



Aspects of robust canonical correlation analysis, principal components and association

Jorge G. Adrover¹ · Stella M. Donato²

Received: 6 August 2021 / Accepted: 17 January 2023 / Published online: 20 February 2023
© The Author(s) under exclusive licence to Sociedad de Estadística e Investigación Operativa 2023

Abstract

Principal component analysis (PCA) and canonical correlation analysis (CCA) are dimension-reduction techniques in which either a random vector is well approximated in a lower dimensional subspace or two random vectors from high dimensional spaces are reduced to a new pair of low dimensional vectors after applying linear transformations to each of them. In both techniques, the closeness between the higher dimensional vector and the lower representations is under concern, measuring the closeness through a robust function. Robust SM-estimation has been treated in the context of PCA and CCA showing an outstanding performance under casewise contamination, which encourages the study of asymptotic properties. We analyze consistency and asymptotic normality for the SM-canonical vectors. As a by-product of the CCA derivations, the asymptotics for PCA can also be obtained. A classical measure of robustness as the influence function is analyzed, showing the usual performance of S-estimation in different statistical models. The general ideas behind SM-estimation in either PCA or CCA are specially tailored to the context of association, rendering robust measures of association between random variables. By these means, a robust correlation measure is derived and the connection with the association measure provided by S-estimation for bivariate scatter is analyzed. On the other hand, we also propose a second robust correlation measure which is reminiscent of depth-based procedures.

Keywords Canonical correlation analysis · S-estimation · M-scales · Robust association

Jorge G. Adrover: Research partially supported by Grants PICT 0821 and 0397 from ANPCYT, Grant 05/B424 Secyt, UNC, and Grant 20020100100276 from Secyt, UBA, Argentina. Stella M. Donato: Research partially supported by Grants PICT 0397 from ANPCYT and Secyt, UNC, Argentina.

✉ Jorge G. Adrover
jorge.adrover@unc.edu.ar

Stella M. Donato
stella.donato@fce.uncu.edu.ar

¹ FAMAF, Universidad Nacional de Córdoba, CIEM and CONICET, Córdoba, Argentina

² Facultad de Ciencias Económicas, Universidad Nacional de Cuyo, Mendoza, Argentina

Mathematics Subject Classification 62H20 · 62H25 · 62F35

1 Introduction

Principal component analysis (PCA) and canonical correlation analysis (CCA) are two-dimensional reduction techniques of widespread use in statistics. For a random vector \mathbf{z} in the Euclidean space of dimension m , with positive definite covariance matrix Σ , PCA looks for the spectral decomposition of Σ , the eigenvectors $\mathbf{t}_1^1, \dots, \mathbf{t}_m^1$ associated with the corresponding m -uple of eigenvalues $(\gamma_1^1, \dots, \gamma_m^1)$ in decreasing order, $\gamma_j^1 \geq \gamma_{j+1}^1 > 0$, for all $1 \leq j \leq m - 1$ that is,

$$\Sigma = \mathbf{T}^1 \mathbf{\Delta}^1 (\mathbf{T}^1)^t = \sum_{i=1}^m \gamma_i^1 \mathbf{t}_i^1 (\mathbf{t}_i^1)^t, \tag{1}$$

where $\mathbf{T}^1 \in \mathbb{R}^{m \times m}$ is an orthonormal matrix whose columns are $\mathbf{t}_j^1, j = 1, \dots, m$ and $\mathbf{\Delta}^1 = \text{diag}(\gamma_1^1, \dots, \gamma_m^1)$. The variables $(\mathbf{t}_1^1)^t (\mathbf{z} - E\mathbf{z}), \dots, (\mathbf{t}_m^1)^t (\mathbf{z} - E\mathbf{z})$ are usually referred as principal components. The eigenvalues and eigenvectors can be also obtained through an optimization scheme (Seber 2004, p.181). On the other hand, the principal components are the best linear predictors for $\mathbf{z} - E\mathbf{z}$ when looking for linear combinations $\sum_{k=1}^p (\mathbf{a}_k^t (\mathbf{z} - E\mathbf{z})) \mathbf{a}_k$ based on an orthonormal set $\{\mathbf{a}_1, \dots, \mathbf{a}_p, \dots, \mathbf{a}_m\}$, $p < m$. Let $\mathbf{z} \sim F$, then the principal components solve the optimization problem

$$\begin{aligned} (\boldsymbol{\mu}_z, V_p) &= \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^m, V} E_F \|\mathbf{z} - \boldsymbol{\mu} - \mathbf{P}_V (\mathbf{z} - \boldsymbol{\mu})\|^2 \\ &= \arg \min_{\boldsymbol{\mu} \in \mathbb{R}^m, V} E_F \|P_{V^\perp} (\mathbf{z} - \boldsymbol{\mu})\|^2, \end{aligned} \tag{2}$$

where \mathbf{P}_V stands for the orthogonal projection on a subspace V of dimension $p < m$, $V = \langle \mathbf{a}_1, \dots, \mathbf{a}_p \rangle$ means that V is generated by the orthonormal set $\{\mathbf{a}_1, \dots, \mathbf{a}_p\}$ and $V^\perp = \langle \mathbf{a}_{p+1}, \dots, \mathbf{a}_m \rangle$ denotes the orthogonal complement of V . Then, $(\boldsymbol{\mu}_z, V_p)$ for (2) are given by $\boldsymbol{\mu}_z = E\mathbf{z}$, $V_p = \langle \mathbf{t}_1^1, \dots, \mathbf{t}_p^1 \rangle$ and $\mathbf{P}_{V_p} (\mathbf{z}) = \sum_{k=1}^p ((\mathbf{t}_k^1)^t (\mathbf{z} - E\mathbf{z})) \mathbf{t}_k^1$.

