Statistica Sinica **32** (2022), 1269-1293
doi:https://doi.org/10.5705/ss.202020.0084

ROBUST SMOOTHED CANONICAL CORRELATION ANALYSIS FOR FUNCTIONAL DATA

Graciela Boente[1,3] and Nadia L. Kudraszow[2,3]

[1]*Universidad de Buenos Aires,* [2]*Universidad Nacional de La Plata and* [3]*CONICET, Argentina*

Abstract: We provide robust estimators for the first canonical correlation and directions of random elements on Hilbert separable spaces by using robust association and scale measures, combined with basis expansions and/or penalizations as a regularization tool. Under regularity conditions, the resulting estimators are consistent. The finite-sample performance of our proposal is illustrated by means of a simulation study that shows that, as expected, the robust method outperforms the existing classical procedure when the data are contaminated. A real data example is also presented.

Key words and phrases: Canonical correlation analysis, functional data, robust estimation, smoothing techniques.

1. Introduction

In recent years, data collected in the form of functions or curves have received considerable attention in fields such as chemometrics, economics, environmental studies, image recognition, spectroscopy, and many others. These data are known in the literature as functional data; see Ramsay and Silverman (2005) for a complete overview. As is well known, functional data are intrinsically infinite-dimensional, and this structure is a source of information. Therefore, even when recorded at a finite grid of points, functional observations should be considered as random elements of some functional space rather than as multivariate observations. In this manner, some of the theoretical and numerical challenges posed by the high dimensionality may be solved. This framework led to the extension of some classical multivariate analysis concepts, such as dimension-reduction techniques, to the context of functional data, usually using some regularization tool.

This paper focuses on canonical correlation analysis, where data consist of pairs of random curves. The aim of this analysis is to identify and quantify the relation between the observed functions. Under a Gaussian model,

Corresponding author: Nadia L. Kudraszow, Departamento de Matemática, Universidad Nacional de La Plata, CMaLP-CONICET, (1900) La Plata, Argentina. E-mail: nkudraszow@mate.unlp.edu.ar.

Leurgans, Moyeed, and Silverman (1993) showed that the natural extension of multivariate estimators to the functional scenario fails, motivating the need for a regularization technique that involves smoothing using a penalty term. In addition, He, Müller and Wang (2003) provided conditions that ensure the existence and proper definition of canonical directions and correlations for processes that support a Karhunen–Loéve expansion, and Cupidon et al. (2007) derived the asymptotic distribution of correlations and regularized functional canonical variations. An alternative way to get around the ill-posed problem related to functional canonical correlation analysis is to use a finite basis expansion. Proposals based on this approach are discussed in He, Müller and Wang (2004) and Ramsay and Silverman (2005). More precisely, these authors proposed to performing a regularization step projecting the observed curves on a finite number of basis functions, before computing the smooth canonical correlations and directions in the basis expansion domain.

The aforementioned papers use the Pearson correlation as a measure of the association between the observed functions. However, the Pearson correlation is known to be sensitive to atypical observation, and this sensitivity is inherited by procedures based on it (see Taskinen et al. (2006)). To the best of our knowledge, when considering the analysis of functional canonical correlation, the only proposal of estimators resistant to anomalous observations is that studied by Alvarez, Boente and Kudraszow (2019), who implemented the regularization by projecting random processes on a finite number of functions in an orthonormal basis.

Our aim is to introduce two families of robust consistent estimators for performing functional canonical correlation analysis. For the first family, regularization is based on a roughness penalty term, while for the second, a dimension-reduction technique is also applied. The remainder of the paper is organized as follows. In Section 2, we state some notation and preliminary definitions, and briefly describe the two classical approaches for regularized functional canonical correlation analysis. Section 3 presents our robust proposals. Consistency results are stated in Section 4. The results of a numerical study comparing the robust proposals with their classical counterparts for clean and contaminated samples are given in Section 5, together with an analysis of a real data example. Some final comments are given in Section 6. All proofs are deferred to the Appendix or the online supplementary Material.

2. Preliminaries

Let $\mathcal{H}$ be a separable Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|^2 = \langle\cdot,\cdot\rangle$. Let (X, Y) be a random element of the Hilbert space $\mathcal{H}\times\mathcal{H}$ defined in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. In the product space $\mathcal{H}\times\mathcal{H}$, we define the usual inner product $\langle(u_1, v_1), (u_2, v_2)\rangle_{\mathcal{H}\times\mathcal{H}} = \langle u_1, u_2\rangle + \langle v_1, v_2\rangle$. When $(X, Y)^\top$ has a finite second moment, that is, $\mathbb{E}(\|X\|^2 + \|Y\|^2) < \infty$, we use $\boldsymbol{\Gamma}_{XX} : \mathcal{H} \to \mathcal{H}$, $\boldsymbol{\Gamma}_{YY} : \mathcal{H} \to \mathcal{H}$, $\boldsymbol{\Gamma}_{XY} : \mathcal{H} \to \mathcal{H}$ and $\boldsymbol{\Gamma}_{YX} : \mathcal{H} \to \mathcal{H}$ to denote the covariance and cross-covariance operators, respectively. More precisely, for any $u_1, u_2 \in \mathcal{H}$, $v \in \mathcal{H}$, we have that $\mathrm{Cov}(\langle u_1, X\rangle, \langle u_2, X\rangle) = \langle u_1, \boldsymbol{\Gamma}_{XX} u_2\rangle$ and $\mathrm{Cov}(\langle u_1, X\rangle, \langle v, Y\rangle) = \langle u_1, \boldsymbol{\Gamma}_{XY} v\rangle$, and similarly for $\boldsymbol{\Gamma}_{YY}$ and $\boldsymbol{\Gamma}_{YX}$.

2.1. The classical approaches

Canonical correlation analysis, developed originally for multivariate data, has been extended successfully to accommodate functional data by Leurgans, Moyeed, and Silverman (1993), as follows.

Assume that the observed data $\{(X_i, Y_i)^\top, i = 1, \ldots, n\}$ are independent realizations of a bivariate stochastic process $(X, Y)^\top \in \mathcal{H}\times\mathcal{H}$. When $(X, Y)^\top$ has finite second moments, a nonsmooth approach to the problem of functional canonical correlation is to search for functions u and v in $\mathcal{H}$ such that the linear combinations $\langle u, X\rangle$ and $\langle v, Y\rangle$ have a maximum squared correlation; that is, the objective is to find $u \neq 0$, $v \neq 0$ that maximize

$$\mathcal{L}(u, v) = \mathrm{Corr}^2(\langle u, X\rangle, \langle v, Y\rangle) = \frac{\langle u, \boldsymbol{\Gamma}_{XY} v\rangle^2}{\langle u, \boldsymbol{\Gamma}_{XX} u\rangle \langle v, \boldsymbol{\Gamma}_{YY} v\rangle}, \tag{2.1}$$

where the ratio $\langle u, \boldsymbol{\Gamma}_{XY} v\rangle^2 / (\langle u, \boldsymbol{\Gamma}_{XX} u\rangle \langle v, \boldsymbol{\Gamma}_{YY} v\rangle)$ is equal to zero when $\langle u, \boldsymbol{\Gamma}_{XX} u\rangle = 0$ or $\langle v, \boldsymbol{\Gamma}_{YY} v\rangle = 0$. In particular, Leurgans, Moyeed, and Silverman (1993) considered the case $\mathcal{H} = L^2(\mathcal{I})$. They also assumed that there are two bases of $\mathcal{H}$ composed of the functional canonical coordinates, which are a generalization of the vector canonical coordinates, that ensure the existence of a solution to the nonsmooth approach.

Leurgans, Moyeed, and Silverman (1993) proved that it is not possible to consider a sample version of the problem of maximizing $\mathcal{L}(u, v)$. Therefore, they proposed estimating the first canonical variables by maximizing, restricted to not null "smooth elements" of $\mathcal{H}$, the estimated canonical correlation penalized by a "penalty operator".

As mentioned in the Introduction, two possibilities may be considered to introduce regularization. One approach is to consider, as in Leurgans, Moyeed, and Silverman (1993), a roughness penalty that gives a measure of the smoothness of a function. The second considers a sieve approximation, eventually combined with a penalty term. We briefly review both methods.

Let $D : \mathcal{H}_{\mathrm{S}} \to \mathcal{H}$ be a linear operator, which we refer to as the *differentiator*. Here $\mathcal{H}_{\mathrm{S}}$ is the subset of *smooth elements* of $\mathcal{H}$; that is, $u \in \mathcal{H}_{\mathrm{S}}$ if $\|Du\| < \infty$. Using D, we define the symmetric positive semi-definite bilinear form $\lceil \cdot, \cdot \rceil$: $\mathcal{H}_{\mathrm{S}} \times \mathcal{H}_{\mathrm{S}} \to \mathbb{R}$, where $\lceil u, v \rceil = \langle Du, Dv \rangle$. The *penalization operator* is then defined as $\Psi : \mathcal{H}_{\mathrm{S}} \to \mathbb{R}$, $\Psi(u) = \lceil u, u \rceil$, and the penalized inner product as $\langle u, v \rangle_\tau = \langle u, v \rangle + \tau \lceil u, v \rceil$.

Remark 1. The most common setting for functional data corresponds to the situation where $\mathcal{H} = L^2(\mathcal{I})$ and

$$\mathcal{H}_{\mathrm{S}} = \left\{ u \in L^2(\mathcal{I}),\ u \text{ is twice differentiable, and } \int_{\mathcal{I}} u''(t)^2 dt < \infty \right\}.$$

In this case, it is usual to consider $Du = u''$ and $\lceil u, v \rceil = \int_{\mathcal{I}} u''(t)\, v''(t) dt$, therefore $\Psi(u) = \int_{\mathcal{I}} u''(t)^2 dt$.

Denote $\mathcal{H}_{\mathrm{S}}^0 := \{u \in \mathcal{H}_{\mathrm{S}} : u \neq 0\}$. Given u and v in $\mathcal{H}_{\mathrm{S}}^0$, Leurgans, Moyeed, and Silverman (1993) defined the population penalized squared correlation, $\mathcal{L}_{\boldsymbol{\tau}}(u, v)$, as

$$\begin{aligned}
\mathcal{L}_{\boldsymbol{\tau}}(u, v) &= \frac{\mathrm{Cov}^2(\langle u, X \rangle, \langle v, Y \rangle)}{\{\mathrm{Var}(\langle u, X \rangle) + \tau_1 \Psi(u)\} \{\mathrm{Var}(\langle v, Y \rangle) + \tau_2 \Psi(v)\}} \\
&= \frac{\langle u, \mathbf{\Gamma}_{XY} v \rangle^2}{\{\langle u, \mathbf{\Gamma}_{XX} u \rangle + \tau_1 \Psi(u)\} \{\langle v, \mathbf{\Gamma}_{YY} v \rangle + \tau_2 \Psi(v)\}},
\end{aligned}$$

where $\boldsymbol{\tau} = (\tau_1, \tau_2)$. The so-called *smoothed canonical correlation analysis* (SCCA) by Leurgans, Moyeed, and Silverman (1993), corresponds to maximizing $\mathcal{L}_{\boldsymbol{\tau}}(u, v)$ over $u, v \in \mathcal{H}_{\mathrm{S}}^0$. In this way, for the sample $\{(X_i, Y_i)^\top, i = 1, \ldots, n\}$, the authors proposed performing an SCCA by replacing the population quantities with their sample counterparts, that is, by maximizing the penalized squared sample correlation

$$\begin{aligned}
\widehat{\mathcal{L}}_{\boldsymbol{\tau}}(u, v) &= \frac{\widehat{\mathrm{Cov}}^2(\langle u, X \rangle, \langle v, Y \rangle)}{\left(\widehat{\mathrm{Var}}(\langle u, X \rangle) + \tau_1 \Psi(u)\right) \left(\widehat{\mathrm{Var}}(\langle v, Y \rangle) + \tau_2 \Psi(v)\right)} \qquad (2.2) \\
&= \frac{\left\langle u, \widehat{\mathbf{\Gamma}}_{XY} v \right\rangle^2}{\left(\left\langle u, \widehat{\mathbf{\Gamma}}_{XX} u \right\rangle + \tau_1 \Psi(u)\right) \left(\left\langle v, \widehat{\mathbf{\Gamma}}_{YY} v \right\rangle + \tau_2 \Psi(v)\right)},
\end{aligned}$$

where $\widehat{\text{Cov}}$ and $\widehat{\text{Var}}$ stand for the sample covariance and variance, respectively computed by replacing the corresponding bivariate or univariate distributions with the empirical ones, and $\widehat{\mathbf{\Gamma}}_{XX}$, $\widehat{\mathbf{\Gamma}}_{YY}$, and $\widehat{\mathbf{\Gamma}}_{XY}$ denote the sample covariance and cross-covariance operators.

