$ and $\mathcal{G}_Y$ the set of indices exceeding the upper whisker of the skewed-adjusted boxplot of the residuals $(R^2_{X,i})_{1\le i\le n}$ and $(R^2_{Y,i})_{1\le i\le n}$ respectively. The observations with indices in the sets $\mathcal{G}_X$, $\mathcal{G}_Y$ or $\mathcal{G}_X \cup \mathcal{G}_Y$ are considered as potential influential observations which, from now on, are called *outliers* as shorthand.

The detection rule based on cross-predictions is based on the following property that can be found in Yohai and García Ben [45]. Let $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{y} \in \mathbb{R}^q$ be centred random vectors and assume that $\mathrm{E}\|\mathbf{x}\|^2_{\mathbb{R}^p} < \infty$ and $\mathrm{E}\|\mathbf{y}\|^2_{\mathbb{R}^q} < \infty$. Given a random vector $\mathbf{z} \in \mathbb{R}^\ell$, denote as $\mathbf{y}^*_{\mathbf{z}}$ the best linear predictor for $\mathbf{y}$ based on $\mathbf{z}$, i.e., $\mathbf{y}^*_{\mathbf{z}} = \mathrm{E}(\mathbf{y}\mathbf{z}^\top)\mathrm{E}(\mathbf{z}\mathbf{z}^\top)^{-1}\mathbf{z}$. Then, when each component z_j of $\mathbf{z} = (z_1, \ldots, z_\ell)$ is a linear function of $\mathbf{x}$, the random vector $\mathbf{z}$ that minimizes the determinant of the matrix $\mathrm{E}(\mathbf{y} - \mathbf{y}^*_{\mathbf{z}})(\mathbf{y} - \mathbf{y}^*_{\mathbf{z}})^\top$ is given by the first ℓ canonical variables related to $\mathbf{x}$.

As in Section 3.4, given $\{\delta_i\}_{i\ge1}$ and $\{\eta_j\}_{j\ge1}$ orthonormal bases of $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively and a sample $\{(X_i, Y_i)^\top\}_{i=1}^n$ set $\mathbf{x}_i = \left(\langle X_i, \delta_1\rangle_{\mathcal{H}_1}, \ldots, \langle X_i, \delta_{p_n}\rangle_{\mathcal{H}_1}\right)^\top$ and $\mathbf{y}_i = (\langle Y_i, \eta_1\rangle_{\mathcal{H}_2}, \ldots, \langle Y_i, \eta_{q_n}\rangle_{\mathcal{H}_2})^\top$. Denote $\mathbf{x}_i^{(c)} = \mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\mathbf{x}}$ and $\mathbf{y}_i^{(c)} = \mathbf{y}_i - \hat{\boldsymbol{\mu}}_{\mathbf{y}}$, the centred observations where $\hat{\boldsymbol{\mu}}_{\mathbf{x}}$ and $\hat{\boldsymbol{\mu}}_{\mathbf{y}}$ are robust location estimates, such as the spatial median computed from the samples $\{\mathbf{x}_i\}_{1\le i\le n}$ and $\{\mathbf{y}_i\}_{1\le i\le n}$ respectively. Furthermore, let $\hat{\mathbf{a}}_1$ and $\hat{\mathbf{b}}_1$ be defined as in (14) and compute the sample of the centred first canonical variables $\hat{u}_i^{(c)} = \hat{\mathbf{a}}_1{}^\top\mathbf{x}_i^{(c)}$ and $\hat{v}_i^{(c)} = \hat{\mathbf{b}}_1{}^\top\mathbf{y}_i^{(c)}$. It is worth noticing that $\hat{u}_i^{(c)} = \langle X_i, \hat{\Phi}_1\rangle_{\mathcal{H}_1} - \hat{\mathbf{a}}_1{}^\top\hat{\boldsymbol{\mu}}_X$ provides an approximation for the centred canonical variate in the space $\mathcal{H}_1$ given by $\hat{U}_i^{(c)} = \langle X_i^{(c)}, \hat{\Phi}_1\rangle_{\mathcal{H}_1} = \langle X_i, \hat{\Phi}_1\rangle_{\mathcal{H}_1} - \langle\hat{\mu}_X, \hat{\Phi}_1\rangle_{\mathcal{H}_1}$ and similarly for $\hat{v}_i^{(c)}$. Using for each observation the centred canonical variates, we obtain the best robust linear predictors of $\mathbf{x}^{(c)}$ and $\mathbf{y}^{(c)}$ based on $\hat{v}^{(c)}$ and $\hat{u}^{(c)}$ respectively, denoted as $\hat{\mathbf{x}}^*_{\hat{v}}$ and $\hat{\mathbf{y}}^*_{\hat{u}}$. As before, one expects that an influential observation will be poorly predicted causing large values of at least one of the two squared residuals norms $r^2_{X,i} = \|\mathbf{x}_i^{(c)} - \hat{\mathbf{x}}^*_{\hat{v},i}\|^2_{\mathbb{R}^p}$ and $r^2_{Y,i} = \|\mathbf{y}_i^{(c)} - \hat{\mathbf{y}}^*_{\hat{u},i}\|^2_{\mathbb{R}^q}$. Three different sets of indices are used to identify the possible atypical data. As above, $\mathcal{G}_X$ and $\mathcal{G}_Y$ indicate the sets of indices exceeding the upper whisker of the skewed-adjusted boxplot of the residuals $(r^2_{X,i})_{1\le i\le n}$ and $(r^2_{Y,i})_{1\le i\le n}$ respectively, while $\mathcal{G}_{X+Y}$ stands for those indices exceeding the upper whisker of the skewed-adjusted boxplot of the sample $(r^2_{X,i} + r^2_{Y,i})_{1\le i\le n}$. Finally, any observation with index in the sets $\mathcal{G}_X$, $\mathcal{G}_Y$, $\mathcal{G}_X \cup \mathcal{G}_Y$ or $\mathcal{G}_{X+Y}$ is considered as a potential atypical data.

It is worth noticing that a detection rule based on the bagplot of $(\hat{U}_i^{(c)}, \hat{V}_i^{(c)})$ may also be considered as a diagnostic tool. We refer to Alvarez [3] for readers interested on the performance of this detection method.

6. Monte Carlo study

In this section, we numerically explore the finite sample behaviour of the proposed estimators for different association measures and different choices for the approximating subspaces when the Hilbert spaces are L^2-spaces. More precisely, we report the results of a Monte Carlo study designed to compare the performance, for Gaussian and contaminated data, of the first canonical association and direction estimators defined in (11) and (12) when using the Pearson correlation and two robust association measures. In all cases, we performed $NR = 1000$ replications.

6.1. The estimators

As mentioned above, the estimators defined in (11) depend both in the association measure to be maximized and in the bases generating the approximating spaces. Beyond the Pearson correlation ρ_{CL}, we report here the results obtained using the association measure defined through the M-scatter matrix $\rho_{\mathbf{V}_\mathrm{M}}$ defined in (2), where $\mathbf{V}_\mathrm{M}$ is computed using Huber's score function with tuning constant $k_1 = (\chi^2_{2,0.9})^{1/2}$, and the normalized Spearman coefficient ρ_{SP} given in (3). The results for other association measures can be seen in Alvarez [3].