Let us denote the subset $\mathcal{O}_{r,m} = \{\mathbf{A} \in \mathbb{R}^{r \times m} : \mathbf{A}\mathbf{A}^t = \mathbf{I}_r\}$ and $p' = m - p$. According to (2), it is easy to see that

$$\min_{\boldsymbol{\mu}, V} E_F \|\mathbf{z} - \boldsymbol{\mu} - \mathbf{P}_V (\mathbf{z} - \boldsymbol{\mu})\|^2 = \min_{(\mathbf{D}, \mathbf{a}) \in \mathcal{B}_{p',m}^1} E_F \|\mathbf{D}\mathbf{z} - \mathbf{a}\|^2 \tag{3}$$

with

$$\mathcal{B}_{p',m}^1 = \{(\mathbf{D}, \mathbf{a}) : \mathbf{D} \in \mathcal{O}_{p',m}, \mathbf{a} \in \mathbb{R}^{p'}\}. \tag{4}$$

To downweight outlying observations in (3), Maronna (2005) defined SM-estimators for principal vectors in PCA through the equations

$$\sigma^1(\mathbf{D}, \mathbf{a}) = \min \left\{ s > 0 : E_F \chi \left(\frac{\|\mathbf{Dz}-\mathbf{a}\|^2}{s} \right) \leq \delta \right\},$$

$$(\mathbf{D}_{SM}^1, \mathbf{a}_{SM}^1) = \arg \min_{(\mathbf{D}, \mathbf{a}) \in \mathcal{B}_{p',m}^1} \sigma^1(\mathbf{D}, \mathbf{a}), \tag{5}$$

with $\chi : [0, \infty) \rightarrow [0, 1]$. He also considered SL-estimators for principal vectors by minimization of an L-scale rather than an M-scale as in (5).

CCA was proposed by Hotelling (1936) to determine the relationship between two sets of variables obtained by transforming the vectors \mathbf{x} and \mathbf{y} into two vectors \mathbf{z} and \mathbf{w} in lower dimensions, whose association has been greatly strengthened (see Das and Sen 1998 for a very thorough account on CCA and their wide variety of applications). In recent years, CCA has also gained popularity as a method for the analysis of genomic data, since it has the potential to be a powerful tool for identifying relationships between genotype and gene expression. It has also been used in geostatistical applications (see Furrer and Genton 2011). CCA is closely related to multivariate regression when the vectors \mathbf{x} and \mathbf{y} are not treated symmetrically (see Yohai and García Ben 1980). Given the two random vectors \mathbf{x} and \mathbf{y} of dimensions p and q , respectively, the joint covariance matrix is given by

$$\Sigma = \begin{pmatrix} E(\mathbf{x} - E\mathbf{x})(\mathbf{x} - E\mathbf{x})^t & E(\mathbf{x} - E\mathbf{x})(\mathbf{y} - E\mathbf{y})^t \\ E(\mathbf{y} - E\mathbf{y})(\mathbf{x} - E\mathbf{x})^t & E(\mathbf{y} - E\mathbf{y})(\mathbf{y} - E\mathbf{y})^t \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \tag{6}$$

$$\det(\Sigma_{xx}) > 0 < \det(\Sigma_{yy}), 0 < r = \text{rank}(\Sigma_{xy}) \leq \min(p, q) = s.$$

CCA seeks sets $\{\alpha_1, \dots, \alpha_r\} \subset \mathbb{R}^p$ and $\{\beta_1, \dots, \beta_r\} \subset \mathbb{R}^q$, respectively, to yield uncorrelated standardized linear combinations of the variables in \mathbf{x} and the variables in \mathbf{y} that are maximally correlated with each other. We can define the canonical vectors $\alpha_j, \beta_j, j = 1, \dots, r$ (except for the signs) as solutions to an optimization problem (Seber 2004, p. 258). Suppose we take unit length vectors $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^p \times \mathbb{R}^q$ such that $Var(\mathbf{a}^t \mathbf{x}) = 1 = Var(\mathbf{b}^t \mathbf{y})$ and $Corr(\mathbf{b}^t \mathbf{y}, \beta_j^t \mathbf{y}) = 0 = Corr(\mathbf{a}^t \mathbf{x}, \alpha_j^t \mathbf{x}), j = 1, 2, \dots, k - 1$, where Var and $Corr$ stand for the variance and the correlation operators for random variables. With this constraint, we choose (α_k, β_k) to yield the maximum squared correlation between $\mathbf{a}^t \mathbf{x}$ and $\mathbf{b}^t \mathbf{y}$. If ρ_k stands for the positive correlation between $\alpha_k^t \mathbf{x}$ and $\beta_k^t \mathbf{y}$ (the k-th canonical correlation), then $\rho_k^2 = (Corr(\alpha_k^t \mathbf{x}, \beta_k^t \mathbf{y}))^2$ and one gets a decreasing sequence of squared canonical correlations, $\rho_1^2 \geq \dots \geq \rho_r^2$. α_k and β_k will be unique (apart from signs) if the canonical correlations are distinct. It is well known that the optimization problem is equivalent to solving the eigensystem

$$\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \alpha_k = \rho_k^2 \alpha_k, k = 1, \dots, r \tag{7}$$

$$\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \beta_k = \rho_k^2 \beta_k, k = 1, \dots, r, \tag{8}$$