As mentioned in the Introduction, to address the dimensionality problems of functional canonical correlation analysis, He, Müller and Wang (2004) and Ramsay and Silverman (2005) propose an alternative to the SCCA that employs dimension-reduction techniques, that is, following a sieve approach. More precisely, these authors implement regularization by first projecting the sample's curves on a finite number of elements of an orthonormal basis. In this way, given $\{\xi_i\}_{i\geq 1}$, a suitable orthonormal basis for $\mathcal{H}$, let $\mathcal{H}_d$ be the subspace of $\mathcal{H}$ spanned by $\{\xi_1, \ldots, \xi_d\}$. Then, if we take $d = d_n$ such that $d_n \to \infty$, the sequence of increasing subspaces $\mathcal{H}_{d_n}$ approximates $\mathcal{H}$. From now on, we assume the basis elements are smooth, and so $\mathcal{H}^0_{d_n} := \{u \in \mathcal{H}_{d_n} : u \neq 0\} \subset \mathcal{H}^0_{\text{S}}$. For simplicity, we consider only the case where the same basis is used to approximate both canonical direction estimators.

For the sample $\{(X_i, Y_i)^\top, i = 1, \ldots, n\}$, Ramsay and Silverman (2005) defined the SCCA restricted to the basis expansion domain as the maximization of $\widehat{\mathcal{L}}_{\boldsymbol{\tau}}(u, v)$ over $\mathcal{H}^0_d \times \mathcal{H}^0_d$. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top$ be the coefficients' vectors of u and v in the considered basis, and let $\mathbf{x} = (\langle X, \xi_1\rangle, \ldots, \langle X, \xi_d\rangle)^\top$ and $\mathbf{y} = (\langle Y, \xi_1\rangle, \ldots, \langle Y, \xi_d\rangle)^\top$. It is easily seen that, in the basis expansion domain, the SCCA of the given data is carried out by maximizing the following over $\boldsymbol{\alpha}, \boldsymbol{\beta} \neq \mathbf{0}$:

$$\widehat{\mathcal{L}}^d_{\boldsymbol{\tau}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\widehat{\text{Cov}}^2(\boldsymbol{\alpha}^\top \mathbf{x}, \boldsymbol{\beta}^\top \mathbf{y})}{\left(\widehat{\text{Var}}(\boldsymbol{\alpha}^\top \mathbf{x}) + \tau_1 \sum_{i,j}^d \alpha_i\, \alpha_j\, \lceil \xi_i, \xi_j \rceil\right)\left(\widehat{\text{Var}}(\boldsymbol{\beta}^\top \mathbf{y}) + \tau_2 \sum_{i,j}^d \beta_i\, \beta_j\, \lceil \xi_i, \xi_j \rceil\right)}. \tag{2.3}$$

The maximizers of (2.3) are the coefficient's vectors of the estimated leading canonical directions in the considered basis.

Frequently used bases for functional data are the Fourier, polynomial, splines, and wavelet bases. The practitioner can also use a data-driven basis, such as the one composed of the eigenfunctions of the covariance operators. The number of basis elements, d, should be chosen sufficiently large to ensure that the regularization is controlled by the choice of the smoothing parameter $\boldsymbol{\tau}$ rather than by the dimensionality d.

2.2. Co-association measures

As is well known, the estimators obtained by maximizing $\widehat{\mathcal{L}}_{\boldsymbol{\mathcal{T}}}(u,v)$ are very sensitive to the presence of outliers, because they are based on the sample version of the covariance operators. This suggests that more resistant association measures are needed to get reliable estimations; see, for instance, Alfons, Croux and Filzmoser (2017), who provide robust canonical correlation estimators for multivariate data and a discussion on bivariate association measures. Association measures are an alternative to, and include, the Pearson correlation. In our setting, we seek robust alternatives to the covariance between two random variables because we are penalizing the two variances appearing in the denominator of (2.2). Clearly, a resistant measure can be constructed from a robust association measure and a robust scale estimator. However, other possible choices can be considered. We first give a definition that provides a general framework for robust counterparts of the usual covariance.

Given two univariate random variables U and V, let $F_{(U,V)}$, F_U, and F_V denote the distributions of $(U,V)^\top$, U, and V, respectively. A *bivariate co-association measure* γ between U and V, denoted as $\gamma(F_{(U,V)})$, is a functional defined over the space of bivariate distributions, such that

(i) $\gamma(F_{(U,V)}) = \gamma(F_{(V,U)})$,

(ii) $\gamma(F_{(aU+b,cV+d)}) = a\,c\,\gamma(F_{(U,V)})$, where a, b, c, and d are real constants.

To simplify the notation, we write $\gamma(U,V)$ instead of $\gamma(F_{(U,V)})$ from now on.

Furthermore, if a bivariate co-association measure γ also satisfies the condition

(iii) $\gamma^2(U,V) \leq \gamma(U,U)\,\gamma(V,V)$,

a measure of association may be defined as $\rho\,(U,V) = \gamma\,(U,V)/\sqrt{\gamma\,(U,U)\,\gamma\,(V,V)}$. Clearly, the covariance between two random variables is a co-association measure that satisfies **(i)**–**(iii)**, and its related association measure is the Pearson correlation.

As mentioned above, to provide a robust counterpart of (2.2), robust scale estimators are also needed. To recall the definition of a scale functional, denote $\mathcal{G}$ as the set of all univariate distributions. A scale functional $\sigma : \mathcal{G} \to [0, +\infty)$ is a location-invariant and scale-equivariant functional, that is, $\sigma(F_{aU+b}) = |a|\sigma(F_U)$, for all real numbers a and b (see Maronna et al. (2019)). Two well-known examples of scale functionals are the standard deviation and the median absolute deviation about the median, $\text{MAD}(F_U) = c \text{ median}\,(|U - \text{median}(U)|)$. The normalization constant c, used in the MAD, can be chosen so that its empirical or sample version is consistent for a scale parameter of interest. Typically, one chooses $c = 1/\Phi^{-1}(0.75)$ so that the MAD is equal to the standard deviation at a normal distribution. More generally, any M-scale estimator can be calibrated to provide Fisher-consistent estimators at the normal distribution, that is, $\sigma(\Phi) = 1$, with Φ being the standard normal distribution. As above, when there is no confusion, we write $\sigma(U)$ instead of $\sigma(F_U)$.

Given a bivariate co-association functional γ and a scale functional σ, one can define the related association measure ρ as $\rho(U,V) = \gamma(U,V)/\{\sigma(U)\sigma(V)\}$, if $\rho^2(U,V) \leq 1$, for any two univariate variables U and V. Conversely, given an association measure ρ and a scale functional σ, the related co-association is given by $\gamma(U,V) = \rho(U,V)\ \sigma(U)\sigma(V)$.

Examples of such association measures can be constructed from a bivariate robust scatter functional $\mathbf{W} = \mathbf{W}(U,V)$, which provides a more resistant alternative to the classical covariance matrix $\mathbf{\Sigma} = \mathrm{Cov}(U,V)$. The association measure induced by a bivariate scatter matrix $\mathbf{W}$ is given by

$$\rho(U,V) = \frac{\mathbf{W}_{12}(U,V)}{\{\mathbf{W}_{11}(U,V)\mathbf{W}_{22}(U,V)\}^{1/2}}, \tag{2.4}$$

where $\mathbf{W}_{ij}(U,V)$ is the (i,j)th element of the scatter matrix $\mathbf{W}(U,V)$. One possible choice for $\mathbf{W}(U,V)$ is the M-scatter estimator defined by Maronna (1976), because it provides an efficient estimator that is also highly robust in the bivariate case. Another possible choice is to consider the orthogonalized Gnanadesikan–Kettenring covariance proposed by Maronna and Zamar (2002). When using M-estimators or the orthogonalized Gnanadesikan–Kettenring covariance, the corresponding co-association measure is defined by taking $\gamma(U,V) = \mathbf{W}_{12}(U,V)$ and the related scale estimators as $\sigma(U) = \sqrt{\mathbf{W}_{11}(U,V)}$ and $\sigma(V) = \sqrt{\mathbf{W}_{22}(U,V)}$. Note that $\rho^2(U,V) \leq 1$ when $\mathbf{W}$ is positive semi-definite, which is satisfied by both estimators mentioned above.

Taking into account that $\mathrm{Cov}(U,V) = (\alpha\beta/4)(\mathrm{SD}^2(U/\alpha + V/\beta) - \mathrm{SD}^2(U/\alpha - V/\beta))$, for all $\alpha \neq 0$ and $\beta \neq 0$, where $\mathrm{SD}(\cdot)$ stands for the standard deviation, Gnanadesikan and Kettenring (1972) define a family of co-association functionals, replacing the standard deviation with a robust scale σ and taking $\alpha = \sigma(U)$ and $\beta = \sigma(V)$. More precisely, given a scale functional σ, the co-association measure $\gamma^\star$ is defined as $\gamma^\star(U,V) = \sigma(U)\,\sigma(V)\,(\sigma_+^2 - \sigma_-^2)/4$, with

$$\sigma_+^2 = \sigma^2\left(\frac{U}{\sigma(U)} + \frac{V}{\sigma(V)}\right) \qquad \sigma_-^2 = \sigma^2\left(\frac{U}{\sigma(U)} - \frac{V}{\sigma(V)}\right). \tag{2.5}$$

In order to obtain a highly robust estimator of the correlation between two real random variables, the association measure $\rho^\star(U,V)$ is defined as $\rho^\star(U,V) = (\sigma_+^2 - \sigma_-^2)/4$. However, the resulting measure is not bounded between -1 and 1, because the co-association measure does not satisfy **(iii)**. To ensure an association measure in the valid range, Gnanadesikan and Kettenring (1972) define the association measure $\rho_{\rm GK}$ as $\rho_{\rm GK}(U,V) = (\sigma_+^2 - \sigma_-^2)/(\sigma_+^2 + \sigma_-^2)$, with σ_+^2 and σ_-^2 defined in (2.5), which lies in the range $[-1,1]$. The related a co-association measure is defined as $\gamma_{\rm GK}(U,V) = \sigma(U)\,\sigma(V)\,\rho_{\rm GK}(U,V)$.