Two different sieve bases are considered: the cubic B-spline basis as fixed basis and the basis of functional principal directions as an adaptive one. The elements of the B-spline basis are orthonormalized before applying the algorithm to compute the estimators. On the other hand, when ρ is the Pearson correlation ρ_{CL}, the principal direction basis on each space is chosen as the eigenfunctions of the sample covariance operators of X and Y respectively. For the robust association measures, robust principal direction estimators are considered. Taking into account that, when performing cross-validation, the principal directions need to be computed each time an observation is removed, we need to choose a robust and fast procedure to compute the principal direction estimators. The spherical principal directions defined in Locantore et al. [36] and studied in Gervini [21] achieve this goal, since they provide a simple and fast method to obtain estimators of the functional principal directions.

The canonical direction estimators were computed as described in Section 3.4 using the alternate GRID algorithm implemented through the function `ccaGrid`. It is worth mentioning that an algorithm using the centred data as possible directions may also be considered to estimate the first canonical directions, following the ideas used by Croux and Ruiz-Gazen [13] for the estimation of the first principal component. However, as shown in Alvarez [3], the GRID algorithm leads to better results. For that reason, we omit the results obtained using the centred data as candidates and we refer to Alvarez [3] for further discussions.

To select the dimension of the approximating spaces, we use the criterion defined in Section 3.3, where for computational simplicity, we have only considered the same dimensions in both spaces, that is, $\mathcal{R} = \{(p, p) : p \in \mathcal{R}^\star\}$, so $\mathbf{r}_n$ equals

(p, p) in (13). The set $\mathcal{R}^\star$ is taken as $\mathcal{R}^\star = \{3, \ldots, 11\}$ for both bases. It is worth noticing that the possible values of the dimension p start in 3 when using B-splines since we are using cubic splines. On the other hand, for the simulation model to be described in Section 6.2, the eigenvalues related to the first three principal directions are equal, so the principal directions are not uniquely defined for $p < 3$. Once the value $\widehat{\mathbf{r}} = (\hat{p}, \hat{p})$ in (13) is obtained, the estimators of the canonical weights are computed using the algorithm described below, leading to the first canonical weights estimators $(\hat{\Phi}_1, \hat{\Psi}_1)$ and the maximum canonical association estimator $\hat{\rho}_1$. We also computed the estimator $\widetilde{\rho}_1 = \widetilde{\rho}_{\hat{p},1}$ defined as $\widetilde{\rho}_1 = \sqrt{RCV_{\widehat{\mathbf{r}}}}$.

6.2. Simulation settings

Our simulation model is similar to the one considered in He et al. [26]. For each replication, we generate independent samples $\{(X_i, Y_i)^\top\}_{i=1}^n \subset \mathcal{H}_1 \times \mathcal{H}_2$ of size $n = 100$ with $\mathcal{H}_j = L^2[0, 50]$. The processes were observed over an equispaced grid of 50 points $t_j, j = 1, \ldots, 50$. Hence, the inner products $\langle X_i, u\rangle_{\mathcal{H}_1}$ and $\langle Y_i, v\rangle_{\mathcal{H}_2}$ were approximated as sums over the design points $\{t_j\}_{1\le j\le 50}$.

The clean data sets, denoted C_0, were generated with the same distribution as the Gaussian random element $(X, Y)^\top \in \mathcal{H}_1 \times \mathcal{H}_2$, given by $X(t) = \sum_{j=1}^m \xi_j f_j(t)$ and $Y(t) = \sum_{j=1}^m \zeta_j f_j(t)$, where $\{f_j\}_{j\ge1}$ is the Fourier basis of $L^2[0, 50]$ and $m = 21$. The scores $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)^\top$ and $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_m)^\top$ are m-dimensional normally distributed random vectors, $(\boldsymbol{\xi}^\top, \boldsymbol{\zeta}^\top)^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^\top & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

with $\boldsymbol{\Sigma}_{22} = \boldsymbol{\Sigma}_{11} = 10\,\mathrm{diag}(1, 1, 1, 0.75, \ldots, 0.75^{m-3})$ and $\boldsymbol{\Sigma}_{12} = \mathrm{diag}(7, 3, 1, 0, \ldots, 0)$.

Taking into account that the process (X, Y) is Gaussian and that all the considered association measures are Fisher-consistent at the bivariate normal distribution, the target quantities to be estimated do not depend on the selected association measure and are equal to the canonical weights and correlations defined in He et al. [24]. Hence, they will be simply denoted as ρ_ℓ and (Φ_ℓ, Ψ_ℓ). For the process described above, we have $\rho_1 = 0.7,\ \rho_2 = 0.3,\ \rho_3 = 0.1$ and $\rho_\ell = 0$ if $\ell > 3$, whereas the canonical weights are $\Phi_\ell(t) = \Psi_\ell(t) = f_\ell(t)$, for $\ell = 1, 2, 3$.

Besides the central Gaussian model, we have considered two contaminated situations that can be described as follows

C_1: $(X_i, Y_i)^\top$ are i.i.d. with the same distribution as $(1 - B)(X, Y)^\top + BW(f_2, f_2)^\top$, where $B \sim \mathcal{B}(1, 0.1)$, $W \sim \mathcal{N}(25, 1)$ and $W, B, (X, Y)^\top$ are all independent of each other. This contamination corresponds to a strong contamination in the direction of the second canonical direction of $(X, Y)^\top$.

C_2: $(X_i, Y_i)^\top$ are i.i.d. and the process X_i and Y_i are such that

$$X_i \sim (1 - B)X + B\left(\xi_1 f_1 + W\frac{f_3 + f_4}{\sqrt{2}} + 0.1\,\xi_3 f_3 + 0.1\,\xi_4 f_4 + \sum_{j=5}^{21} \xi_j f_j\right)$$

$$Y_i \sim (1 - B)Y + B\left(\zeta_1 f_1 + W\frac{f_3 + f_4}{\sqrt{2}} + 0.1\,\zeta_3 f_3 + 0.1\,\zeta_4 f_4 + \sum_{j=5}^{21} \zeta_j f_j\right)$$

where $B \sim \mathcal{B}(1, 0.1), W \sim \mathcal{N}(25, 0.01)$ independent of B, X and Y and $(\boldsymbol{\xi}^\top, \boldsymbol{\zeta}^\top)^\top \sim \mathcal{N}(0, \boldsymbol{\Sigma})$. This contamination corresponds to a strong contamination in the direction of a linear combination of the third and fourth canonical weights of $(X, Y)^\top$.

It is worth noticing that, when ρ is the Pearson correlation, $\Phi_\ell(P) = \Psi_\ell(P) = f_\ell$ for $\ell = 1, 2, 3$ under C_0, but not necessarily for the contaminated distributions. More precisely, under C_1 the order between f_1 and f_2 is reversed when using $\rho = \rho_{\mathrm{CL}}$, so the first canonical directions are $\Phi_1(P) = \Psi_1(P) = f_2$ for this association measure, while f_1 corresponds to the second canonical weights.