which makes the search computationally more tractable. Classical estimators are obtained by replacing in (7) and (8) by the sample covariance matrix. A robust counterpart of (7) and (8) can be easily performed by solving the linear system, for $k = 1, \dots, r$,

$$\begin{aligned} \Sigma^{(R)} &= \begin{pmatrix} \Sigma_{xx}^{(R)} & \Sigma_{xy}^{(R)} \\ \Sigma_{yx}^{(R)} & \Sigma_{yy}^{(R)} \end{pmatrix}, \\ \Sigma_{xx}^{(R)-1} \Sigma_{xy}^{(R)} \Sigma_{yy}^{(R)-1} \Sigma_{yx}^{(R)} \alpha_k^{(R)} &= \left(\rho_k^{(R)}\right)^2 \alpha_k^{(R)}, \\ \Sigma_{yy}^{(R)-1} \Sigma_{yx}^{(R)} \Sigma_{xx}^{(R)-1} \Sigma_{xy}^{(R)} \beta_k^{(R)} &= \left(\rho_k^{(R)}\right)^2 \beta_k^{(R)}, \end{aligned} \tag{9}$$

with $\Sigma^{(R)}$ a robust dispersion estimator.

The canonical variables $\mathbf{z} = (\alpha_1^t(\mathbf{x} - E\mathbf{x}), \dots, \alpha_r^t(\mathbf{x} - E\mathbf{x}))^t$ and $\mathbf{w} = (\beta_1^t(\mathbf{y} - E\mathbf{y}), \dots, \beta_r^t(\mathbf{y} - E\mathbf{y}))^t$ are also the best linear combinations to predict each other by making the mean squared loss $E_F \|\mathbf{z} - \mathbf{w}\|^2$ as small as possible (see Seber 2004, p. 260), since they solve the optimization problem

$$(\mathbf{A}_C, \mathbf{B}_C, \mu_x, \mu_y) = \arg \min_{(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \mu, \nu) \in \mathcal{C}} E_F \left\| \bar{\mathbf{A}}(\mathbf{x} - \mu) - \bar{\mathbf{B}}(\mathbf{y} - \nu) \right\|^2 \tag{10}$$

with

$$\mathcal{C} = \left\{ (\bar{\mathbf{A}}, \bar{\mathbf{B}}, \mu, \nu) : \bar{\mathbf{A}} \in \mathbb{R}^{r \times p}, \bar{\mathbf{B}} \in \mathbb{R}^{r \times q}, \mu \in \mathbb{R}^p, \nu \in \mathbb{R}^q, \bar{\mathbf{A}} \Sigma_{xx} \bar{\mathbf{A}}^t = \mathbf{I}_r = \bar{\mathbf{B}} \Sigma_{yy} \bar{\mathbf{B}}^t \right\}, \tag{11}$$

\mathbf{I}_r an $r \times r$ identity matrix, $\mu_x = E\mathbf{x}$ and $\mu_y = E\mathbf{y}$. The subscript C stands for Classical.

Adrover and Donato (2015) introduced SM-estimators for canonical vectors in CCA as follows. Given the matrices $\bar{\mathbf{A}} \in \mathbb{R}^{r \times p}$ and $\bar{\mathbf{B}} \in \mathbb{R}^{r \times q}$, let us take $\mathbf{A} = \bar{\mathbf{A}} \Sigma_{xx}^{1/2}$, $\mathbf{B} = \bar{\mathbf{B}} \Sigma_{yy}^{1/2}$, with Σ_{xx} and Σ_{yy} given in (6), $\mathbf{D} = (\mathbf{A} - \mathbf{B}) \in \mathbb{R}^{r \times m}$, $m = p + q$ and the random vector $\mathbf{z} = (\mathbf{x}^t \Sigma_{xx}^{-1/2}, \mathbf{y}^t \Sigma_{yy}^{-1/2})^t$. By reformulating (10) and (11) for the standardized vectors $\Sigma_{xx}^{-1/2} \mathbf{x}$ and $\Sigma_{yy}^{-1/2} \mathbf{y}$, we have

$$\min_{(\bar{\mathbf{A}}, \bar{\mathbf{B}}, \mu, \nu) \in \mathcal{C}} E_F \left\| \bar{\mathbf{A}}\mathbf{x} - \bar{\mathbf{B}}\mathbf{y} - (\bar{\mathbf{A}}\mu - \bar{\mathbf{B}}\nu) \right\|^2 = \min_{(\mathbf{D}, \mathbf{a}) \in \mathcal{B}_{r,m}^0} E_F \|\mathbf{D}\mathbf{z} - \mathbf{a}\|^2 \tag{12}$$

with $\mathcal{B}_{r,m}^0 = \{(\mathbf{D}, \mathbf{a}) : \mathbf{a} \in \mathbb{R}^r, \mathbf{D} = (\mathbf{A} - \mathbf{B}) \in \mathbb{R}^{r \times m}, \mathbf{A} \in \mathcal{O}_{r,p}, \mathbf{B} \in \mathcal{O}_{r,q}\}$.