Remark 2. We say that $(U,V) \sim \mathcal{E}_2(\boldsymbol{\mu},\boldsymbol{\Sigma},\varphi)$ if $\mathbf{Z} = (U,V)^\top$ is elliptically distributed with location $\boldsymbol{\mu}$, scatter matrix $\boldsymbol{\Sigma}$, and characteristic generator function φ, that is, the characteristic function of $\mathbf{Z}$ is equal to $\psi_{\mathbf{Z}}(\mathbf{t}) = \exp(i\boldsymbol{\mu}^\top\mathbf{t})\varphi(\mathbf{t}^\top\boldsymbol{\Sigma}\mathbf{t})$. As mentioned in Section 2.1 of Alvarez, Boente and Kudraszow (2019), if the robust scatter functional $\mathbf{W}$ is affine-equivariant, the association measure defined in (2.4) is Fisher-consistent for elliptical families; that is, $\rho(U,V) = \Sigma_{12}/\sqrt{\Sigma_{11}\,\Sigma_{22}}$. In particular, the association measure induced by the M-scatter estimator defined by Maronna (1976) is Fisher-consistent at any elliptical distribution. Furthermore, even when the scatter matrix defined in Maronna and Zamar (2002) is not affine equivariant, the association measure ρ given in (2.4) is also Fisher-consistent at any elliptical distribution.

When the scale function $\sigma(\cdot)$ is calibrated to be Fisher-consistent at the normal distribution, $\gamma^\star$ and $\gamma_{\rm GK}$ are Fisher-consistent at the bivariate normal distribution. When considering elliptical distributed random vectors $(U,V) \sim \mathcal{E}_2(\boldsymbol{\mu},\boldsymbol{\Sigma},\varphi)$, it is well known that for any robust scale functional, there exists a constant $c > 0$ such that, for any $a,b \in \mathbb{R}$, $\sigma^2(aU+bV) = c(a^2\Sigma_{11} + b^2\Sigma_{22} + 2ab\Sigma_{12})$ (see, e.g., Maronna et al. (2019)). Straightforward arguments show that, in this situation, $\sigma_+^2 = 2(1+\Sigma_{12}/\sqrt{\Sigma_{11}\,\Sigma_{22}})$ and $\sigma_-^2 = 2(1-\Sigma_{12}/\sqrt{\Sigma_{11}\,\Sigma_{22}})$; thus $\rho^\star$ and $\rho_{\rm GK}$ are also Fisher-consistent at elliptical distributions.

3. Robust Approaches for Smoothed Canonical Correlation Analysis

Throughout this paper, $P_Z[u]$ denotes the distribution of $\langle u, Z\rangle$ when $Z \sim P_Z$, and $P_{(X,Y)}[u,v]$ stands for the joint distribution of $(\langle u,X\rangle, \langle v,Y\rangle)^\top$ when $(X,Y)^\top \sim P_{(X,Y)}$. Furthermore, given a sample $Z_1,\ldots,Z_n$, we write $P_{n,Z}[u]$ for the empirical distribution of $\langle u,Z_1\rangle,\ldots,\langle u,Z_n\rangle$, and $P_{n,(X,Y)}[u,v]$ for that of the bivariate sample $(\langle u,X_i\rangle, \langle v,Y_i\rangle)^\top$, for $1 \le i \le n$.

Let $\gamma_{\rm R}$ and $\sigma_{\rm R}$ be robust co-association and scale functionals, respectively, defining a measure of association; that is, $\gamma_{\rm R}^2(U,V) \le \sigma_{\rm R}^2(U)\,\sigma_{\rm R}^2(V)$. Henceforth, $\gamma_{XY}(u,v) = \gamma_{\rm R}\left(P_{(X,Y)}[u,v]\right)$ and $\sigma_Z(u) = \sigma_{\rm R}(P_Z[u])$, and their sample versions are denoted as $g_n(u,v) = \gamma_{\rm R}(P_{n,(X,Y)}[u,v])$ and $s_{n,Z}^2(u) = \sigma_{\rm R}^2(P_{n,Z}[u])$, respectively. When $\gamma_{\mathrm{R}}(U,V) = \rho_{\mathrm{R}}(U,V)\ \sigma_{\mathrm{R}}(U)\,\sigma_{\mathrm{R}}(V)$ for some association measure ρ_{R}, we write $r_n(u,v) = \rho_{\mathrm{R}}(P_{n,(X,Y)}[u,v])$ and $\rho_{XY}(u,v) = \rho_{\mathrm{R}}\left(P_{(X,Y)}[u,v]\right)$. Furthermore, given any $u,v \in \mathcal{H}$, denote $\mathcal{L}_{\mathrm{R}}(u,v)$ as the robust population squared measure of association between $\langle u, X\rangle$ and $\langle v, Y\rangle$, and $\mathcal{L}_{\boldsymbol{\tau},\mathrm{R}}(u,v)$ as its smoothed version; that is,

$$\mathcal{L}_{\mathrm{R}}(u,v) = \frac{\gamma_{XY}^2(u,v)}{\sigma_X^2(u)\sigma_Y^2(v)} \quad \text{and} \quad \mathcal{L}_{\boldsymbol{\tau},\mathrm{R}}(u,v) = \frac{\gamma_{XY}^2(u,v)}{\{\sigma_X^2(u)+\tau_1\Psi(u)\}\{\sigma_Y^2(v)+\tau_2\Psi(v)\}}\,,$$

where we define $\mathcal{L}_{\mathrm{R}}(u,v) = 0$ when $\sigma_X^2(u) = 0$ or $\sigma_Y^2(v) = 0$. Note that $\mathcal{L}_{\mathrm{R}}$ is the robust counterpart of $\mathcal{L}(u,v)$ in (2.1). Moreover, if γ_{R} is related to an association measure ρ_{R} and the scale functional σ_{R} as $\gamma_{\mathrm{R}}(U,V) = \rho_{\mathrm{R}}(U,V)\ \sigma_{\mathrm{R}}(U)\,\sigma_{\mathrm{R}}(V)$, then $\mathcal{L}_{\mathrm{R}}(u,v) = \rho_{\mathrm{R}}^2\left(P_{(X,Y)}[u,v]\right)$. We refer to the supremum of $\mathcal{L}_{\mathrm{R}}(u,v)$ as the first or maximum canonical association.

As mentioned in Section 2.1, when the co-association measure γ_{R} and the scale functional σ_{R} are taken as the covariance and the standard deviation, respectively, the functional canonical correlation is an ill-posed problem and some regularization is needed. Similarly, when considering a general co-association and scale functionals, it is not possible to consider a sample version of the problem of maximizing $\mathcal{L}_{\mathrm{R}}(u,v)$. More precisely, Proposition 3.1 of Alvarez, Boente and Kudraszow (2019) shows that when $\dim(\mathcal{H}) = \infty$, there are directions such that the empirical association measure $\widehat{\mathcal{L}}_{\mathrm{R}}(u,v) = g_n^2(u,v)/\{s_{n,X}^2(u)\,s_{n,Y}^2(v)\}$ is equal to one. Thus, the proposal of Leurgans, Moyeed, and Silverman (1993) can easily be adapted, using the sample version of $\mathcal{L}_{\boldsymbol{\tau},\mathrm{R}}$, to obtain more stable estimators. To simplify our notation, in what follows, we avoid the subscript R when defining the canonical directions and their estimators. The robust canonical direction functionals and their smoothed versions are defined, respectively, as $(\phi_1,\psi_1) = \mathrm{argmax}_{u,v\in\mathcal{H}_{\mathrm{S}}^0}\,\mathcal{L}_{\mathrm{R}}(u,v)$ and $(\phi_{\boldsymbol{\tau},1},\psi_{\boldsymbol{\tau},1}) = \mathrm{argmax}_{u,v\in\mathcal{H}_{\mathrm{S}}^0}\,\mathcal{L}_{\boldsymbol{\tau},\mathrm{R}}(u,v)$. The sample counterparts of $(\phi_{\boldsymbol{\tau},1},\psi_{\boldsymbol{\tau},1})$ are obtained using the sample versions of the robust co-association and scale functionals; that is, the smoothed robust canonical correlation estimators are given by

$$\begin{aligned}(\widehat{\phi}_{\boldsymbol{\tau},1},\widehat{\psi}_{\boldsymbol{\tau},1}) &= \underset{u,v\in\mathcal{H}_{\mathrm{S}}^0}{\mathrm{argmax}}\,\frac{g_n^2(u,v)}{\{s_{n,X}^2(u)+\tau_1\Psi(u)\}\{s_{n,Y}^2(v)+\tau_2\Psi(v)\}}\\ &= \underset{u,v\in\mathcal{H}_{\mathrm{S}}^0}{\mathrm{argmax}}\,\widehat{\mathcal{L}}_{\boldsymbol{\tau},\mathrm{R}}(u,v)\,. \end{aligned} \tag{3.1}$$

In the same way, the proposal of Ramsay and Silverman (2005) based on regularization by means of both orthonormal bases and a penalization parameter can be easily adapted to be robust maximizing $\widehat{\mathcal{L}}_{\boldsymbol{\tau},\mathrm{R}}$ over $\mathcal{H}_d^0$. Therefore, the smoothed robust canonical correlation estimators in the basis expansion domain are given by

$$(\widetilde{\phi}_{\boldsymbol{\kappa},1}, \widetilde{\psi}_{\boldsymbol{\kappa},1}) = \underset{u,v\,\in\mathcal{H}_d^0}{\operatorname{argmax}}\, \widehat{\mathcal{L}}_{\boldsymbol{\tau},\mathrm{R}}(u,v)\,, \tag{3.2}$$

where $\boldsymbol{\kappa} = (\boldsymbol{\tau}, d)$.

Note that the above maximizations have no unique solution, and that any scalar multiplication of a solution is also a solution. For that reason, conditions over the norms of the directions or the variances of the projections are usually imposed in order to achieve identifiability up to a sign. With this equivalence in mind, we have that (ϕ_1, ψ_1) is the pair of leading robust canonical directions of the model, and $(\widehat{\phi}_{\tau,1}, \widehat{\psi}_{\tau,1})$ and $(\widetilde{\phi}_{\boldsymbol{\kappa},1}, \widetilde{\psi}_{\boldsymbol{\kappa},1})$, given in (3.1) and (3.2), respectively, are its estimators. Note that an unsmoothed robust version of (3.2), that is, when $\tau_1 = \tau_2 = 0$, was studied in Alvarez, Boente and Kudraszow (2019).

4. Consistency

As in Leurgans, Moyeed, and Silverman (1993), to derive consistency results for smoothed robust canonical correlation estimators, it is enough to consider the special case where $\tau_1 = \tau_2 = \tau$; thus, we focus on this case henceforth. Let $\mathcal{N}$ denote the null space of $\lceil\cdot,\cdot\rceil$ and $\mathcal{N}^\perp$ its orthogonal complement. The following assumptions are needed to obtain the desired convergence results.

C1 There exists a constant $c > 0$ and a self-adjoint, positive, compact operator $\boldsymbol{\Gamma} : \mathcal{H}\times\mathcal{H} \to \mathcal{H}\times\mathcal{H}$ such that

$$\boldsymbol{\Gamma} = \begin{pmatrix} \boldsymbol{\Gamma}_{11} & \boldsymbol{\Gamma}_{12} \\ \boldsymbol{\Gamma}_{21} & \boldsymbol{\Gamma}_{22} \end{pmatrix}, \tag{4.1}$$

and for any $u, v \in \mathcal{H}$, $\sigma_X^2(u) = c\,\langle u, \boldsymbol{\Gamma}_{11} u\rangle$, $\sigma_Y^2(v) = c\,\langle v, \boldsymbol{\Gamma}_{22} v\rangle$ and $\gamma_{XY}(u,v) = c\,\langle u, \boldsymbol{\Gamma}_{12} v\rangle$. In addition, the eigenfunctions of $\boldsymbol{\Gamma}_{11}$ and $\boldsymbol{\Gamma}_{22}$ fall in $\mathcal{H}_{\mathrm{S}}$.