6.3. Simulation results

For each situation, we evaluate the performance of the first canonical directions and the maximum canonical association. To compare the performance of the first canonical weight estimators $(\hat{\Phi}_1, \hat{\Psi}_1)$ of (Φ_1, Ψ_1), we compute

- a global goodness of fit measure, denoted as MISE, considered also in He et al. [26], which is the average over replications of $\|\hat{\Phi}_1 - \Phi_1\|^2 + \|\hat{\Psi}_1 - \Psi_1\|^2$.
- the average over replications of the absolute Pearson correlation between the canonical variates computed only for the non-atypical data, that is, the average of

$$\hat{\rho}_X = \hat{\rho}_{\mathrm{CL},XX,\mathrm{CLEAN}}(\Phi_1, \hat{\Phi}_1) = \frac{\left|\sum_{i=1}^n (1 - B_i)U_i\hat{U}_i\right|}{\sqrt{\sum_{i=1}^n (1 - B_i)\hat{U}_i^2\ \sum_{i=1}^n (1 - B_i)U_i^2}},$$

  where $U_i = \langle X_i, \Phi_1\rangle$ and $\hat{U}_i = \langle X_i, \hat{\Phi}_1\rangle$. This measure provides a way to quantify how the proposals fit the good observations. A similar measure was computed for $\hat{\Psi}_1$.

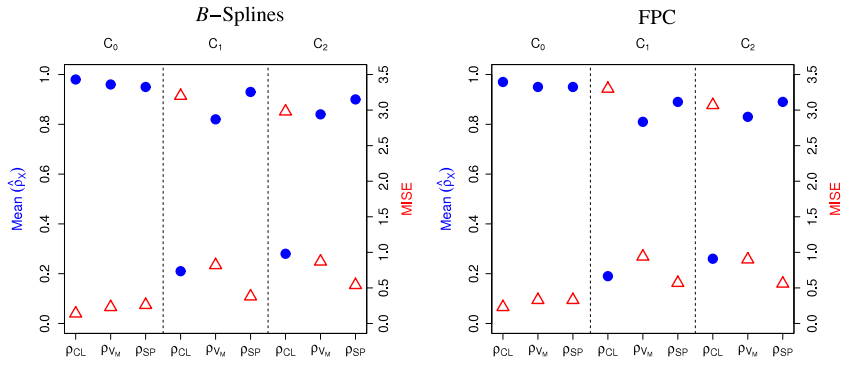

Fig. 2. MISE of the canonical weight estimators (in red triangles) and average of $\hat{\rho}_X = \hat{\rho}_{\text{CL},XX,\text{CLEAN}}(\Phi_1, \hat{\Phi}_1)$ (in blue solid points), when using the Pearson correlation ρ_{CL}, the association measure $\rho_{\mathbf{V}_M}$ defined in (2) and the normalized Spearman coefficient ρ_{SP} given in (3). The considered sieve bases are the cubic *B*-spline basis and the basis of estimated functional principal directions, labelled *B*-Splines and FPC, respectively.

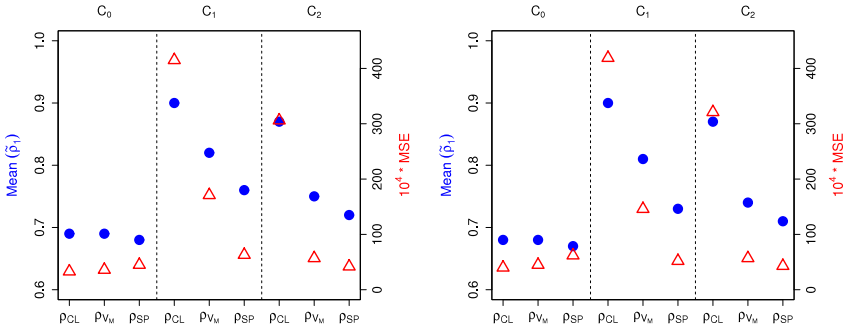

Fig. 3. Mean (in blue solid points) and mean square error (MSE) (in red triangles) of the estimators $\widetilde{\rho}_1$, when using the Pearson correlation ρ_{CL}, the association measure $\rho_{\mathbf{V}_M}$ defined in (2) and the normalized Spearman coefficient ρ_{SP} given in (3). The considered sieve bases are the cubic *B*-spline basis and the basis of estimated functional principal directions, labelled *B*-Splines and FPC, respectively.

Taking into account that the results obtained for $\hat{\Psi}_1$ are similar to those obtained for $\hat{\Phi}_1$, we only report the summary measures for the first component of the canonical direction estimators. More precisely, Table 1 reports the MISE and averages over replications of $\hat{\rho}_X$ for different association measures and bases. We also report the median of the dimension $\hat{p}$ obtained by cross-validation to illustrate how the contaminations affect the dimension of the chosen linear spaces.

The better performance of the estimators computed using a robust association measure is better illustrated in Figs. 2 and 3, where, for simplicity, the functional principal direction basis is denoted FPC. More precisely, Fig. 2 plots in red triangles and blue solid circles the MISE and the average over replications of $\hat{\rho}_X = \hat{\rho}_{\text{CL},XX,\text{CLEAN}}(\Phi_1, \hat{\Phi}_1)$ respectively.

Note that for clean data the average of $\hat{\rho}_X$ is close to 1, while the values of the MISE are small, as expected. This fact is more evident in Fig. 2 which reveals that under C_0 all the procedures lead to similar results. On the other hand, for contaminated samples the estimators based on Pearson's correlation are strongly affected by the presence of outliers. Indeed, in this case Fig. 2 shows that the average values of $\hat{\rho}_{\text{CL},XX,\text{CLEAN}}(\Phi_1, \hat{\Phi}_1)$ move away from 1 approaching 0 and the MISE grows taking values very close to 4, which is the maximum possible value. On the other hand, both robust estimators give more resistant results. However, the proposal based on the Spearman coefficient stands out, since it has the lowest MISE value and the average over replications of $\hat{\rho}_{\text{CL},XX,\text{CLEAN}}(\Phi_1, \hat{\Phi}_1)$ is the one closer to 1 for the considered bases in both contaminated scenarios.

To summarize the behaviour of a given estimator $\hat{\rho}_1$ of the maximum canonical association ρ_1, we compute the mean over replications of the obtained values as well as the mean square error (MSE). Table 2 reports the mean and mean square error (multiplied by 10^4) of the estimators $\hat{\rho}_1$ and $\widetilde{\rho}_1 = \widetilde{\rho}_{\hat{p},1}$. On the other hand, Fig. 3 shows the mean and MSE of the estimators $\widetilde{\rho}_1$ in blue solid points and red triangles respectively revealing the sensitivity of the classical procedure based on the Pearson correlation that increases its mean square error due to its increased bias. The same conclusion can be obtained from the estimator $\hat{\rho}_1$ reported in Table 2. Indeed, for clean data the mean values of the estimates $\hat{\rho}_1$ are close to ρ_1 and the MSE are low and very similar to each other. However, it should be noted that the estimates $\hat{\rho}_1$ present a positive bias in all cases, while the estimator $\widetilde{\rho}_{\hat{p},1}$ is slightly negatively biased. In this scenario, the absolute bias of $\widetilde{\rho}_{\hat{p},1}$ is lower than that of $\hat{\rho}_1$, leading to MSE values which are in general smaller than those of $\hat{\rho}_1$.

For contaminated samples, both estimators $\hat{\rho}_1$ and $\widetilde{\rho}_{\hat{p},1}$ are strongly affected, moving away from ρ_1, when the Pearson correlation is used. The best procedure is the one based on the normalized Spearman coefficient giving the smallest mean square errors and only a small increase on the estimators bias. As for uncontaminated samples, the MSE of $\widetilde{\rho}_{\hat{p},1}$ is somewhat

Table 1
MISE and mean over replications of $\hat{\rho}_X = \hat{\rho}_{\text{CL},XX,\text{CLEAN}}(\Phi_1, \hat{\Phi}_1)$ for different contamination settings, when using the Pearson correlation ρ_{CL}, the association measure $\rho_{\mathbf{v}_M}$ defined in (2) and the normalized Spearman coefficient ρ_{SP} given in (3). The median of the dimension $\hat{p}$ obtained by cross-validation is also reported. The considered sieve bases are the cubic *B*-spline and the estimated functional principal directions bases.