Since the covariance matrix for the standardized random vector \mathbf{z} is given by

$$\mathbf{M} = \begin{pmatrix} \mathbf{I}_p & \Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1/2} \\ \Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1/2} & \mathbf{I}_q \end{pmatrix}, \tag{13}$$

to evaluate the “largeness” of the “residuals” $\|A\tilde{x} - B\tilde{y} - a\|^2$, an M-scale $\sigma = \sigma(A, B, a)$ is computed implicitly through

$$\sigma(A, B, a) = \min \left\{ \sigma > 0 : E_F \chi \left(\frac{\|A\tilde{x} - B\tilde{y} - a\|^2}{\sigma} \right) \leq \delta \right\}, \tag{14}$$

with $\chi : [0, \infty) \rightarrow [0, 1]$, $\Sigma_{xx}^{(R)}$ and $\Sigma_{yy}^{(R)}$ robust dispersion estimators for Σ_{xx} and Σ_{yy} , respectively, $\tilde{x} = (\Sigma_{xx}^{(R)})^{-1/2} x$, $\tilde{y} = (\Sigma_{yy}^{(R)})^{-1/2} y$ and $((A - B), a) \in \mathcal{B}_{r,m}^0$. Then, the robust standardized SM-canonical vectors are defined through the equation

$$(A_{SM}^o, B_{SM}^o, a_{SM}^o) = \arg \min_{(A,B,a) \in \mathcal{B}_{r,m}^0} \sigma(A, B, a), \tag{15}$$

and the final SM-canonical vectors are defined as

$$A_{SM} = A_{SM}^o (\Sigma_{xx}^{(R)})^{-1/2}, B_{SM} = B_{SM}^o (\Sigma_{yy}^{(R)})^{-1/2}.$$

If we have a random sample z_1, \dots, z_n and $F = F_n$ stands for the empirical distribution function based on z_1, \dots, z_n , the sample version of the estimates is simply obtained by replacing the population expectation by the empirical expectation, that is, $\hat{\sigma}$ is the robust scale based on the sample given in (14) and $(\hat{A}_{SM}^o, \hat{B}_{SM}^o, \hat{a}_{SM}^o)$ the solutions to (15) using $\hat{\sigma}$, that is

$$(\hat{A}_{SM}^o, \hat{B}_{SM}^o, \hat{a}_{SM}^o) = \arg \min_{(A,B,a) \in \mathcal{B}_{r,m}^0} \hat{\sigma}(A, B, a), \tag{16}$$

$$\hat{A}_{SM} = \hat{A}_{SM}^o (\hat{\Sigma}_{xx}^{(R)})^{-1/2}, \hat{B}_{SM} = \hat{B}_{SM}^o (\hat{\Sigma}_{yy}^{(R)})^{-1/2}. \tag{17}$$

The algorithm to compute the SM-estimators is easily derived from the fact that we have a constrained minimization and the Lagrange multipliers method applies (Adrover and Donato 2015). In the context of sparsity, (10) was also considered by Wilms and Croux (2015) as well as a robust proposal in Wilms and Croux (2016).

The outstanding robust performance of (5) and (15) suggests the study of asymptotic properties. Li and Chen (1985) dealt with a robust procedure by considering a robust dispersion S rather than the Var operator in the optimization scheme for PCA, Cui et al. (2003) obtained the asymptotic distribution of the procedure, and Croux and Ruiz-Gazen (2005) tackled the problem of influence function. Drašković et al. (2019) deals with the derivation of the asymptotic behavior for robust PCA in the context of complex elliptically symmetric distributions based on the spectral decomposition of the Maronna’s monotone multivariate dispersion estimator (Maronna 1976), extending previous results (Tyler 1981; Boente 1987; Croux and Haesbroeck 2000). Croux et al. (2017) considered trimmed estimators in the PCA context (in special the SL-estimator given in Maronna 2005 is also included) studying theoretical properties such as consistency, influence function and breakdown point. ten Berge (1979) and Lemma 3 in

Adrover and Donato (2015) explored the close relationship between PCA and CCA which is given by the fact that the principal vectors of \mathbf{M} defined in (13) comprise the transformed canonical vectors of Σ . Thus, asymptotics and influence function for the SM-estimator given by Maronna (2005) in the context of PCA is easily derived from the arguments used in the CCA case included in this paper and therefore omitted.

Anderson (1999) derived the asymptotic distribution for the canonical correlation and the canonical vectors when sampling is from the normal distribution. Taskinen et al. (2006) stated asymptotic properties for CCA based on robust estimators of the covariance matrix as in (9). Alfons et al. (2017) treated asymptotic properties for projection pursuit estimators (see Branco et al. 2005), asymptotic distribution and influence function.

In Sect. 2, we establish the consistency under elliptical distributions for the SM-estimators given in (16) and (17). In Sect. 3, the asymptotic distribution is derived for the SM-estimators and Sect. 4 analyzes the influence function (IF) for the proposal. In Sect. 5, we revise the concept of association between random variables by analyzing some proposals for robust correlation measures which exploit the concept of residual smallness as in (2) and (10). In Sect. 6, we include some concluding remarks. Some relevant proofs are deferred to the ‘‘Appendix’’.

2 Consistency of SM-estimators for CCA

In the multivariate location and dispersion model (MLDM), we have an m -dimensional random vector $\mathbf{z} = (z_1, \dots, z_m)^t$ with distribution $F_{\mu, \Sigma}(B) = F_0(\Sigma^{-1/2}(B - \mu))$, where F_0 is a known distribution in \mathbb{R}^m , B is a Borel set in \mathbb{R}^m , $\mu \in \mathbb{R}^m$ and $\Sigma \in S_m$, the set of $m \times m$ positive definite matrices. An important case is the family of elliptical distributions. The elliptical model allows for a great variety of distributions which comprises the majority of the distributions used in practice, not only the multivariate normal distribution but also distributions without finite moments. We say that an m -dimensional random vector has an elliptical distribution if it has a density of the form

$$f(\mathbf{z}, \mu_0, \Sigma_0) = \frac{1}{(\det \Sigma_0)^{1/2}} f_0((\mathbf{z} - \mu_0)^t \Sigma_0^{-1} (\mathbf{z} - \mu_0)), \quad (18)$$