C2 There exist functions ϕ_1 and ψ_1 in $\mathcal{H}_{\mathrm{S}}$ such that, for any $u, v \in \mathcal{H}$, we have $\mathcal{L}_{\mathrm{R}}(u,v) \le \mathcal{L}_{\mathrm{R}}(\phi_1,\psi_1) = \rho_1^2$. Furthermore, there exists $0 \le \rho_2 < \rho_1$, such that $\mathcal{L}_{\mathrm{R}}(u,v) \le \rho_2$, for any $u \in \mathcal{H}$ and $v \in \mathcal{H}$ such that $\gamma_{\mathrm{R}}(P_{(X,X)}[u,\phi_1]) = \gamma_{\mathrm{R}}(P_{(Y,Y)}[v,\psi_1]) = 0$. Furthermore, assume that $\|\phi_1\| = 1$ and $\|\psi_1\| = 1$, and that (ϕ_1,ψ_1) is unique up to change of sign.

C3 (a) For any $u \in \mathcal{N}$, $u \neq 0$, $\sigma_X(u) \neq 0$, and $\sigma_Y(u) \neq 0$.

(b) $\mathcal{N}$ is finite dimensional and there exists $d > 0$ such that $\Psi(u) = \lceil u, u\rceil > d\|u\|^2$, for all $u \in \mathcal{N}^\perp$.

Note that in **C1**, we may assume without loss of generality that $c = 1$, redefining $\mathbf{\Gamma}$ as $c\ \mathbf{\Gamma}$. From now on, we denote $\|u\|^2_{1,\tau} = \sigma^2_X(u) + \tau \lceil u, u \rceil = \langle u, \mathbf{\Gamma}_{11} u\rangle + \tau \lceil u, u \rceil$ and $\|v\|^2_{2,\tau} = \sigma^2_Y(v) + \tau \lceil v, v \rceil = \langle v, \mathbf{\Gamma}_{22} v\rangle + \tau \lceil v, v \rceil$. Furthermore, let

$$\begin{aligned} C_{n,X} &= \sup_{\|u\|_{1,\tau_n}=1} \left| s^2_{n,X}(u) - \sigma^2_X(u) \right| , \\ C_{n,Y} &= \sup_{\|v\|_{2,\tau_n}=1} \left| s^2_{n,Y}(v) - \sigma^2_Y(v) \right| , \\ C_{n,XY} &= \sup_{\|u\|_{1,\tau_n}=\|v\|_{2,\tau_n}=1} \left| g_n(u,v) - \gamma_{XY}(u,v) \right| . \end{aligned}$$

We also need the following assumption, which is related to the convergence of the scale and co-association estimators.

C4 The smoothing parameter $\tau = \tau_n \geq 0$ is such that $\tau_n \to 0$, $\max(C_{n,X}, C_{n,Y}) \xrightarrow{a.s.} 0$, and one of the following hold:

(a) $C_{n,XY} \xrightarrow{a.s.} 0$ as $n \to \infty$, and there exists a constant $A > 0$ such that for any $u, v \in \mathcal{H}^0_{\mathrm{S}}$, we have that

$$\widehat{\mathcal{L}}_{\tau,\mathrm{R}}(u,v) = \frac{g^2_n(u,v)}{\{s^2_{n,X}(u) + \tau\Psi(u)\}\{s^2_{n,Y}(v) + \tau\Psi(v)\}} \leq A.$$

(b) The co-association measure is such that $g_n(u,v) = r_n(u,v) s_{n,X}(u)\, s_{n,Y}(v)$, and we have that $\theta_n = \sup_{\|u\|=\|v\|=1} |r_n(u,v) - \rho_{XY}(u,v)| \xrightarrow{a.s.} 0$.

Note that the condition $\widehat{\mathcal{L}}_{\tau,\mathrm{R}}(u,v) \leq A$, for any $u, v \in \mathcal{H}^0_{\mathrm{S}}$, clearly holds with $A = 1$ when $\gamma_{\mathrm{R}}(U,V) = \rho_{\mathrm{R}}(U,V)\, \sigma_{\mathrm{R}}(U)\, \sigma_{\mathrm{R}}(V)$, for some association measure ρ_{R}, as is the case with the classical setting, because $r_n(u,v) \leq 1$. Note too that $r_n(u,v)$ and $\rho_{XY}(u,v)$ are scale invariant, so we also have that $\theta_n = \sup_{\|u\|_{1,\tau_n}=\|v\|_{2,\tau_n}=1} |r_n(u,v) - \rho_{XY}(u,v)|$.

Assume that **C2** holds, and define $\lambda_0 = \rho^2_1$, $\lambda_\tau = \sup_{u,v \in \mathcal{H}^0_{\mathrm{S}}} \mathcal{L}_{\tau,\mathrm{R}}(u,v)$, and $\widehat{\lambda}_\tau = \sup_{u,v \in \mathcal{H}^0_{\mathrm{S}}} \widehat{\mathcal{L}}_{\tau,\mathrm{R}}(u,v)$. As in Section 3, we denote the values maximizing $\mathcal{L}_{\mathrm{R}}(u,v)$, $\mathcal{L}_{\tau,\mathrm{R}}(u,v)$, and $\widehat{\mathcal{L}}_{\tau,\mathrm{R}}(u,v)$ by (ϕ_1, ψ_1), $(\phi_{\tau,1}, \psi_{\tau,1})$, and $(\widehat{\phi}_{\tau,1}, \widehat{\psi}_{\tau,1})$, respectively.

The convergence of the estimators $(\widehat{\phi}_{\tau,1}, \widehat{\psi}_{\tau,1})$ to the first population canonical directions, (ϕ_1, ψ_1), is the convergence with respect to the association measure induced by γ_{R} and σ_{R}, that is analogous to the Γ-norm convergence defined in Leurgans, Moyeed, and Silverman (1993). This convergence means that the canonical variates obtained from $(\widehat{\phi}_{\tau,1}, \widehat{\psi}_{\tau,1})$ for a given random element $(X, Y)^\top$

behave as those obtained from (ϕ_1, ψ_1), which is a desirable property for the estimated canonical directions. To clarify the convergence considered, given $u_1, u_2, v_1, v_2 \in \mathcal{H}$, define the quantities

$$\mathcal{L}_{\mathrm{R}}^X(u_1, u_2) = \frac{\gamma_{\mathrm{R}}^2(P_{(X,X)}[u_1, u_2])}{\sigma_X^2(u_1)\sigma_X^2(u_2)} \quad \text{and} \quad \mathcal{L}_{\mathrm{R}}^Y(v_1, v_2) = \frac{\gamma_{\mathrm{R}}^2(P_{(Y,Y)}[v_1, v_2])}{\sigma_Y^2(v_1)\sigma_Y^2(v_2)} \,.$$

For any pair of sequences $(u_n)_{n\in\mathbb{N}} \subseteq \mathcal{H}$, $(v_n)_{n\in\mathbb{N}} \subseteq \mathcal{H}$, we say that (u_n, v_n) converges to $(u, v) \in \mathcal{H}$ in the $\mathcal{L}_{\mathrm{R}}$-norm if $\mathcal{L}_{\mathrm{R}}^X(u, u_n) \to 1$ and $\mathcal{L}_{\mathrm{R}}^Y(v, v_n) \to 1$.

The following theorem (the proof is given in the Appendix) shows that the robust estimators of the canonical directions given in (3.1) are consistent.

Theorem 1. *Let $(X_1, Y_1)^\top, \ldots, (X_n, Y_n)^\top$ be independent and identically distributed (i.i.d.) with the same distribution as $(X, Y)^\top \sim P_{(X,Y)}$. Assume that* **C1** *–***C2***,* **C3***(a), and* **C4** *hold. Then, we have that*

- ***(a)*** $\widehat{\lambda}_\tau \xrightarrow{a.s.} \lambda_0 = \rho_1^2$, *so the estimate of the maximum canonical association is consistent,*
- ***(b)*** $\mathcal{L}_{\mathrm{R}}(\widehat{\phi}_{\tau,1}, \widehat{\psi}_{\tau,1}) \xrightarrow{a.s.} \lambda_0$, *and*
- ***(c)*** $\mathcal{L}_{\mathrm{R}}^X(\widehat{\phi}_{\tau,1}, \phi_1) \xrightarrow{a.s.} 1$ *and* $\mathcal{L}_{\mathrm{R}}^Y(\widehat{\psi}_{\tau,1}, \psi_1) \xrightarrow{a.s.} 1$.

In order to get consistency results for the robust smoothed canonical correlation estimators in the basis expansion domain, it is necessary to adopt additional notation and assumptions. Let $\boldsymbol{\kappa} = \boldsymbol{\kappa}_n = (\tau_n, d_n)$, $\widetilde{\lambda}_{\boldsymbol{\kappa}} = \sup_{u,v\in\mathcal{H}_{d_n}^0} \widehat{\mathcal{L}}_{\tau,\mathrm{R}}(u, v)$, and $\lambda_{\boldsymbol{\kappa}} = \sup_{u,v\in\mathcal{H}_{d_n}^0} \mathcal{L}_{\tau,\mathrm{R}}(u, v)$, with maximizers $(\widetilde{\phi}_{\boldsymbol{\kappa},1}, \widetilde{\psi}_{\boldsymbol{\kappa},1})$ and $(\phi_{\boldsymbol{\kappa},1}, \psi_{\boldsymbol{\kappa},1})$. Let

$$\begin{aligned}
D_{n,X} &= \sup_{u\in\mathcal{H}_{d_n},\, \|u\|_{1,\tau_n}=1} \left|s_{n,X}^2(u) - \sigma_X^2(u)\right| \,, \\
D_{n,Y} &= \sup_{v\in\mathcal{H}_{d_n},\, \|v\|_{2,\tau_n}=1} \left|s_{n,Y}^2(v) - \sigma_Y^2(v)\right| \,, \\
D_{n,XY} &= \sup_{u,v\in\mathcal{H}_{d_n},\, \|u\|_{1,\tau_n}=\|v\|_{2,\tau_n}=1} \left|g_n(u, v) - \gamma_{XY}(u, v)\right| \,,
\end{aligned}$$

and let $\Pi_{\mathcal{H}_{d_n}}$ be the orthogonal projection operator onto $\mathcal{H}_{d_n}$.

C5 The basis $\{\xi_i\}_{i\geq 1} \subset \mathcal{H}_{\mathrm{S}}^0$ and $d = d_n$ is such that $d_n \to \infty$. The smoothing parameter $\tau = \tau_n \geq 0$ is such that $\tau_n \to 0$, $\max(D_{n,X}, D_{n,Y}) \xrightarrow{a.s.} 0$, and one of the following hold:

(a) $D_{n,XY} \xrightarrow{a.s.} 0$ as $n \to \infty$ and there exists a constant $A > 0$ such that for any $u, v \in \mathcal{H}_{d_n}^0$, we have that $\widehat{\mathcal{L}}_{\tau,\mathrm{R}}(u, v) \leq A$.