	C_0			C_1			C_2		
ρ	$\hat{p}$	$\hat{\rho}_X$	MISE	$\hat{p}$	$\hat{\rho}_X$	MISE	$\hat{p}$	$\hat{\rho}_X$	MISE
	B-splines								
ρ_{CL}	3	0.98	0.14	3	0.21	3.20	5	0.28	2.98
$\rho_{\mathbf{v}_M}$	3	0.96	0.23	4	0.82	0.82	5	0.84	0.87
ρ_{SP}	4	0.95	0.26	4	0.93	0.38	5	0.90	0.54
	Functional principal direction basis								
ρ_{CL}	4	0.97	0.23	3	0.19	3.30	4	0.26	3.07
$\rho_{\mathbf{v}_M}$	5	0.95	0.33	5	0.81	0.94	5	0.83	0.90
ρ_{SP}	4	0.95	0.33	4	0.89	0.57	5	0.89	0.56

Table 2
Mean and mean square error (multiplied by 10^4) of the maximum canonical association estimators $\hat{\rho}_1$ and $\widetilde{\rho}_1 = \widetilde{\rho}_{\hat{p},1}$ for different contamination settings, when using the Pearson correlation ρ_{CL}, the association measure $\rho_{\mathbf{v}_M}$ defined in (2) and the normalized Spearman coefficient ρ_{SP} given in (3). The median of the dimension $\hat{p}$ obtained by cross-validation is also reported. The considered sieve bases are the cubic *B*-spline and the estimated functional principal directions bases.

	C_0					C_1					C_2				
		Mean		$10^4\times$MSE			Mean		$10^4\times$MSE			Mean		$10^4\times$MSE	
ρ	$\hat{p}$	$\hat{\rho}_1$	$\widetilde{\rho}_1$	$\hat{\rho}_1$	$\widetilde{\rho}_1$	$\hat{p}$	$\hat{\rho}_1$	$\widetilde{\rho}_1$	$\hat{\rho}_1$	$\widetilde{\rho}_1$	$\hat{p}$	$\hat{\rho}_1$	$\widetilde{\rho}_1$	$\hat{\rho}_1$	$\widetilde{\rho}_1$
	B-splines														
ρ_{CL}	3	0.73	0.69	34	33	3	0.91	0.90	437	415	5	0.89	0.87	373	306
$\rho_{\mathbf{v}_M}$	3	0.74	0.69	45	36	4	0.86	0.82	288	171	5	0.82	0.75	177	57
ρ_{SP}	4	0.75	0.68	53	45	4	0.80	0.76	124	63	5	0.79	0.72	104	42
	Functional principal direction basis														
ρ_{CL}	4	0.73	0.68	36	40	3	0.91	0.90	435	419	4	0.89	0.87	356	321
$\rho_{\mathbf{v}_M}$	5	0.74	0.68	48	45	5	0.86	0.81	287	146	5	0.82	0.74	167	57
ρ_{SP}	4	0.75	0.67	55	62	4	0.79	0.73	116	52	5	0.78	0.71	92	43

lower than the one obtained with $\hat{\rho}_1$, in all cases. We can explain this result by the nature of the introduced contamination, which tends to increase the values of the correlation estimators. Recall that for clean samples $\widetilde{\rho}_{\hat{p},1}$ has a negative bias, while $\hat{\rho}_1$ has a positive bias. In this case, as a result of the considered contamination, the estimators $\widetilde{\rho}_{\hat{p},1}$ and $\hat{\rho}_1$ give larger values resulting in estimators $\widetilde{\rho}_{\hat{p},1}$ with smaller bias than $\hat{\rho}_1$.

We also study the performance of the two methods proposed in Section 5 to detect influential observations. For each detection method, Tables 3 and 4 report the average sensitivity and specificity over the 1000 replications. Recall that sensitivity is the proportion of actual atypical trajectories that are correctly flagged as such, while specificity is the proportion of non-atypical curves correctly identified as not atypical. An ideal method will simultaneously maintain high sensitivity and specificity. As described in Section 5, the detection rule based on prediction flags as outliers the observations with indices in the sets $\mathcal{G}_X$, $\mathcal{G}_Y$ or $\mathcal{G}_X \cup \mathcal{G}_Y$ defined therein, leading to the detection methods denoted AT_X, AT_Y and $\text{AT}_{X\text{o}Y}$ respectively in Table 3. On the other hand, the cross-prediction method defines three sets of indices $\mathcal{G}_X$, $\mathcal{G}_Y$ and $\mathcal{G}_{X+Y}$ based on the skewed-adjusted boxplot of the cross-prediction residuals norms. Any observation with index in the sets $\mathcal{G}_X$, $\mathcal{G}_Y$, $\mathcal{G}_X \cup \mathcal{G}_Y$ or $\mathcal{G}_{X+Y}$ is considered as a potential atypical data, leading to the detection rules labelled by AT_X, AT_Y, $\text{AT}_{X\text{o}Y}$ and AT_{X+Y} respectively in Table 4.

As shown in Tables 3 and 4, all detection rules have high specificity levels under all scenarios, even though the rules based on $\text{AT}_{X\text{o}Y}$ identify more data as atypical, resulting in a slightly lower specificity but a higher sensitivity. This behaviour is more clearly observed in Figures S.1 and S.2 in the supplementary file.

For the method based on prediction, the detection rule $\text{AT}_{X\text{o}Y}$ seems to be the one with the best performance, since the decrease in specificity with respect to those labelled AT_X or AT_Y is small compared to the increase in sensitivity obtained. Analogous results are observed in the method based on cross-predictions when using the detection rules AT_{X+Y} and $\text{AT}_{X\text{o}Y}$. In almost all cases, the detection rule $\text{AT}_{X\text{o}Y}$ corresponding to the method based on prediction overcome in sensitivity the rules based on cross-prediction.

For the scenarios considered here, the detection rule $\text{AT}_{X\text{o}Y}$ based on prediction, using the functional principal direction basis and the transformed Spearman correlation leads to specificity results similar to those obtained with *B*-splines with only a slightly smaller sensitivity. Taking into account that the functional principal direction basis is adaptive, the practitioner may prefer to use as a diagnostic tool the rule $\text{AT}_{X\text{o}Y}$ based on prediction, the association measure ρ_{SP} and the functional principal direction basis, when considering a real data set. However, taking into account the better overall performance of the detection rule based on *B*-splines and the fact that the *B*-spline basis is sufficiently rich to represent most data sets, we still recommend using this basis for diagnostic purposes.

Table 3
Specificity and sensitivity for the detection rule based on the prediction errors squared norm $R^2_{X,i} = \|X_i^{(c)} - \hat{X}_i^{(c)}\|^2_{\mathcal{H}_1}$ or $R^2_{Y,i} = \|Y_i^{(c)} - \hat{Y}_i^{(c)}\|^2_{\mathcal{H}_2}$. The estimators are obtained using the association measure $\rho_{\mathbf{v}_M}$ defined in (2) or the normalized Spearman coefficient ρ_{SP} given in (3) combined with two possible sieve bases: the cubic B-spline basis and the basis of estimated functional principal directions.