where $f_0 : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ (it is denoted by $\mathbf{z} \sim E_m(\mu_0, \Sigma_0)$). If $\mathbf{z} \sim E_m(\mathbf{0}, \mathbf{I})$, then $\mathbf{a}^t \mathbf{z}$ has the same distribution for all $\mathbf{a} \in S^{m-1} = \{\mathbf{a} \in \mathbb{R}^m : \|\mathbf{a}\| = 1\}$. In case of having $\mathbf{z} = (\mathbf{x}^t, \mathbf{y}^t)^t \sim E_m(\mu_0, \Sigma_0)$, with the location and dispersion parameters partitioned as

$$\mu_0 = \begin{pmatrix} \mu_{0,x} \\ \mu_{0,y} \end{pmatrix} \text{ and } \Sigma_0 = \begin{pmatrix} \Sigma_{0,xx} & \Sigma_{0,xy} \\ \Sigma_{0,yx} & \Sigma_{0,yy} \end{pmatrix}$$

respectively, then $\mathbf{x} \sim E_m(\mu_{0,x}, \Sigma_{0,xx})$ and $\mathbf{y} \sim E_m(\mu_{0,y}, \Sigma_{0,yy})$. Let us now take the random vector $\mathbf{z}_0 = (\mathbf{x}^t \Sigma_{0,xx}^{-1/2}, \mathbf{y}^t \Sigma_{0,yy}^{-1/2})^t$, then $\mathbf{z}_0 \sim E_m(\tilde{\mu}_0, \mathbf{M}_0)$, with the location

and dispersion parameters partitioned as

$$\tilde{\mu}_0 = \begin{pmatrix} \mu_{0,x} \Sigma_{0,xx}^{-1/2} \\ \mu_{0,y} \Sigma_{0,yy}^{-1/2} \end{pmatrix} \text{ and } \mathbf{M}_0 = \begin{pmatrix} \mathbf{I}_p & \Sigma_{0,xx}^{-1/2} \Sigma_{0,xy} \Sigma_{0,yy}^{-1/2} \\ \Sigma_{0,yy}^{-1/2} \Sigma_{0,yx} \Sigma_{0,xx}^{-1/2} & \mathbf{I}_q \end{pmatrix}.$$

If $\mathbf{z} \sim E_m(\mu_0, \Sigma_0)$ has finite second moments, then $E\mathbf{z} = \mu_0$, the covariance matrix Σ and Σ_0 are equal up to a constant, that is, $\Sigma = c \Sigma_0$ for some positive constant c , and $\mathbf{M}_0 = \mathbf{M}$ with \mathbf{M} given in (13). For the sake of simplicity, we will only keep the notation \mathbf{M} for either \mathbf{M} or \mathbf{M}_0 since they coincide in case of having finite second moments. From now on, we will use the symbols either Σ_{xx} and Σ_{yy} to refer to the multivariate dispersion parameters at the elliptical model. The possibility of considering a general elliptical distribution rather than a multivariate normal distribution let deal with a broader scenario for modeling.

Let us take the spectral decomposition for \mathbf{M} , $\mathbf{M} = \sum_{i=1}^{p+q} \gamma_i^0 \mathbf{t}_i^0 (\mathbf{t}_i^0)^t$, with eigenvalues $\gamma_1^0 \geq \dots \geq \gamma_{p+q}^0 \geq 0$ and eigenvectors $\{\mathbf{t}_j^0\}_{j=1}^{p+q}$, $(\mathbf{t}_j^0)^t \mathbf{t}_j^0 = \delta_{ij}$, $1 \leq i, j \leq p + q$, with δ_{ij} the Kronecker delta. Then, $\mathbf{M} = \mathbf{P}_0 \Gamma_0 \mathbf{P}_0^t$, with $\Gamma_0 = \text{diag}(\gamma_1^0, \gamma_2^0, \dots, \gamma_{p+q}^0)$ and \mathbf{P}_0 an orthogonal matrix whose columns are $\mathbf{t}_1^0, \dots, \mathbf{t}_{p+q}^0$. Take $\mathbf{v}_i^0 \in \mathbb{R}^p$ and $\mathbf{w}_i^0 \in \mathbb{R}^q$, $i = 1, \dots, p + q$, as $\mathbf{t}_i^0 = ((\mathbf{v}_i^0)^t, (\mathbf{w}_i^0)^t)^t$. Then, let us call

$$\begin{aligned} \mathbf{A}_o &= \left(\frac{\mathbf{v}_{p+q-r+1}^0}{\|\mathbf{v}_{p+q-r+1}^0\|}, \dots, \frac{\mathbf{v}_{p+q}^0}{\|\mathbf{v}_{p+q}^0\|} \right)^t \in \mathbb{R}^{r \times p}, \\ \mathbf{B}_o &= \left(\frac{\mathbf{w}_{p+q-r+1}^0}{\|\mathbf{w}_{p+q-r+1}^0\|}, \dots, \frac{\mathbf{w}_{p+q}^0}{\|\mathbf{w}_{p+q}^0\|} \right)^t \in \mathbb{R}^{r \times q}, \\ \mathbf{a}_o &= \mathbf{A}_o \Sigma_{xx}^{-1/2} \mu_x - \mathbf{B}_o \Sigma_{yy}^{-1/2} \mu_y, \\ \sigma_o &= \sigma(\mathbf{A}_o, \mathbf{B}_o, \mathbf{a}_o). \end{aligned}$$

A zero $(r \times s)$ -matrix is a matrix all of whose entries are zero and we denote it as $\mathbf{0}_{r \times s}$. Given square matrices $\mathbf{N}_i \in \mathbb{R}^{p_i \times p_i}$, $i = 1, \dots, k$, set the matrix $\mathbf{N} = \text{diag}(\mathbf{N}_1, \dots, \mathbf{N}_k) \in \mathbb{R}^{\sum_{i=1}^k p_i \times \sum_{i=1}^k p_i}$ whose entries off the blocks $\mathbf{N}_1, \dots, \mathbf{N}_k$ are zero. $\{\mathbf{f}_1^{(s)}, \dots, \mathbf{f}_s^{(s)}\}$ stands for the canonical basis in \mathbb{R}^s .