(b) The co-association measure is such that $g_n(u,v) = r_n(u,v)s_{n,X}(u)$ $s_{n,Y}(v)$, and we have that $\delta_n = \sup_{u,v\in\mathcal{H}^0_{d_n},\, \|u\|_{1,\tau_n}=\|v\|_{2,\tau_n}=1} |r_n(u,v) - \rho_{XY}(u,v)|$.

C6 $\sigma^2_X(u) : \mathcal{H} \longrightarrow \mathbb{R}$, $\sigma^2_Y : \mathcal{H} \longrightarrow \mathbb{R}$, and $\gamma_{XY} : \mathcal{H}\times\mathcal{H} \longrightarrow \mathbb{R}$ are continuous in ϕ_1, ψ_1, and (ϕ_1,ψ_1), respectively.

Note that assumption **C5** is slightly weaker than **C4**.

The following theorem shows that the robust estimators of the canonical directions given in (3.2) are consistent (see the Appendix for the proof).

Theorem 2. *Let* $(X_1,Y_1)^\top,\ldots,(X_n,Y_n)^\top$ *be i.i.d. with the same distribution as* $(X,Y)^\top \sim P_{(X,Y)}$. *Assume that* **C1**–**C3***(a) and* **C5**–**C6** *hold,* $\tau_n\Psi(\Pi_{\mathcal{H}_{d_n}}\phi_1)\to 0$, *and* $\tau_n\Psi(\Pi_{\mathcal{H}_{d_n}}\psi_1)\to 0$. *Then, we have that*

(a) $\widetilde{\lambda}_{\boldsymbol{\kappa}} \xrightarrow{a.s.} \lambda_0 = \rho_1^2$,

(b) $\mathcal{L}_{\mathrm{R}}(\widetilde{\phi}_{\boldsymbol{\kappa},1},\widetilde{\psi}_{\boldsymbol{\kappa},1}) \xrightarrow{a.s.} \lambda_0$, *and*

(c) $\mathcal{L}^X_{\mathrm{R}}(\widetilde{\phi}_{\boldsymbol{\kappa},1},\phi_1) \xrightarrow{a.s.} 1$ *and* $\mathcal{L}^Y_{\mathrm{R}}(\widetilde{\psi}_{\boldsymbol{\kappa},1},\psi_1) \xrightarrow{a.s.} 1$.

4.1. Some general comments

Assumptions **C2** and **C3** are similar to assumptions 3 and 4 in Leurgans, Moyeed, and Silverman (1993). In particular, **C3**(b) corresponds to the first part of assumption 4 in Leurgans, Moyeed, and Silverman (1993). **C3** is satisfied, for example, when the roughness penalty is the integrated squared second derivative, subject to periodic boundary conditions. Note that the assumptions $\tau\Psi(\Pi_{\mathcal{H}_{d_n}}\phi_1)\to 0$ and $\tau\Psi(\Pi_{\mathcal{H}_{d_n}}\psi_1)\to 0$ in Theorem 2 are satisfied when $\{\Psi(\Pi_{\mathcal{H}_d}\phi_1)\}_{d\in\mathbb{N}}$ and $\{\Psi(\Pi_{\mathcal{H}_d}\psi_1)\}_{d\in\mathbb{N}}$ are bounded.

When γ_{R} is the covariance and σ_{R} is the standard deviation SD, **C1** holds, with $\boldsymbol{\Gamma}_{11} = \boldsymbol{\Gamma}_{XX}$, $\boldsymbol{\Gamma}_{22} = \boldsymbol{\Gamma}_{YY}$, $\boldsymbol{\Gamma}_{12} = \boldsymbol{\Gamma}_{XY}$, and $c = 1$. Note 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, and Theorem 4.8 of He, Müller and Wang (2003) provides a characterization of the canonical directions using an eigen-analysis of the cross-correlation operator, under mild assumptions. As discussed in the Supplementary Material, this moment requirement may be relaxed when other association measures are considered. More precisely, in Section S.2 of the Supplementary Material, we discuss what the target quantities represent for elliptical processes when second moments do not exist.

Recall that a desirable property is that the measures of co-association and scale defining $\mathcal{L}_{\mathrm{R}}$ determine the same canonical directions, which are the target ones, at least for a given distribution family. This property, known as Fisher-consistency, is strongly connected with **C1** and **C2**. In particular, if γ_{R} is related to an association measure ρ_{R} and the scale functional σ_{R} by $\gamma_{\mathrm{R}}(U,V) = \rho_{\mathrm{R}}(U,V)\ \sigma_{\mathrm{R}}(U)\,\sigma_{\mathrm{R}}(V)$, then $\mathcal{L}_{\mathrm{R}}(u,v) = \rho^2_{\mathrm{R}}(P[\langle u, X\rangle,\ \langle v, Y\rangle])$. Thus, for elliptical processes, **C1** is a consequence of the Fisher-consistency of ρ_{R} (see Section S.1 of the Supplementary Material). Some examples of association measures that are Fisher-consistent for elliptical distributed vectors are discussed in Remark 2.

The following lemma gives conditions ensuring that the convergences in **C4** hold, that is, that $C_{n,X} \xrightarrow{a.s.} 0$, $C_{n,Y} \xrightarrow{a.s.} 0$, and $C_{n,XY} \xrightarrow{a.s.} 0$.

Lemma 1. *Let* $(X_1, Y_1)^\top, \ldots, (X_n, Y_n)^\top$ *be i.i.d. with the same distribution as* $(X,Y)^\top \sim P_{(X,Y)}$. *Let* $\eta_n = \sup_{\|u\|=\|v\|=1} |g_n(u,v) - \gamma_{XY}(u,v)|$, $\zeta_n = \sup_{\|u\|=1} |s^2_{n,X}(u) - \sigma^2_X(u)|$, *and* $\nu_n = \sup_{\|v\|=1} |s^2_{n,Y}(v) - \sigma^2_Y(v)|$. *If* **C1** *and* **C3** *hold, and* $\tau_n \to 0$ *and* $\tau_n^{-1} \max(\zeta_n, \nu_n, \eta_n) \xrightarrow{a.s.} 0$ *as* $n \to \infty$, *then we have that* $C_{n,X} \xrightarrow{a.s.} 0$, $C_{n,Y} \xrightarrow{a.s.} 0$, *and* $C_{n,XY} \xrightarrow{a.s.} 0$.

Note that Lemma 1 and Theorem 1 allow us to derive strong consistency results for the canonical directions defined in Leurgans, Moyeed, and Silverman (1993). Effectively, define $L(t) = \log\ \max(t, e)$ and $LL(t) = L(L(t))$, for any $t > 0$. Moreover, denote $LLn = LL(n)$, so that $LLn = \log\,\log n$ for $n \geq 3$, and $LLn = 1$ for $n = 1, 2$. Let $Z = (X,Y)^\top$ and $\mathbf{\Gamma}_{ZZ} = \mathbb{E}\left[\{Z - \mathbb{E}(Z)\} \otimes \{Z - \mathbb{E}(Z)\}\right]$ be its covariance operator. Note that $\mathbf{\Gamma}_{ZZ}$ is a self-adjoint continuous linear operator over $\mathcal{H} \times \mathcal{H}$; moreover, it is a Hilbert–Schmidt operator. For simplicity, $\mathcal{F}$ will stand for the Hilbert space of such operators with an inner product defined by $\langle \mathbf{\Gamma}_1, \mathbf{\Gamma}_2\rangle_{\mathcal{F}} = \mathrm{trace}(\mathbf{\Gamma}_1^* \mathbf{\Gamma}_2) = \sum_{j=1}^\infty \langle \mathbf{\Gamma}_1 z_j, \mathbf{\Gamma}_2 z_j\rangle_{\mathcal{H}\times\mathcal{H}}$, where $\{z_j : j \geq 1\}$ is any orthonormal basis of $\mathcal{H} \times \mathcal{H}$ and $\mathbf{\Gamma}_1^*$ is the adjoint of $\mathbf{\Gamma}_1$. Furthermore, define $V = \{Z - \mathbb{E}(Z)\} \otimes \{Z - \mathbb{E}(Z)\} - \mathbf{\Gamma}_{ZZ}$, which is a zero mean random element in $\mathcal{F}$. Then, if $\mathbb{E}\left\{\|V\|^2_{\mathcal{F}} / LL(\|V\|_{\mathcal{F}})\right\} < \infty$ and $\mathbb{E}\left(\langle V, F\rangle^2_{\mathcal{F}}\right) < \infty$, for any $F \in \mathcal{F}$, the law of iterated logarithms in Hilbert spaces obtained in Acosta and Kuelbs (1983) allows us to conclude that the assumptions in Lemma 1 hold when $\tau_n \sqrt{n/LLn} \to \infty$. Hence, under **C2** and **C3**, the canonical directions are consistent in the $\mathbf{\Gamma}_{ZZ}-$norm.

5. Numerical Results

In this section, we report the results of a small simulation study conducted to compare the finite-sample behavior of the proposed robust estimators with that of the classical ones, that is, those based on the Pearson correlation. We also present an analysis of the Spanish weather data set considered in Dai and Genton (2019). For the robust estimators, we consider the association measure induced from a bivariate robust M-scatter functional, described in Section 2.2, computed using Huber's score function with tuning constant $k_1 = (\chi^2_{2,0.9})^{1/2}$. The classical and robust estimators are labeled as CL and ROB, respectively, in all tables and figures.

5.1. Monte Carlo study

Our simulation model is similar to that considered in He, Müller and Wang (2004) and Alvarez, Boente and Kudraszow (2019). For each replication, we generate independent samples $\{(X_i, Y_i)^\top\}_{i=1}^n \subset \mathcal{H} \times \mathcal{H}$ of size $n = 100$ with $\mathcal{H} = L^2[0, 50]$. The processes are observed over an equispaced grid of 50 points t_j, for $j = 1, \ldots, 50$. Hence, the inner products $\langle X_i, u\rangle_{\mathcal{H}}$ and $\langle Y_i, v\rangle_{\mathcal{H}}$ are approximated as sums over the design points $\{t_j\}_{1\leq j\leq 50}$.

The clean data sets, denoted by C_0, are generated from the same distribution as the Gaussian random element $(X, Y)^\top \in \mathcal{H} \times \mathcal{H}$, given by

$$X(t) = \sum_{j=1}^{m} \eta_j f_j(t) \qquad \text{and} \qquad Y(t) = \sum_{j=1}^{m} \zeta_j f_j(t)\,, \tag{5.1}$$

where $\{f_j\}_{j\geq 1}$ is the Fourier basis of $L^2[0, 50]$ and $m = 21$. The scores $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m)^\top$ and $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_m)^\top$ are normally distributed random vectors, $(\boldsymbol{\eta}^\top, \boldsymbol{\zeta}^\top)^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, with $\boldsymbol{\Sigma}$ a matrix with diagonal blocks $\boldsymbol{\Sigma}_{22} = \boldsymbol{\Sigma}_{11} = 10\,\text{diag}(1, 1, 1, 0.75, \ldots, 0.75^{m-3})$ and an off-diagonal block $\boldsymbol{\Sigma}_{12} = \text{diag}(7, 3, 1, 0, \ldots, 0)$.

Note that for uncontaminated samples, the target quantities do not depend on the selected co-association measure, and are equal to the canonical weights and correlations defined in He, Müller and Wang (2004); that is, 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 $\ell = 1, 2, 3$.