ρ	C_0			$C_{1,0.1}$						$C_{2,0.1}$					
	Specificity			Specificity			Sensitivity			Specificity			Sensitivity		
	AT_X	AT_Y	AT_{XoY}	AT_X	AT_Y	AT_{XoY}	AT_X	AT_Y	AT_{XoY}	AT_X	AT_Y	AT_{XoY}	AT_X	AT_Y	AT_{XoY}
	B-splines														
$\rho_{\mathbf{v}_M}$	0.991	0.991	0.982	0.998	0.998	0.997	0.887	0.888	0.917	0.999	0.999	0.998	0.933	0.929	0.966
ρ_{SP}	0.990	0.991	0.981	0.999	0.999	0.998	0.964	0.978	0.994	0.999	0.999	0.998	0.984	0.978	0.997
	Functional principal direction basis														
$\rho_{\mathbf{v}_M}$	0.990	0.991	0.982	0.999	0.999	0.998	0.887	0.884	0.920	0.999	0.999	0.998	0.914	0.911	0.953
ρ_{SP}	0.991	0.991	0.982	0.999	0.999	0.998	0.949	0.956	0.973	0.999	0.999	0.998	0.980	0.976	0.995

Table 4
Specificity and sensitivity for the detection rule based on the cross-prediction errors squared norm $r^2_{X,i} = \|\mathbf{x}_i^{(c)} - \hat{\mathbf{x}}^*_{\hat{v},i}\|^2_{\mathbb{R}^p}$ and $r^2_{Y,i} = \|\mathbf{y}_i^{(c)} - \hat{\mathbf{y}}^*_{\hat{u},i}\|^2_{\mathbb{R}^q}$. The estimators are obtained using the association measure $\rho_{\mathbf{v}_M}$ defined in (2) or the normalized Spearman coefficient ρ_{SP} given in (3) combined with two possible sieve bases: the cubic B-spline basis and the basis of estimated functional principal directions.

ρ	C_0				$C_{1,0.1}$								$C_{2,0.1}$							
	Specificity				Specificity				Sensitivity				Specificity				Sensitivity			
	AT_X	AT_Y	AT_{XoY}	AT_{X+Y}	AT_X	AT_Y	AT_{XoY}	AT_{X+Y}	AT_X	AT_Y	AT_{XoY}	AT_{X+Y}	AT_X	AT_Y	AT_{XoY}	AT_{X+Y}	AT_X	AT_Y	AT_{XoY}	AT_{X+Y}
	B-splines																			
$\rho_{\mathbf{v}_M}$	0.992	0.992	0.984	0.991	0.999	0.999	0.998	0.999	0.812	0.815	0.869	0.883	0.999	0.999	0.998	0.999	0.834	0.814	0.907	0.912
ρ_{SP}	0.992	0.992	0.984	0.991	0.999	0.999	0.998	0.999	0.931	0.932	0.975	0.978	0.999	0.999	0.998	0.999	0.858	0.862	0.942	0.953
	Functional principal direction basis																			
$\rho_{\mathbf{v}_M}$	0.992	0.992	0.983	0.989	0.998	0.998	0.997	0.998	0.748	0.748	0.821	0.821	0.999	0.999	0.998	0.999	0.890	0.883	0.943	0.942
ρ_{SP}	0.991	0.991	0.983	0.990	0.999	0.999	0.997	0.999	0.870	0.873	0.929	0.919	0.999	0.999	0.998	0.999	0.959	0.957	0.988	0.989

7. Example: the writing data set

To illustrate the performance of the proposed first canonical directions estimators and of the atypical data detection rules, we consider the writing data set described in the Introduction. We only analyse the 186 data related to the speed of the pen on the horizontal and vertical axis when writing the letter "e" denoted $X(t)$ and $Y(t)$ respectively.

To identify potential atypical observations, we use as detection rules those that turn out to be the most effective ones in the simulation study described in Section 6. Table 5 reports the indices corresponding to observations detected as outliers/influential by the prediction and cross-prediction methods using the detection rules labelled AT_{XoY} and AT_{X+Y} respectively. The estimates of the first canonical weights were computed as described in Section 6.1, using the normalized Spearman coefficient ρ_{SP} defined in (3) combined with B-splines and with the functional principal direction basis. The dimension of the approximating spaces was selected by the cross-validation criterion given in (13) with $\mathcal{R} = \{(p, p),\ 3 \le p \le 11\}$, when using B-splines and $\mathcal{R} = \{(p, p),\ 1 \le p \le 11\}$ for the functional principal direction basis. When the maximum was attained at 11, the possible values of p were enlarged up to 20. For the B-spline basis, the pair satisfying (13) has coordinates $\hat{p} = 8$, while for the functional principal direction basis the maximum is attained at $\hat{p} = 5$. It is worth noticing that for the B-spline basis, when $5 \le p \le 8$, the values of $RCV_{(p,p)}$ are very close to the maximum $RCV_{(\hat{p},\hat{p})}$ (see Figure S.3 in the supplementary file available on-line). In order to show that, for B-splines, with a smaller dimension of the approximating spaces we can also obtain reliable results, Table 5 reports the results corresponding to dimension $p = 5$. Observations 33, 38, 139 and 175 are detected as atypical or influential curves for all bases and detection methods. On the other hand, when using the functional principal direction basis, the detection rule based on predictions detects as atypical some additional observations.

In order to evaluate the influence of the potential atypical data on the estimators based on the Pearson correlation, the classical estimates of the canonical weights were computed after removing the detected atypical observations. More precisely, when using B-splines, the observations with indices in $\mathcal{I}_1 = \{33, 38, 139, 175\}$ are excluded from the analysis. On the other hand, when projecting on the functional principal direction basis, taking into account that the observations in $\mathcal{I}_1$ are also detected by the cross-prediction method, we compute the estimators based on the data set without the observations in $\mathcal{I}_1$ and those obtained omitting the trajectories with indices in $\mathcal{I}_2 = \{7, 33, 38, 113, 139, 154, 175, 137, 140\}$. We denote as $\hat{\Phi}_{CL,1}^{-\mathcal{I}_j}$ the classical first canonical weight estimator in the X space obtained after removing the observations indexed in $\mathcal{I}_j$, while $\hat{\Phi}_{CL,1}$ or $\hat{\Phi}_{SP,1}$ stand for the estimators computed with the whole data set using the Pearson correlation or the normalized Spearman coefficient respectively. The results for the maximal canonical association are reported in the supplementary file, while the results obtained for the first canonical weight associated to Y are omitted, since they are similar to those corresponding to X.

Table 5
Atypical observations detected by the prediction and cross-prediction methods using normalized Spearman coefficient combined with the cubic B-spline basis or the basis of estimated functional principal directions. The dimension of the approximating subspaces is also reported.