If $\mathbf{w} \in \mathbb{R}^m$ and \mathcal{L}_m stands for the set of distributions of \mathbf{w} denoted by $\mathcal{L}(\mathbf{w})$, we call a multivariate location and dispersion functional to an application $(\mathbf{T}, S): \mathcal{L}_m \rightarrow \mathbb{R}^m \times \mathbb{R}^{m \times m}$ such that (i) $S(\mathcal{L}(\mathbf{w})) \in S_m$, (ii) it is affine equivariant, i.e., given a nonsingular matrix $\mathbf{G} \in \mathbb{R}^{m \times m}$ and a vector $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{T}(\mathcal{L}(\mathbf{G}\mathbf{w} + \mathbf{b})) = \mathbf{G}\mathbf{T}(\mathcal{L}(\mathbf{w})) + \mathbf{b}$, $S(\mathcal{L}(\mathbf{G}\mathbf{w} + \mathbf{b})) = \mathbf{G}S(\mathcal{L}(\mathbf{w}))\mathbf{G}^t$.

Let us take location and dispersion functionals (\mathbf{T}, S) such that $(\mathbf{T}_x, S_x) = (\mathbf{T}(\mathcal{L}(\mathbf{x})), S(\mathcal{L}(\mathbf{x})))$ and $(\mathbf{T}_y, S_y) = (\mathbf{T}(\mathcal{L}(\mathbf{y})), S(\mathcal{L}(\mathbf{y})))$, respectively. Take $\mathbf{x}_c = \mathbf{x} - \mathbf{T}_x$, $\bar{\mathbf{x}} = \mathbf{S}_x^{-1/2} \mathbf{x}_c$, $\mathbf{y}_c = \mathbf{y} - \mathbf{T}_y$, $\bar{\mathbf{y}} = \mathbf{S}_y^{-1/2} \mathbf{y}_c$ and $\mathbf{z}_c = \mathbf{z} - \mathbf{T}_z$.

The functional (\mathbf{T}, \mathbf{S}) for the location and dispersion parameters at MLDM is said to be Fisher consistent if $\mathbf{T}(F_{\mu, \Sigma}) = \mu$ and $\mathbf{S}(F_{\mu, \Sigma}) = \Sigma$. Adrover and Donato (2015) gave the definition of a Fisher consistent CCA functional, and they showed the Fisher consistency of the SM-estimator for CCA. For this purpose, Fisher consistent functionals for Σ_{xx} and Σ_{yy} are required.

In order to deal with SM-estimators well defined as well as consistent and asymptotically normal, some conditions are required:

C0 \mathbf{M} has eigenvalues $(\gamma_1^0, \dots, \gamma_m^0) \in \Gamma$, with

$$\Gamma = \{(\gamma_1, \dots, \gamma_{p+q}) : \gamma_1 > \dots > \gamma_{r+1} \geq \dots \geq \gamma_{p+q-r} > \dots > \gamma_{p+q} \geq 0\}.$$

C1 $\chi(\cdot)$ is nondecreasing.

C2 $\chi(x)$ is left continuous for $x > 0$.

C3 $\chi(0) = 0$.

C4 χ is continuous in 0.

C5 $\lim_{x \rightarrow \infty} \chi(x) = 1$.

C6 There exists $c_0 \in (0, \infty)$ such that $\chi(x) < 1$ if $0 \leq x < c_0$.

C7 $\chi(x) = 1$ for $c_0 < x < \infty$, c_0 as in (C6).

In order to obtain the asymptotic behavior of the estimates, it is also assumed the following conditions regarding the model density f_0 as well as the parameters Σ_{xx} and Σ_{yy} , which are estimated before the SM-procedure applies.

C8 f_0 is nonincreasing, and there exist $\xi \leq \infty$ such that $f_0(x) > 0$ if $x < \xi$ and $f_0(x) = 0$ if $x > \xi$.

C9 Let ξ be as in C8. The functions $f_0(\cdot)$ and $\chi(\cdot)$ have at least a common point of strict monotonicity, that is, there exists $d < \xi$ and a nondegenerate interval I such that $d \in I$ and for all $u, v \in I$ with $u < d < v$ it holds that $\chi(u) < \chi(d) < \chi(v)$ and $f_0(u) > f_0(d) > f_0(v)$.

C10 Let ξ and d be as in C8 and C9 and Γ as in C0. Let $\{\lambda_1, \dots, \lambda_r\} \subset \Gamma$ such that $\lambda_j \geq \lambda_{j+1}$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$. Let $\sigma(\Lambda)$ be such that

$$\int \chi(\mathbf{w}^t \Lambda \mathbf{w} / \sigma(\Lambda)) f_0(\mathbf{w}^t \mathbf{w}) d\mathbf{w} = \delta$$

and $\sigma_0 = \min_{\Lambda} \{\sigma(\Lambda)\}$, then $d\sigma_0 / \gamma_{m-1}^0 < \xi$.