As in Alvarez, Boente and Kudraszow (2019), two contamination models, denoted by C_1 and C_2, are studied so that the estimated canonical directions might be affected. The trajectories obtained from these contaminations are denoted by $\{(X_i^{(c)}, Y_i^{(c)})\}_{i=1}^n$. Outliers were introduced using a Bernoulli random variable $B_i \sim \mathcal{B}(1, 0.1)$, for $1 \leq i \leq n$, which corresponds to 10% outliers. To construct trajectories with patterns different from the clean ones, we consider random variables $W_i \sim \mathcal{N}(25, \sigma^2)$, such that W_i, B_i, $(X_i, Y_i)^\top$ are independent. Under C_2, we additionally generate $(\boldsymbol{\eta}_i^\top, \boldsymbol{\zeta}_i^\top)^\top = (\eta_{i,1}, \ldots, \eta_{i,m}, \zeta_{i,1}, \ldots, \zeta_{i,m})^\top \sim \mathcal{N}(0, \boldsymbol{\Sigma})$, independent of W_i, B_i and $(X_i, Y_i)^\top$.

When $B_i = 0$, the generated trajectories correspond to clean ones, that is,

$(X_i^{(c)}, Y_i^{(c)})^\top \sim (X,Y)^\top$ given in (5.1). When $B_i = 1$, the curve is contaminated, as follows. Under C_1, we define $(X_i^{(c)}, Y_i^{(c)})^\top = W_i(f_2, f_2)^\top$ and take $\sigma^2 = 1$, while under C_2, $\sigma^2 = 0.01$ and $(X_i^{(c)}, Y_i^{(c)})$ are given by

$$X_i^{(c)} = \eta_{i,1}\, f_1 \,+\, W_i \frac{f_3 + f_4}{\sqrt{2}} \,+\, 0.1\, \eta_{i,3}\, f_3 \,+\, 0.1\, \eta_{i,4}\, f_4,$$
$$Y_i^{(c)} = \zeta_{i,1}\, f_1 \,+\, W_i \frac{f_3 + f_4}{\sqrt{2}} \,+\, 0.1\, \zeta_{i,3}\, f_3 \,+ 0.1\, \zeta_{i,4}\, f_4.$$

Note that C_1 is a strong contamination in the direction of the second canonical direction of $(X,Y)^\top$. In C_2, we contaminate in the direction of a linear combination of the third canonical weight and the fourth element of the basis which, for clean samples, corresponds to a canonical direction with null canonical correlation. In all cases, we performed $NR = 1{,}000$ replications.

We present the results for the smoothed canonical estimators in the basis expansion domain defined through (3.2); those corresponding to the procedure given in (3.1) are relegated to the Supplementary Material. Two bases $\{\xi_i\}_{i\geq 1}$ are considered in this numerical study: the cubic B-spline basis, and the Fourier basis (the same basis used to generate the data). The elements of the B-spline basis are orthonormalized before applying the algorithm to compute the estimators. Because the samples are generated using the first $m = 21$ elements of the Fourier basis, for this basis, the dimension of the approximating spaces is selected as $d = 5, 9, 13$. For the cubic B-spline basis, we also considered $d = 20$ as a possible dimension for the linear subspace.

The selected penalization operator is the L^2 norm of the second derivative; that is, $\Psi(u) = \int_0^{50} u''(t)^2 dt$ (see Remark 1). As in Leurgans, Moyeed, and Silverman (1993), the discretization of the roughness penalty is computed over the same design points $\{t_j\}_{1\leq j\leq 50}$, and the second derivative of u at t_i is approximated by $\{u(t_{i+1}) - 2u(t_i) + u(t_{i-1})\}/(t_{i+1} - t_i)^2$, where $u(t_0) = u(t_{50})$ and $u(t_{51}) = u(t_1)$. The values $\lceil \xi_i, \xi_j \rceil = \int_0^{50} \xi_i''(t)\, \xi_j''(t) dt$, for $i, j = 1, \ldots, d$, are evaluated using the described approximation for the second derivative in each element of the basis, and then approximating the integral by sums. To evaluate the influence of the penalty parameter, different values of τ are considered. More precisely, we compare the performance when the smoothing parameter is equal to $\tau = 20, 30$, or 40. In this preliminary study, we analyze the performance for fixed values of (τ, d) only, even though, in practice, a robust cross-validation criterion is recommended, see Alvarez, Boente and Kudraszow (2019).

The smoothed estimators $(\widetilde{\phi}_{\boldsymbol{\kappa},1}, \widetilde{\psi}_{\boldsymbol{\kappa},1})$ given by (3.2) are computed using the algorithm described in the Supplementary Material. For each situation, to study

Table 1. MISE for SCCA when using the Fourier basis and different contamination settings.

$d =$		5			9			13		
$\tau =$		20	30	40	20	30	40	20	30	40
Model	Method									
C_0	CL	0.13	0.13	0.13	0.31	0.27	0.25	0.33	0.29	0.26
	ROB	0.18	0.19	0.18	0.42	0.37	0.34	0.44	0.38	0.34
C_1	CL	3.19	3.19	3.19	3.22	3.21	3.21	3.22	3.21	3.21
	ROB	0.79	0.79	0.79	1.06	1.00	0.96	1.09	1.01	0.97
C_2	CL	3.02	3.02	3.02	3.08	3.06	3.05	3.08	3.07	3.06
	ROB	1.02	1.04	1.02	1.41	1.30	1.26	1.43	1.34	1.28

Table 2. MISE for SCCA when using cubic B-splines and different contamination settings.

$d =$		5			9			13			20		
$\tau =$		20	30	40	20	30	40	20	30	40	20	30	40
Model	Method												
C_0	CL	0.13	0.13	0.13	0.30	0.26	0.24	0.33	0.28	0.26	0.34	0.29	0.26
	ROB	0.18	0.18	0.18	0.38	0.34	0.31	0.43	0.37	0.33	0.44	0.38	0.34
C_1	CL	3.19	3.19	3.19	3.21	3.21	3.21	3.22	3.21	3.21	3.22	3.22	3.20
	ROB	0.79	0.78	0.78	1.03	0.98	0.96	1.08	1.02	0.98	1.08	1.03	0.98
C_2	CL	3.01	3.01	3.01	3.06	3.05	3.04	3.07	3.05	3.04	3.07	3.07	3.04
	ROB	0.99	0.98	0.97	1.29	1.22	1.19	1.32	1.24	1.17	1.35	1.28	1.24

the performance of the first canonical weight estimators, we compute the average over replications of $\|\widetilde{\phi}_{\boldsymbol{\kappa},1}-\phi_1\|^2+\|\widetilde{\psi}_{\boldsymbol{\kappa},1}-\psi_1\|^2$, denoted as the MISE. Tables 1 and 2 report the results when considering the Fourier and B-spline bases, respectively, for each possible combination of the dimension and penalty parameters. We also calculate the measures defined in Alvarez, Boente and Kudraszow (2019) based on the absolute Pearson correlation of the canonical variates over non-atypical data. The results obtained with these measures are similar to those given by the MISE; thus, they are relegated to the Supplementary Material.

Taking into account that the MISE is nonnegative and expected to have a skewed distribution, Figure S.2 in the Supplementary Material presents skewed-adjusted boxplots, as defined in Hubert and Vandervieren (2008), to display the obtained results. The red and blue boxes correspond to the classical and robust procedures, respectively.

As expected, when no outliers are present, all procedures are comparable, leading to small values of the MISE, with the robust procedure performing slightly worse than the nonrobust method, owing to the efficiency loss. In addition, taking into account the smoothness of the true directions, the choice $d = 5$ leads to average values of the MISE that are around half those obtained for larger dimensions.

The stability and advantage of the robust proposal over the classical one when outliers are present in the data can be see in Tables 1 and 2 and Figure S.2. Even though the MISE of the estimators computed using a robust association measure is enlarged under the considered contaminations with respect to those obtained for clean samples, its performance is much better than that of the classical ones. Note that the maximum value of the MISE is equal to four when the estimators are orthogonal to the target directions. Note that, under C_1 and C_2, independently of the basis dimension and penalty parameter, the classical method based on the sample covariances and variances leads to average values of the MISE larger than three. In contrast, the largest values for the average of the MISE when considering the proposed robust method are equal to 1.43 and 1.35 for the Fourier and B-spline bases (see Tables 1 and 2), respectively. Figure S.2 also reveals that, even when large values of the MISE are obtained in some samples for the robust estimators, the median never exceeds 1.032. Indeed, the angle between the target and the robustly estimated directions has a median smaller than 47 degrees for any of the considered contaminations and bases. Note when the basis dimension is equal to five, the reported results are quite similar, independently of the value of τ, for clean and contaminated samples. In contrast, as the dimension increases, the penalty shows its advantage.

5.2. Real data analysis

To illustrate the performance of the estimators of the first canonical directions defined in this paper, we consider the Spanish weather data from the `R` package `fda.usc`, also studied in Dai and Genton (2019). This data set contains geographic information from 73 weather stations in Spain that recorded meteorology information for the period 1980–2009. We only analyze the data related to temperature and wind speed, denoted by $X(t)$ and $Y(t)$, respectively, and focus on the estimated canonical weights using the penalized classical and robust procedures on the basis expansion domain, taking the Fourier basis.

The basis dimension d and the smoothing parameter τ were obtained by cross-validation. For the classical procedure, we considered the criterion defined in He, Müller and Wang (2004), while for the robust one, we adapted the robust criterion defined in Alvarez, Boente and Kudraszow (2019). More precisely, denote $(\widetilde{\phi}_{\boldsymbol{\kappa},1}^{(-i)}, \widetilde{\psi}_{\boldsymbol{\kappa},1}^{(-i)})$ as the first canonical direction estimators defined using (3.2), computed without the ith observation, and let $U_{\boldsymbol{\kappa},1}^{(i)} = \langle \widetilde{\phi}_{\boldsymbol{\kappa},1}^{(-i)}, X_i \rangle$

and $V_{\boldsymbol{\kappa},1}^{(i)} = \langle \widetilde{\psi}_{\boldsymbol{\kappa},1}^{(-i)}, Y_i \rangle$ denote the canonical variates of the ith subject. The robust cross-validation procedure maximizes over $\boldsymbol{\kappa} = (\tau, d)$ in a set of candidates $\mathcal{K} = \mathcal{T} \times \mathcal{D}$, the quantity $RCV(\boldsymbol{\kappa})$, defined as

$$RCV(\boldsymbol{\kappa}) = \rho^2 \left\{ \frac{1}{n} \sum_{i=1}^{n} \Delta_{(U_{\boldsymbol{\kappa},1}^{(i)}, V_{\boldsymbol{\kappa},1}^{(i)})} \right\}, \tag{5.2}$$

where $\Delta_{(a,b)}$, denotes the bivariate probability measure giving all its mass to the point (a, b) and ρ is the same association measure considered in the estimation step. Note that we choose the same dimension and penalizing parameter on both the temperature and wind velocity spaces.

For the basis expansion, we considered only odd dimensions between 5 and 21, that is, $\mathcal{D} = \{d = 2k + 1\,,\, 2 \leq k \leq 9\}$, while the grid $\mathcal{T}$ for possible values of τ consists of 20 equally spaced values between zero and one as well as values between 2 and 40 with step 2. To perform the maximization, we first analyzed the performance of $RCV(\boldsymbol{\kappa})$ when d is fixed and τ varies on $\mathcal{T}$, leading to a value $\widehat{\tau}(d)$ maximizing $RCV(d, \tau)$. We then determined the maximum of $RCV(d, \widehat{\tau}(d))$ over $\mathcal{D}$.