B-splines		
p	Detection rule	Index of the detected observation
5	Prediction method $\text{AT}_{X\circ Y}$	33, 38, 139, 175
5	Cross-prediction method AT_{X+Y}	33, 38, 139, 175
Functional principal direction basis		
$\hat{p}$	Detection rule	Index of the detected observation
5	Prediction method $\text{AT}_{X\circ Y}$	7, 33, 38, 113, 139, 154, 175, 137, 140
5	Cross-prediction method AT_{X+Y}	33, 38, 139, 175

Table 6
Association between canonical variables and absolute cosine of the angle between the robust and classical first canonical weight estimates computed with B-splines or with the functional principal direction basis using the Pearson correlation and the transformed Spearman coefficient. We label as $-\mathcal{I}_j$ the results obtained when the classical estimator is computed after removing the observations indexed in $\mathcal{I}_j$ with $\mathcal{I}_1 = \{33, 38, 139, 175\}$ and $\mathcal{I}_2 = \{7, 33, 38, 113, 139, 154, 175, 137, 140\}$.

B-splines					
$\hat{\rho}_{\text{SP,CL}}$	$\hat{\rho}_{\text{SP,CL}}^{-\mathcal{I}_1}$		$\hat{c}_{\text{SP,CL}}$	$\hat{c}_{\text{SP,CL}}^{-\mathcal{I}_1}$	
0.71	1.00		0.09	0.99	
Functional principal direction basis					
$\hat{\rho}_{\text{SP,CL}}$	$\hat{\rho}_{\text{SP,CL}}^{-\mathcal{I}_1}$	$\hat{\rho}_{\text{SP,CL}}^{-\mathcal{I}_2}$	$\hat{c}_{\text{SP,CL}}$	$\hat{c}_{\text{SP,CL}}^{-\mathcal{I}_1}$	$\hat{c}_{\text{SP,CL}}^{-\mathcal{I}_2}$
0.95	0.97	0.97	0.79	0.86	0.85

Table 6 gives some summary measures that illustrate the behaviour of the first canonical directions estimates of X. To compare the performance of the classical estimators with those obtained using the normalized Spearman coefficient, we report the absolute cosine of the angle between the robust and the classical first canonical weight estimates. As mentioned above, the estimators based on the Pearson correlation are computed with the complete sample and after removing the observations detected as atypical. We denote as $\hat{c}_{\text{SP,CL}}$ the absolute cosine of the angle between $\hat{\Phi}_{\text{SP},1}$ and $\hat{\Phi}_{\text{CL},1}$ and as $\hat{c}_{\text{SP,CL}}^{-\mathcal{I}_j}$ that of the angle between $\hat{\Phi}_{\text{SP},1}$ and $\hat{\Phi}_{\text{CL},1}^{-\mathcal{I}_j}$. It is worth noticing that we cannot compute the measure $\hat{\rho}_{\text{CL},XX,\text{CLEAN}}$ defined in Section 6.3, since the abnormal data were not artificially introduced but detected by our diagnostic rules. For that reason, taking into account that the Spearman coefficient provides a reliable measure when atypical data arise in the sample, we report the empirical normalized Spearman correlation between the robust and classical canonical variables corresponding to such canonical directions. More precisely, we report the values $\hat{\rho}_{\text{SP,CL}} = \rho_{\text{SP}}(P_n[\langle\hat{\Phi}_{\text{SP},1}, X\rangle, \langle\hat{\Phi}_{\text{CL},1}, X\rangle])$ and $\hat{\rho}_{\text{SP,CL}}^{-\mathcal{I}_j} = \rho_{\text{SP}}(P_n[\langle\hat{\Phi}_{\text{SP},1}, X\rangle, \langle\hat{\Phi}_{\text{CL},1}^{-\mathcal{I}_j}, X\rangle])$.

The obtained results show that, when considering the estimates based on B-splines and computed with the complete data set, the robust and classical canonical weight estimators are far from each other, since the association measure $\hat{\rho}_{\text{SP,CL}}$ is far from 1 and the absolute cosine $\hat{c}_{\text{SP,CL}}$ is close to 0 implying that the directions are almost orthogonal. On the other hand, the canonical variable $\langle\hat{\Phi}_{\text{SP},1}, X\rangle$ and the classical one obtained when the potential *outliers* are removed, $\langle\hat{\Phi}_{\text{CL},1}^{-\mathcal{I}_1}, X\rangle$, attain the largest possible empirical association and the absolute cosine $\hat{c}_{\text{SP,CL}}^{-\mathcal{I}_1}$ is also close to 1. It is also worth noticing that, when the whole data set is considered and the estimators are computed using the Pearson coefficient and B-splines, the maximum of $RCV_{(p,p)}$ is attained at $\hat{p} = 16$ with value $RCV_{(\hat{p},\hat{p})} = 0.94$, while the values of $RCV_{(p,p)}$ for $5 \leq p \leq 10$ are between 0.78 and 0.85, so a smaller dimension cannot be considered in this case. On the other hand, after removing the data with indices in $\mathcal{I}_1$ the maximum is attained at $\hat{p} = 5$. Hence, the dimension of the approximating space and the canonical weights computed with the Pearson correlation coefficient after removing the observations with indices in $\mathcal{I}_1$ are closer to the canonical weights computed with the Spearman coefficient using the whole data set. Hence, as expected, the robust procedure leads to more reliable results without excluding atypical data.

A different phenomenon is observed for the functional principal direction basis. In this situation, the classical and robust first canonical variates in the space X, computed with the whole sample, are highly associated since $\rho_{\text{SP}}(P_n[\langle\hat{\Phi}_{\text{SP},1}, X\rangle, \langle\hat{\Phi}_{\text{CL},1}, X\rangle]) = 0.95$, while the angle between the robust and classical canonical weights is close to 38°. After removing the potential atypical observations, the cosine of the angle slightly increases resulting in an angle around 30°. Again, as expected, the estimators based on the normalized Spearman coefficient computed with the complete data set give estimates close to those obtained with the Pearson correlation coefficient after removing the observations detected as atypical.

To highlight the behaviour of the detected observations, Fig. 4 gives a plot of the data trajectories. The thin grey lines in the background correspond to the complete set of observations, and they are included for visual reference, while the potential

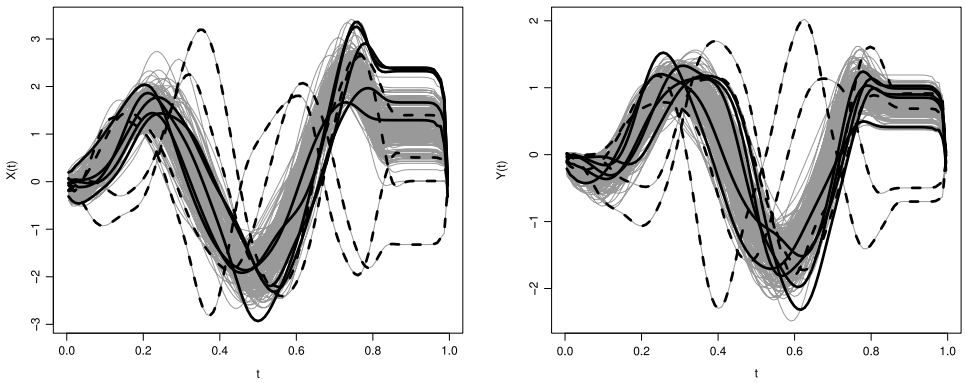

Fig. 4. Speed of the pen on the horizontal and vertical axes. Black dashed curves correspond to atypical observations with indices in $\mathcal{I}_1$, while the black solid curves to those with indices belonging to $\mathcal{I}_2 - \mathcal{I}_1$.

atypical trajectories are given in black lines. To help in visualizing both sets of atypical data, Fig. 4 shows the potential outliers with indices in $\mathcal{I}_1$ in black dashed lines, while those corresponding to indices in $\mathcal{I}_2 - \mathcal{I}_1$ are given in solid black lines (recall that $\mathcal{I}_1 \subset \mathcal{I}_2$).