C11 a. Let $\mathbf{z}_1, \dots, \mathbf{z}_n$ be a random sample in \mathbb{R}^m from an elliptical distribution with density (18). Then, $\{\hat{\Sigma}_{xx}^{(R)}\}_{n=p+1}^{\infty} \subseteq S_p$ and $\{\hat{\Sigma}_{yy}^{(R)}\}_{n=q+1}^{\infty} \subseteq S_q$ based on $\mathbf{z}_1, \dots, \mathbf{z}_n$ are consistent estimators to Σ_{xx} and Σ_{yy} , respectively.

C11 b. If $(\mathbf{x}^t, \mathbf{y}^t)^t \sim F$ with density (18), then the multivariate dispersion functionals $\Sigma_{xx}^{(R)}(F) = \Sigma_{xx}$ and $\Sigma_{yy}^{(R)}(F) = \Sigma_{yy}$.

C0 allows for a simpler way of presenting the consistency because of having eigenspaces of dimension 1. The conditions C1–C7 regarding to the loss function χ are similar to the ones considered in the robust literature for redescending estimators (see Davies 1987; Maronna 2005; Maronna et al. 2019). C1 keeps the monotonicity

displayed by the square loss function in the classical case, letting larger residuals have larger weights. C2 allows for the minimum rather than the infimum in the definition of M-scale given in (5). C3 stands for the fact that a zero residual has zero weight. C4 stands for the intuitive fact that residuals coming smaller and smaller cannot have positive weights above a threshold while C3 holds. C5 summarizes the fact of getting a bounded χ , which is instrumental to get robust estimators able to cope with a large proportion of outlying observations. C6 and C7 are required for technical reasons to derive the consistency. C8, C9 and C10 are crucial for the Fisher-consistency: they prevent from having other parameters different from the elliptical model parameters yielding the minimum in (15). C11 stands for the consistency and Fisher consistency for the preliminary estimators and functionals corresponding to the parameters Σ_{xx} and Σ_{yy} . The limitations regarding the loss function χ are not relevant to the practice, since they do not impose restrictions on the data distribution. The loss function affects the asymptotic properties as well as the robustness properties that the estimators will possess.

Next, a useful concept in the robustness literature is included, whose fulfillment is required to ensure that (12) can be solved properly. We say that a sample $\{z_1, \dots, z_n\}$ is in *r-general position* if any linear manifold $Z_n = \{z \in \mathbb{R}^m : Cz=a\}$ with $(C, a) \in \mathcal{B}_{r,m}^1$ has at most $m - r + 1$ points from the sample.

Lemma 1 *Let z_1, \dots, z_n be a random sample in \mathbb{R}^m from an elliptical distribution with density (18). Let us suppose that conditions C1-C7 hold. If the sample $\{z_1, \dots, z_n\}$ is in r-general position and $m - r + 1 < n(1 - \delta)$, then (12) has at least one solution with probability 1.*

To show the consistency of a sequence of SM-estimators, Theorem 4.2, p. 665 of Rao (1962) is required. A useful generalization of this result is as follows.

Lemma 2 *Let z_1, \dots, z_n be a random sample in \mathbb{R}^m from an elliptical distribution F with density (18). Let us suppose that conditions C1-C7 hold. Let $S_{m,o} \subset \mathbb{R}^{m \times m}$ be the set of nonnegative symmetric matrices and F_n stands for the empirical distribution function based on the sample z_1, \dots, z_n . Then, it holds that*

$$P \left(\lim_{n \rightarrow \infty} \sup_{\substack{G \in S_{m,o} \\ \mu \in \mathbb{R}^m \\ \sigma \in (0, \infty)}} \left| E_{F_n} \chi \left(\frac{(z - \mu)^t G (z - \mu)}{\sigma} \right) - E_F \chi \left(\frac{(z - \mu)^t G (z - \mu)}{\sigma} \right) \right| = 0 \right) = 1.$$

In what follows, let $\hat{\Sigma}_{xx}^{(R)}$ and $\hat{\Sigma}_{yy}^{(R)}$ be consistent estimators to Σ_{xx} and Σ_{yy} , respectively. The Fisher consistency, the existence of the estimator for random samples and Lemma 2 entail the consistency of the SM-estimators, and it is omitted.

Theorem 1 *Let z_1, \dots, z_n be a random sample in \mathbb{R}^m from an elliptical distribution F with density (18). Suppose that conditions C0-C11a. hold and $m - r + 1 < n(1 - \delta)$. Let $(\hat{A}_{SM}^o, \hat{B}_{SM}^o, \hat{a}_{SM}^o)$ be solutions of (16), then*

$$\lim_{n \rightarrow \infty} (\hat{A}_{SM}^o, \hat{B}_{SM}^o, \hat{a}_{SM}^o) = (A_o, B_o, a_o) \text{ a. s.,}$$

$$\lim_{n \rightarrow \infty} \left(\hat{A}_{SM}^o \left[\hat{\Sigma}_{xx}^{(R)} \right]^{-1/2}, \hat{B}_{SM}^o \left[\hat{\Sigma}_{yy}^{(R)} \right]^{-1/2}, \hat{a}_{SM}^o \right) = \left(A_o \Sigma_{xx}^{-1/2}, B_o \Sigma_{yy}^{-1/2}, a_o \right) \text{ a. s.}$$