When using the classical procedure, the pair maximizing (5.2) equalled $\widehat{\boldsymbol{\kappa}} = (\widehat{d}, \widehat{\tau}) = (21, 34)$, while for the robust procedure, the maximum was attained at $\widehat{\boldsymbol{\kappa}} = (11, 12)$. The estimators for the first canonical weights are shown in Figure 1. Solid black lines are used for the robust estimators, and solid gray lines are used for the classical ones. The classical estimators are more wiggly than the robust ones, owing to the larger dimension of the basis domain. In addition, the peaks of $\widetilde{\phi}_{\widehat{\boldsymbol{\kappa}},1}$ and $\widetilde{\psi}_{\widehat{\boldsymbol{\kappa}},1}$ in August and mid-November (corresponding to days 240 and 320) for the classical procedure are larger than those of the robust ones. Therefore, the classical procedure suggests that the temperatures and wind speeds in that period should receive a larger positive weight. To complement Figure 1, Table 3 gives the cosine of the angle between the robust and the classical first canonical weight estimates of the first canonical direction estimates of X, labeled $\widetilde{\phi}_{\text{ROB}}$ and $\widetilde{\phi}_{\text{CL}}$, respectively, and those corresponding to Y, labeled $\widetilde{\psi}_{\text{ROB}}$ and $\widetilde{\psi}_{\text{CL}}$, respectively, for simplicity.

We then identified possible influential observations using the bagplot of the canonical variates $U_i = \langle \widetilde{\phi}_{\widehat{\boldsymbol{\kappa}},1}, \widetilde{X}_i \rangle$ and $V_i = \langle \widetilde{\psi}_{\widehat{\boldsymbol{\kappa}},1}, \widetilde{Y}_i \rangle$, where $\widetilde{X}_i = X_i - \widehat{\mu}_X$ and $\widetilde{Y}_i = Y_i - \widehat{\mu}_Y$ are the centered observations, with $\widehat{\mu}_X$ and $\widehat{\mu}_Y$ the spatial medians of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$, respectively. Observations 28, 57, 58, 59, and 73 were detected as influential curves. These observations correspond to the Granada air base, three stations located in Santa Cruz de Tenerife, and the airport

Table 3. Cosine of the angle between the robust and classical first canonical weight estimates. We label as $-\mathcal{I}$ the results obtained when the classical estimator is computed after removing the observations indexed in $\mathcal{I} = \{28, 57, 58, 59, 73\}$. The upper line corresponds to the data-driven dimension, and the second line reports the results obtained when the dimension is equal to $d = 11$ and $\tau = \widehat{\tau}(d)$.

d	$\cos(\widetilde{\phi}_{\mathrm{ROB}}, \widetilde{\phi}_{\mathrm{CL}})$	$\cos(\widetilde{\psi}_{\mathrm{ROB}}, \widetilde{\psi}_{\mathrm{CL}})$	$\cos(\widetilde{\phi}_{\mathrm{ROB}}, \widetilde{\phi}_{\mathrm{CL}}^{-\mathcal{I}})$	$\cos(\widetilde{\psi}_{\mathrm{ROB}}, \widetilde{\psi}_{\mathrm{CL}}^{-\mathcal{I}})$
$\widehat{d}$	0.100	0.549	0.966	0.991
11	0.292	0.911	0.966	0.991

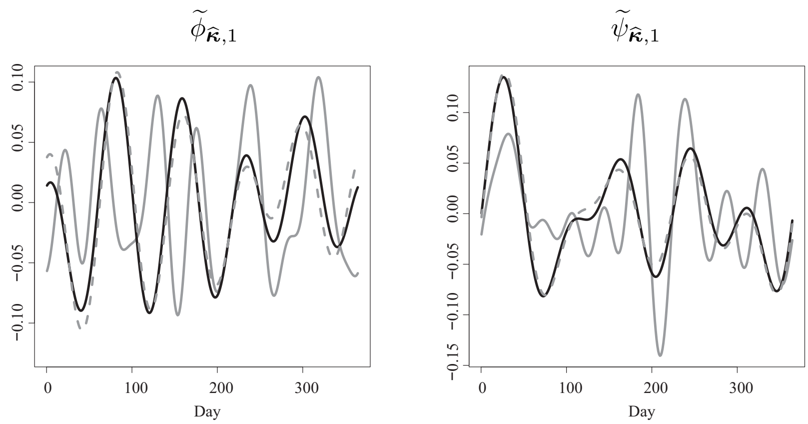

Figure 1. Spanish weather data: Estimates of the canonical weights. The black line corresponds to the robust fit, while the solid and dashed gray lines correspond to the classical estimators computed using the whole data set and without the outliers, respectively.

station in Zaragoza, respectively. In particular, in Zaragoza, the wind speed is almost constant throughout the year. We computed the classical estimators after removing these observations. The optimal value for (d, τ) was $(11, 0.842)$. The dashed lines in Figure 1 correspond to the classical estimators obtained without these possible atypical observations. Note that the classical estimators computed without these potential outliers are very close to the robust ones. In other words, the robust estimator behaves similarly to the classical estimator if one can identify and manually remove suspected outliers.

To determine whether the detected observations affect only the choice of the dimension of the basis, Figure 2 gives the estimates obtained when $d = 11$ for the classical and robust procedures. In this case, the optimal value of τ for the classical estimator is equal to 40. Even though the canonical weights $\widetilde{\psi}_{\widehat{\boldsymbol{\kappa}},1}$ are more similar, those corresponding to temperature show a quite different pattern, specially during the second semester. In particular, the classical estimate has two

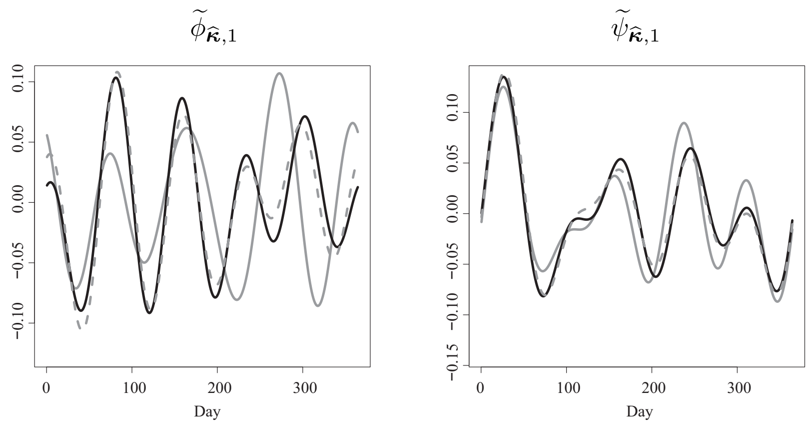

Figure 2. Spanish weather data: Estimates of the canonical weights when $d = 11$. The black line corresponds to the robust fit, while the solid and dashed gray lines correspond to the classical estimators computed using the whole data set and without the outliers, respectively.

"slump" days around days 220 and 320 (mid-August and mid-November), while the robust weight takes a positive value in that region.

6. Conclusion

We have introduced two procedures to obtain robust estimators of canonical directions based on co-association measures and a regularization term involving a roughness penalty. The resulting estimators are consistent under mild conditions. Furthermore, if the process $(X, Y)^\top$ has an elliptical distribution with a finite second moment, the resulting target quantities correspond to the usual canonical correlation and directions. However, if second moments are not assumed, an interpretation of the canonical weights analogous to those given in a classical FCCA, but in terms of the scatter operator, is possible. Finally, our simulation study confirms the poor performance of the SCCA based on Pearson's correlation under contaminated samples. In contrast, the robust procedures based on the association measure defined by an M-dispersion matrix show reliable results for both clean and contaminated samples. We apply our method to a real data set, and confirm that the robust estimators remain reliable, even when the data set contains atypical observations.

Note that we have assumed that X and Y are defined over the same infinite-dimensional space $\mathcal{H}$ to simplify the notation. The extension to the situation in which $X \in \mathcal{H}_1$ and $Y \in \mathcal{H}_2$ is straightforward.

Supplementary Material

The supplementary file contains some comments regarding the Fisher-consistency of the robust canonical directions and their interpretation. It also includes the proofs of Lemma 2 and Proposition 1 and some additional results regarding the simulation study reported in Section 5.

Acknowledgments

The authors wish to thank the associate editor and two anonymous referees for their valuable comments and suggestions. This research was partially supported by grants PICT 2018-00740 from ANPCYT, 20020170100022BA from the Universidad de Buenos Aires at Buenos Aires, Argentina (Graciela Boente), the Spanish Project MTM2016-76969P from the Ministry of Economy and Competitiveness, Spain (MINECO/ AEI/FEDER, UE) (Graciela Boente), and PPID X030 and PID I231 from Universidad Nacional de La Plata, Argentina (Nadia Kudraszow).

Appendix

To prove Theorem 1, we need some preliminary results whose proofs are given in the supplementary file. In particular, Proposition 1 is the key step in the proof of Theorem 1 and provides a sufficient condition for convergence in the $\mathcal{L}_{\mathrm{R}}$-norm. The proof of Lemma 3 is omitted, since it may be derived using similar arguments to those considered in Lemma 4 in Leurgans, Moyeed, and Silverman (1993).

Lemma 2. *Assume that* **C2** *holds. Then, for any u and v in $\mathcal{H}_{\mathrm{S}}$, we have that $\mathcal{L}_{\mathrm{R}}(u,v) \geq \mathcal{L}_{\tau,\mathrm{R}}(u,v)$ and $\lambda_\tau \leq \mathcal{L}_{\mathrm{R}}(\phi_{\tau,1},\psi_{\tau,1})$. Moreover, $\lambda_\tau \to \lambda_0$ as $\tau \to 0$.*

Proposition 1. *Assume that* **C3***(a) and* **C4** *hold, then* $\sup_{u,v\in\mathcal{H}_{\mathrm{S}}^0} |\widehat{\mathcal{L}}_{\tau_n,\mathrm{R}}(u,v) - \mathcal{L}_{\tau_n,\mathrm{R}}(u,v)| \xrightarrow{a.s.} 0$.

Lemma 3. *Assume that* **C1** *and* **C2** *hold and that $\mathcal{L}_{\mathrm{R}}(u_n,v_n) \to \rho_1^2 = \mathcal{L}_{\mathrm{R}}(\phi_1,\psi_1)$ as $n\to\infty$. Then $(u_n,v_n)\to(\phi_1,\psi_1)$ in the $\mathcal{L}_{\mathrm{R}}$-norm.*

Proof of Theorem 1. (a) Standard arguments and Proposition 1 allow to conclude that $|\widehat{\lambda}_\tau - \lambda_\tau| \xrightarrow{a.s.} 0$ which together with Lemma 2 entails that $\widehat{\lambda}_\tau \xrightarrow{a.s.} \lambda_0$. To prove (b) first note that, by Proposition 1, $|\widehat{\mathcal{L}}_{\tau,\mathrm{R}}(\widehat{\phi}_{\tau,1},\widehat{\psi}_{\tau,1}) - \mathcal{L}_{\tau,\mathrm{R}}(\widehat{\phi}_{\tau,1},\widehat{\psi}_{\tau,1})| \xrightarrow{a.s.} 0$. Using $\sim_{a.s.}$ to connect quantities whose difference converges to 0 almost surely, from Lemma 2, we get that

$$\lambda_0 \geq \mathcal{L}_{\mathrm{R}}(\widehat{\phi}_{\tau,1},\widehat{\psi}_{\tau,1}) \geq \mathcal{L}_{\tau,\mathrm{R}}(\widehat{\phi}_{\tau,1},\widehat{\psi}_{\tau,1}) \sim_{a.s.} \widehat{\mathcal{L}}_{\tau,\mathrm{R}}(\widehat{\phi}_{\tau,1},\widehat{\psi}_{\tau,1}) = \widehat{\lambda}_\tau \xrightarrow{a.s.} \lambda_0\,,$$

concluding the proof of (b).
(c) is an immediate consequence of (b) applying Lemma 3.