Most of the trajectories with indices in $\mathcal{I}_2 - \mathcal{I}_1$ correspond to individuals with a higher writing speed over the interval $[0.8, 1]$. It is clear from Fig. 4 that the observations with indices belonging to $\mathcal{I}_1$ correspond to observations far away from the bulk of the data due to their behaviour both in shape and phase. More precisely, these trajectories seem to have a temporal phase shift with respect to the bulk of the data. Effectively, the time when their maximum (minimum) is reached is far from the time in which most of the trajectories reach their maximum (minimum). In particular, two of the trajectories of $\mathcal{I}_1$, corresponding to the observation labelled as 139 and 175, present a distinctly different behaviour within the interval $[0.8, 1]$ (see Figures S.4 and S.5 in the supplementary file). In this two cases, the pen moves slower than the majority on the vertical axis. Furthermore, the individual corresponding to the data labelled 175 has a handwriting of the character "e" slower than most individuals when ending its writing. On the other hand, the observation labelled 33 is atypical since it has a high writing speed on the vertical axis and a low one on the horizontal axis within the interval $[0.8, 1]$. Finally, the maximum of $X(t)$ in the interval $[0, 0.5]$ corresponding to observation 38 clearly exceeds the remaining trajectories. Moreover, this maximum is attained approximately at $t = 0.3$, while for most of the data, the value where the maximum is reached is close to 0.2. As shown in Figure S.6 of the supplementary file, the robust proposal given in this paper was useful to identify potential atypical data which affect the estimation of the first canonical directions. These atypical data correspond to individuals with a clearly different handwriting of the character "e".

8. Concluding remarks

In this paper, we introduce a family of robust estimators for the first canonical weights and the related maximal association for functional data. Using robust association measures, our proposal adapts the projection-pursuit ideas introduced in Alfons et al. [2], Branco et al. [9] and Croux and Filzmoser [12] for multivariate samples with the sieve approach considered in He et al. [26].

Among other contributions, we provide an extension of the result given in Leurgans et al. [35], when a general association measure and not only the Pearson correlation is used. More precisely, we show that the natural extension of the projection-pursuit multivariate estimators considered in Alfons et al. [2] to the functional scenario fails, since directions can be found with empirical canonical association equal to one, motivating our robust proposal which combines robust projection-pursuit with the method of sieves as a smoothing tool.

The robust estimators introduced for the first canonical directions and the maximal association are consistent under mild conditions on the association measure. As in the multivariate case, the proposed estimators are Fisher-consistent for elliptical or Gaussian processes for appropriate choices of the association measure.

Finally, our simulation study confirms the inadequate behaviour of the classical estimators when atypical data arise in the sample, while the robust procedures based on the association measure defined through an M-scatter matrix or the normalized Spearman coefficient lead to more reliable results. In particular, we recommend the procedure based on the normalized Spearman coefficient. As shown in our simulation study, the robust estimators are useful to detect atypical data using the predicted canonical variates. The benefits of considering robust estimators are also illustrated on a real data set where the detection rules reveal the presence of influential observations that would be missed otherwise.

The described procedure can be extended to robustly estimate the subsequent canonical correlations and directions. More precisely, for $k \geq 2$, the kth canonical directions related to the association measure ρ may be defined as $(\Phi_k(P), \Psi_k(P)) = (\Phi_k, \Psi_k) = \text{argmax}_{(u,v)\in\mathcal{B}_k} \rho_{XY}(u, v)$, where $\mathcal{B}_k = \{(u, v) \in \mathcal{S}_1 \times \mathcal{S}_2 : \rho_{XX}(u, \Phi_j) = \rho_{YY}(v, \Psi_j) = 0, \text{ for all } 1 \leq j \leq k-1\}$, while the kth maximal canonical association equals $\rho_k(P) = \rho(P[\langle\Phi_k, X\rangle_{\mathcal{H}_1}, \langle\Psi_k, Y\rangle_{\mathcal{H}_2}])$. The sieves estimators for the kth

canonical directions are defined as $(\hat{\Phi}_k, \hat{\Psi}_k) = \text{argmax}_{(u,v)\in\mathcal{B}_{k,p_n,q_n}} \rho_n(u, v)$, where ρ_n is given in (10) and $\mathcal{B}_{k,p_n,q_n} = \{(u, v) \in \mathcal{S}_{1,p_n} \times \mathcal{S}_{2,q_n} : \rho(P_n[\langle u, X\rangle_{\mathcal{H}_1}, \langle \hat{\Phi}_j, X\rangle_{\mathcal{H}_1}]) = \rho(P_n[\langle v, Y\rangle_{\mathcal{H}_2}, \langle \hat{\Psi}_j, Y\rangle_{\mathcal{H}_2}]) = 0, \text{ for all } 1 \le j \le k-1\}$. Finally, the kth maximal canonical association estimator equals $\hat{\rho}_k = \rho_n\left(\hat{\Phi}_k, \hat{\Psi}_k\right)$. These estimators can be computed using the alternate GRID algorithm as described in Section 3.4 for the first canonical direction estimators. Consistency results for the canonical directions and correlations when $k \ge 2$ are an interesting topic but beyond the scope of this paper. The main difficulties arise by the side null-association conditions $\rho(P_n[\langle u, X\rangle_{\mathcal{H}_1}, \langle \hat{\Phi}_j, X\rangle_{\mathcal{H}_1}]) = \rho(P_n[\langle v, Y\rangle_{\mathcal{H}_2}, \langle \hat{\Psi}_j, Y\rangle_{\mathcal{H}_2}]) = 0$, for $1 \le j \le k-1$. It is worth noticing that, when dealing with the Euclidean case, Jin and Cui [33] impose orthogonality conditions, i.e., $\langle u, \hat{\Phi}_j\rangle_{\mathcal{H}_1} = \langle v, \hat{\Psi}_j\rangle_{\mathcal{H}_2} = 0$ as in principal component analysis to derive the consistency of the estimators loosing the desired null association property between the canonical variates $\langle \hat{\Phi}_j, X\rangle_{\mathcal{H}_1}$, $1 \le j \le k$. These robust canonical direction estimators may be helpful to obtain a resistant estimation procedure for functional canonical regression generalizing the approach considered in He et al. [27].

Functional discrimination has been extensively considered and we refer to Cuevas et al. [15] and Hubert et al. [31] for a depth approach, to Yao et al. [44] for an approach when dealing with sparse data and to Baillo et al. [5] for further discussions. The relation between canonical correlation and discriminant analysis has been widely described in the multivariate setting and also extended to the functional case, see, for instance, Hastie et al. [23] and Ramsay and Silverman [40]. Hence, the robust proposal considered in this paper may be useful to deal with robust functional optimal scoring and discriminant analysis taking Y as a dummy vector coding the group class.

Acknowledgments

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. This research was partially supported by Grants 20020130100279BA from the Universidad de Buenos Aires, PICT 2014–0351 from anpcyt, Argentina and the Spanish Project MTM2016-76969P from the Ministry of Science and Innovation, Spain. G. Boente also wish to thank the Minerva Foundation for its support to present some of this paper results at the International Conference on Robust Statistics 2017.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2018.03.003.

References

[1] A. Alfons, C. Croux, P. Filzmoser, Robust maximum association between data sets: The R package ccaPP, Austral. J. Statist. 45 (2016) 71–79.
[2] A. Alfons, C. Croux, P. Filzmoser, Robust maximum association estimators, J. Amer. Statist. Assoc. 112 (2017) 436–445.
[3] A. Alvarez, Métodos Robustos En Correlación Canónica Funcional, (Ph.D. thesis), Universidad de Buenos Aires, 2017 (in spanish), Available at http://cms.dm.uba.ar/academico/carreras/doctorado/tesisAlvarez.pdf.
[4] G. Aneiros, E.G. Bongiorno, R. Cao, P. Vieu, Functional Statistics and Related Fields, Springer, New York, 2017.
[5] A. Baillo, A. Cuevas, R. Fraiman, Classification methods for functional data, in: F. Ferraty, Y. Romain (Eds.), The Oxford Handbook of Functional Data Analysis, Oxford University Press, 2011, pp. 259–297.
[6] L. Bali, G. Boente, Principal points and elliptical distributions from the multivariate setting to the functional case, Statist. Probab. Lett. 79 (2009) 1858–1865.
[7] G. Boente, M. Salibián-Barrera, $S-$estimators for functional principal component analysis, J. Amer. Statist. Assoc. 110 (2015) 1100–1111.
[8] G. Boente, M. Salibián-Barrera, D.E. Tyler, A characterization of elliptical distributions and some optimality properties of principal components for functional data, J. Multivariate Anal. 131 (2014) 254–264.
[9] J.A. Branco, C. Croux, P. Filzmoser, M.R. Oliveira, Robust canonical correlations: A comparative study, Comput. Statist. 20 (2005) 203–229.
[10] C. Croux, C. Dehon, Analyse canonique base sur des estimateurs robustes de la matrice de covariance, La Revue Stat. Appl. 2 (2002) 5–26.
[11] C. Croux, C. Dehon, Influence functions of the Spearman and Kendall correlation measures, Stat. Methods Appl. 9 (2010) 497–515.
[12] C. Croux, P. Filzmoser, Projection pursuit based measures of association. Research report 0341, Katholieke Universiteit Leuven. Available at https://lirias.kuleuven.be/bitstream/123456789/118289/1/OR_0341.pdf , 2003.
[13] C. Croux, A. Ruiz-Gazen, A fast algorithm for robust principal components based on projection pursuit, in: A. Prat (Ed.), Compstat: Proceedings in Computational Statistics, Physica-Verlag, Heidelberg, 1996, pp. 211–217.
[14] A. Cuevas, A partial overview of the theory of statistics with functional data, J. Statist. Plann. Inference 147 (2014) 1–23.
[15] A. Cuevas, M. Febrero, R. Fraiman, Robust estimation and classification for functional data via projection-based depth notions, Comput. Statist. Data Anal. 22 (2007) 481–496.
[16] J. Cupidon, R. Eubank, D. Gilliam, F. Ruymgaart, Some properties of canonical correlations and variates in infinite dimensions, J. Multivariate Anal. 99 (2008) 1083–1104.
[17] J. Cupidon, D. Gilliam, R. Eubank, F. Ruymgaart, The delta method for analytic functions of random operators with application to functional data, Bernoulli 13 (2007) 1179–1194.
[18] F. Ferraty, Y. Romain, The Oxford Handbook of Functional Data Analysis, Oxford University Press, 2010.
[19] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer, New York, 2006.
[20] P. Filzmoser, C. Dehon, C. Croux, Outlier resistant estimators for canonical correlation analysis, in: J.G. Betlehem, P.G.M. van der Heijden (Eds.), COMPSTAT: Proceedings in Computational Statistics, Physica-Verlag, Heidelberg, 2000, pp. 301–306.
[21] D. Gervini, Robust functional estimation using the spatial median and spherical principal components, Biometrika 95 (2008) 587–600.
[22] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, J. Multivariate Anal. 146 (2016) 1–6.
[23] T. Hastie, A. Buja, R. Tibshirani, Penalized discriminant analysis, Ann. Statist. 23 (1995) 73–102.

[24] G. He, H.G. Müller, J.L. Wang, Extending correlation and regression from multivariate to functional data, in: M. Puri (Ed.), Asymptotics in Statistics and Probability, VSP, Zeist (Netherlands), 2000, pp. 197–210.
[25] G. He, H.G. Müller, J.L. Wang, Functional canonical analysis for square integrable stochastic processes, J. Multivariate Anal. 85 (2003) 54–77.
[26] G. He, H.G. Müller, J.L. Wang, Methods of canonical analysis for functional data, J. Statist. Plann. Inference 122 (2004) 141–159.
[27] G. He, H.G. Müller, J.L. Wang, W. Yang, Functional linear regression via canonical analysis, Bernouilli 16 (2010) 705–729.
[28] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer, New York, 2012.
[29] T. Hsing, R. Eubank, Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators, Wiley, New York, 2015.
[30] M. Hubert, P. Rousseeuw, P. Segaert, Multivariate functional outlier detection, Stat. Methods Appl. 24 (2015) 177–202.
[31] M. Hubert, P. Rousseeuw, P. Segaert, Multivariate and functional classification using depth and distance, Adv. Data Anal. Classif. 11 (2017) 445–466.
[32] M. Hubert, E. Vandervieren, An adjusted boxplot for skewed distributions, Comput. Statist. Data Anal. 52 (2008) 5186–5201.
[33] J. Jin, H. Cui, Asymptotic distributions in the projection pursuit based canonical correlation analysis, Sci. China Math. 53 (2010) 485–498.
[34] G. Karnel, Robust canonical correlation and correspondence analysis, in: The Frontiers of Statistical Scientific and Industrial Applications. Proceedings of ICOSCO-I, Vol. II, American Sciences Press, 1991, pp. 415–420.
[35] S.E. Leurgans, R.A. Moyeed, B.W. Silverman, Canonical correlation analysis when the data are curves, J. Roy. Statist. Soc. Ser B. 55 (1993) 725–740.
[36] 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.
[37] R. Maronna, Robust M−estimators of multivariate location and scatter, Ann. Statist. 4 (1976) 51–67.
[38] R. Maronna, R.D. Martin, V. Yohai, Robust Statistics: Theory and Methods, John Wiley & Sons, New York, 2006.
[39] R. Maronna, R. Zamar, Robust estimates of location and dispersion for high-dimensional datasets, Technometrics 44 (2002) 307–317.
[40] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, Springer, Berlin, 2005.
[41] G.L. Shevlyakov, N.O. Vilchevski, Robustness in Data Analysis: Criteria and Methods, Walter de Gruyter, Utrech, 2001.
[42] S. Taskinen, C. Croux, A. Kankainen, E. Ollila, H. Oja, Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices, J. Multivariate Anal. 97 (2006) 359–384.
[43] B. Williams, M. Toussaint, A. Storkey, A primitive based generative model to infer timing information in unpartitioned handwriting data, in: Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, M. Veloso (Eds.), 2007, pp. 1119-1124.
[44] Y. Yao, F. Wu, J. Zou, Probability–enhanced effective dimension reduction for classifying sparse functional data, TEST 25 (2016) 1–22.
[45] V.J. Yohai, M.G. Ben, Canonical variables as optimal predictors, Ann. Statist. 8 (1980) 865–869.