PROOF OF THEOREM 2. (a) Using **C3**(a) and **C5** and similar arguments to those considered in the proof of Proposition 1, we easily obtain that

$$\sup_{u,v\in\mathcal{H}^0_{d_n}} |\widehat{\mathcal{L}}_{\tau_n,\mathrm{R}}(u,v) - \mathcal{L}_{\tau_n,\mathrm{R}}(u,v)| \xrightarrow{a.s.} 0\ ,$$

which implies $|\widetilde{\lambda}_{\boldsymbol{\kappa}} - \lambda_{\boldsymbol{\kappa}}| \xrightarrow{a.s.} 0$. From Lemma 2, it is easily seen that $\lambda_0 \geq \lambda_{\boldsymbol{\tau}} \geq \lambda_{\boldsymbol{\kappa}} \geq \mathcal{L}_{\tau,\mathrm{R}}(\widetilde{\phi}_{d_n,1}, \widetilde{\psi}_{d_n,1})$, where $\widetilde{\phi}_{d_n,1} = \Pi_{\mathcal{H}_{d_n}}\phi_1/\|\Pi_{\mathcal{H}_{d_n,1}}\phi_1\|$ and $\widetilde{\psi}_{d_n,1} = \Pi_{\mathcal{H}_{d_n}}\psi_1/\|\Pi_{\mathcal{H}_{d_n}}\psi_1\|$ are the standardized orthogonal projections of ϕ_1 and ψ_1 respectively, onto $\mathcal{H}_{d_n}$. Then, the proof of (a) is completed by showing that

$$\mathcal{L}_{\tau,\mathrm{R}}\left(\widetilde{\phi}_{d_n,1}, \widetilde{\psi}_{d_n,1}\right) \to \mathcal{L}_{\mathrm{R}}(\phi_1,\psi_1) = \lambda_0.$$

Since $\|\Pi_{\mathcal{H}_{d_n}}\phi_1\| \to \|\phi_1\|$, $\|\Pi_{\mathcal{H}_{d_n}}\psi_1\| \to \|\psi_1\|$, $\tau\Psi(\Pi_{\mathcal{H}_{d_n}}\phi_1) \to 0$ and $\tau\Psi(\Pi_{\mathcal{H}_{d_n}}\psi_1) \to 0$, by **C6**, we have

$$\frac{\mathcal{L}_{\tau,\mathrm{R}}\left(\widetilde{\phi}_{d_n,1}, \widetilde{\psi}_{d_n,1}\right)}{\mathcal{L}_{\mathrm{R}}(\phi_1,\psi_1)} = \frac{\gamma^2_{XY}\left(\widetilde{\phi}_{d_n,1}, \widetilde{\psi}_{d_n,1}\right)}{\gamma^2_{XY}(\phi_1,\psi_1)} \frac{\sigma^2_X(\phi_1)}{\left\{\sigma^2_X\left(\widetilde{\phi}_{d_n,1}\right) + \tau\Psi\left(\widetilde{\phi}_{d_n,1}\right)\right\}} \frac{\sigma^2_Y(\psi_1)}{\left\{\sigma^2_Y\left(\widetilde{\psi}_{d_n,1}\right) + \tau\Psi\left(\widetilde{\psi}_{d_n,1}\right)\right\}}$$

converges to 1.

Finally, (b) and (c) may be obtained in a similar fashion as (b) and (c) of Theorem 1.

Lemma 4 is needed in the proof of Lemma 1. It corresponds to Lemma 2 in Leurgans, Moyeed, and Silverman (1993) so its proof is omitted.

Lemma 4. *Assume that* **C1**, **C3***(a), and* **C3***(c) hold and let $\ell_1(\tau)$ and $\ell_2(\tau)$ be the smallest eigenvalues of $c\boldsymbol{\Gamma}_{11} + \tau\Psi(\cdot)$ and $c\,\boldsymbol{\Gamma}_{22} + \tau\Psi(\cdot)$, respectively. Then, for $0 < \tau \leq 1$, we have that $\ell_1(\tau) \geq \tau\ell_1(1) > 0$ and $\ell_2(\tau) \geq \tau\ell_2(1) > 0$.*

PROOF OF LEMMA 1. As in Lemma 4, denote $\ell_1(\tau)$ and $\ell_2(\tau)$ as the smallest eigenvalues of $\boldsymbol{\Gamma}_{11} + \tau\Psi(\cdot)$ and $\boldsymbol{\Gamma}_{22} + \tau\Psi(\cdot)$, respectively. Then, we have that for any $u \in \mathcal{H}^0_{\mathrm{S}}$, $\langle u, \boldsymbol{\Gamma}_{11}u\rangle + \tau\Psi(u) \geq \ell_1(\tau)\|u\|^2$, so that $\inf_{u\in\mathcal{H}^0_{\mathrm{S}},\|u\|=1}\{\langle u, \boldsymbol{\Gamma}_{11}u\rangle +$

$\tau\Psi(u)\} \geq \ell_1(\tau)$, which together with the fact that

$$C_{n,X} = \sup_{\|u\|_{1,\tau_n}=1} \left|s^2_{n,X}(u) - \sigma^2_X(u)\right| = \sup_{u\in\mathcal{H}^0_{\mathrm{S}}} \left|\frac{s^2_{n,X}(u) - \sigma^2_X(u)}{\|u\|^2_{1,\tau_n}}\right|$$
$$= \sup_{u\in\mathcal{H}^0_{\mathrm{S}}} \left|\frac{s^2_{n,X}(u) - \sigma^2_X(u)}{\langle u, \mathbf{\Gamma}_{11}u\rangle + \tau_n\Psi(u)}\right| \leq \frac{\sup_{u\in\mathcal{H}^0_{\mathrm{S}},\|u\|=1}\left|s^2_{n,X}(u) - \sigma^2_X(u)\right|}{\inf_{u\in\mathcal{H}^0_{\mathrm{S}},\|u\|=1}\{\langle u, \mathbf{\Gamma}_{11}u\rangle + \tau_n\Psi(u)\}},$$

leads to

$$C_{n,X} \leq \frac{\sup_{u\in\mathcal{H}^0_{\mathrm{S}},\,\|u\|=1}\left|s^2_{n,X}(u) - \sigma^2_X(u)\right|}{\ell_1(\tau_n)} \leq \frac{\zeta_n}{\ell_1(\tau_n)} \leq \frac{\zeta_n}{\tau_n\ell_1(1)} \xrightarrow{a.s.} 0\,.$$

Similar arguments and the fact that $\nu_n/\tau_n \xrightarrow{a.s.} 0$, allow to show that $C_{n,Y} \xrightarrow{a.s.} 0$. Finally,

$$C_{n,XY} = \sup_{\|u\|_{1,\tau_n}=\|v\|_{2,\tau_n}=1} |g_n(u,v) - \gamma_{XY}(u,v)| = \sup_{u,v\in\mathcal{H}^0_{\mathrm{S}}}\left|\frac{g_n(u,v) - \gamma_{XY}(u,v)}{\|u\|_{1,\tau_n}\|v\|_{2,\tau_n}}\right|$$
$$= \sup_{u,v\in\mathcal{H}^0_{\mathrm{S}}}\left|\frac{g_n(u,v) - \gamma_{XY}(u,v)}{\left(\langle u, \mathbf{\Gamma}_{11}u\rangle + \tau_n\Psi(u)\right)^{1/2}\,\left(\langle v, \mathbf{\Gamma}_{22}v\rangle + \tau_n\Psi(v)\right)^{1/2}}\right|$$
$$\leq \frac{\sup_{u,v\in\mathcal{H}^0_{\mathrm{S}},\|u\|=\|v\|=1}|g_n(u,v) - \gamma_{XY}(u,v)|}{\inf_{u\in\mathcal{H}^0_{\mathrm{S}},\|u\|=1}\{\langle u, \mathbf{\Gamma}_{11}u\rangle + \tau_n\Psi(u)\}^{1/2}\,\inf_{v\in\mathcal{H}^0_{\mathrm{S}},\|v\|=1}\{\langle v, \mathbf{\Gamma}_{22}v\rangle + \tau_n\Psi(v)\}^{1/2}}.$$

Using that from Lemma 4 $\ell_j(\tau) \geq \tau\ell_j(1)$, for $j = 1, 2$, we obtain that

$$C_{n,XY} \leq \frac{\sup_{u,v\in\mathcal{H}^0_{\mathrm{S}},\|u\|=\|v\|=1}|g_n(u,v) - \gamma_{XY}(u,v)|}{\left(\ell_1(\tau_n)\,\ell_2(\tau_n)\right)^{1/2}} \leq \frac{\eta_n}{\tau_n\,\left(\ell_1(1)\,\ell_2(1)\right)^{1/2}},$$

which concludes the proof since $\eta_n/\tau_n \xrightarrow{a.s.} 0$.

References

Acosta, A. de and Kuelbs, J. (1983). Some results on the cluster set $C(\{S_n/a_n\})$ and the LIL. *The Annals of Probability* **ll**, 102–122.

Alfons, A., Croux, C. and Filzmoser, P. (2017). Robust maximum association estimators. *Journal of the American Statistical Association* **112**, 436–445.

Alvarez, A., Boente, G. and Kudraszow, N. (2019). Robust sieve estimators for functional canonical correlation analysis. *Journal of Multivariate Analysis* **170**, 46–62.

Cupidon, J., Gilliam, D., Eubank, R. and Ruymgaart, F. (2007). The delta method for analytic functions of random operators with application to functional data. *Bernoulli* **13**, 1179–1194.

Dai, D and Genton, M. (2019). Directional outlyingness for multivariate functional data. *Com-*

putational Statistics and Data Analysis **131**, 50–65.

Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **28**, 81–124.

He, G., Müller, H. G. and Wang, J. L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis* **85**, 54–77.

He, G., Müller, H. G. and Wang, J. L. (2004). Methods of canonical analysis for functional data. *Journal of Statistical Planning and Inference* **122**, 141–159.

Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis* **52**, 5186–5201.

Leurgans, S. E., Moyeed, R. A. and Silverman, B. W. (1993). Canonical correlation analysis when the data are curves. *Journal of the Royal Society, Series B (Methodological)* **55**, 725–740.

Maronna, R. (1976). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics* **4**, 51–67.

Maronna, R., Martin, R. D., Yohai, V. and Salibián-Barrera, M. (2019). *Robust Statistics: Theory and Methods (with R)*. 2nd Edition. John Wiley & Sons, New York.

Maronna, R. and Zamar, R. (2002). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics* **44**, 307–317.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, Berlin.

Taskinen, S., Croux, C., Kankainen, A., Ollila, E. and Oja, H. (2006). Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *Journal of Multivariate Analysis* **97**, 359–384.

Graciela Boente

Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Instituto de Cálculo-CONICET, Ciudad Universitaria, Pabellón 2, 1428, Buenos Aires, Argentina.

E-mail: gboente@dm.uba.ar

Nadia L. Kudraszow

Departamento de Matemática, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, CMaLP-CONICET, Casilla de Correo 172, (1900) La Plata, Argentina.

E-mail: nkudraszow@mate.unlp.edu.ar

(Received March 2020; accepted November 2020)