Proof It is deferred to the ‘‘Appendix’’. □

The previous theorem has shown the consistency of the vectors minimizing the M-scale. To derive the asymptotic behavior of the SM-estimators, we consider the critical points obtained from the constraint minimization

$$h(\mathbf{A}, \mathbf{B}, \mathbf{a}, \boldsymbol{\Theta}, \boldsymbol{\Xi}) = \sigma(\mathbf{A}, \mathbf{B}, \mathbf{a}) + tr(\mathbf{A}\mathbf{A}^t \boldsymbol{\Xi}^t) + tr(\mathbf{B}\mathbf{B}^t \boldsymbol{\Theta}^t) - tr(\boldsymbol{\Theta}^t) - tr(\boldsymbol{\Xi}^t) \tag{19}$$

with $\boldsymbol{\Theta}, \boldsymbol{\Xi} \in \mathbb{R}^{r \times r}$. Thus, we take a Lagrangian whose set of critical points contain the critical points of $h(\mathbf{A}, \mathbf{B}, \mathbf{a}, \boldsymbol{\Theta}, \boldsymbol{\Xi})$ in (19), that is, we consider

$$\sigma(\mathbf{D}, \mathbf{a}) + tr(\mathbf{D}\mathbf{D}^t \boldsymbol{\Lambda}^t) - tr(\boldsymbol{\Lambda}^t), \mathbf{D} = \frac{1}{\sqrt{2}} (\mathbf{A} - \mathbf{B}) \in \mathcal{B}_{r,m}^1, \boldsymbol{\Lambda} \in \mathbb{R}^{r \times r}. \tag{20}$$

Let us introduce some notation for the covariance matrices later used in the derivation of asymptotics. Set the parameters $\tilde{\Sigma} = diag(\boldsymbol{\Xi}^{-1/2}, \boldsymbol{\Theta}^{-1/2})$ with $\boldsymbol{\Xi} \in S_p$ and $\boldsymbol{\Theta} \in S_q$. Let us take the functionals $\tilde{\Sigma}(H) = diag\left(\left[\Sigma_{xx}^{(R)}(H)\right]^{-1/2}, \left[\Sigma_{yy}^{(R)}(H)\right]^{-1/2}\right)$ and $\tilde{\Sigma}_\varepsilon = \tilde{\Sigma}(F_\varepsilon)$. Consider the estimators $\hat{\tilde{\Sigma}} = \tilde{\Sigma}(F_n)$, $\Sigma_{xx}^{(R)}(F_n) = \hat{\Sigma}_{xx}^{(R)}$ and $\Sigma_{yy}^{(R)}(F_n) = \hat{\Sigma}_{yy}^{(R)}$. If we are dealing with Fisher consistent functionals, we have that $\tilde{\Sigma}(F) = \tilde{\Sigma}_o = diag(\Sigma_{xx}^{-1/2}, \Sigma_{yy}^{-1/2})$.

Take $\tilde{\mathbf{z}} = (\tilde{\mathbf{x}}^t, \tilde{\mathbf{y}}^t)^t$, by deriving (20) we get the equivalent system,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \chi' \left(\frac{\|\mathbf{D}\tilde{\mathbf{z}}_i - \mathbf{a}\|^2}{\sigma} \right) (\mathbf{D}\tilde{\mathbf{z}}_i - \mathbf{a}) = \mathbf{0}_r, \\ & [\mathbf{I}_m - \mathbf{D}^t \mathbf{D}] \frac{1}{n} \sum_{i=1}^n \chi' \left(\frac{\|\mathbf{D}\tilde{\mathbf{z}}_i - \mathbf{a}\|^2}{\sigma} \right) \tilde{\mathbf{z}}_i (\mathbf{D}\tilde{\mathbf{z}}_i - \mathbf{a})^t = \mathbf{0}_{m \times r}. \end{aligned} \tag{21}$$

In case of having a sequence of consistent estimators initializing an iterative procedure to come up with a sequence of critical points solving (21), we can ensure that the sequence of critical points is also consistent in the CCA context. This is a result also available for some robust regression methods, and the proof is similar to that of those procedures [see, for instance, Theorem 3.2 in Yohai (1987), Theorem 4.1 in Yohai and Zamar (1988)].

Proposition 1 *Let z_1, \dots, z_n be a random sample in \mathbb{R}^m from an elliptical distribution F with density (18). Suppose that conditions C0-C10 hold, moreover χ is twice differentiable and concave in $[0, c_0]$, c_0 given in C6. If $m - r + 1 < n(1 - \delta)$, $(\hat{A}_o^{(0)}, \hat{B}_o^{(0)}, \hat{a}^{(0)})$ is a sequence of consistent estimators for (A_o, B_o, a_o) and the solutions of (21) $(\hat{A}^o, \hat{B}^o, \hat{a}^o)$ verify that $\hat{\sigma}(\hat{A}^o, \hat{B}^o, \hat{a}^o) \leq \hat{\sigma}(\hat{A}_o^{(0)}, \hat{B}_o^{(0)}, \hat{a}^{(0)})$, then $\lim_{n \rightarrow \infty} (\hat{A}^o, \hat{B}^o, \hat{a}^o) = (A_o, B_o, a_o)$ almost surely.*

3 Asymptotic behavior of SM-estimators for CCA

3.1 Notation

To deal with the asymptotic behavior of the SM-estimators, let us introduce some notation. The parameters in our setting are taken from the sets

$$\begin{aligned} \mathcal{G} &= \mathcal{O}_{r,p} \times \mathcal{O}_{r,q} \times \mathbb{R}^r \times S_p \times S_q \times (0, \infty), \\ \overline{\mathcal{G}} &= \mathbb{R}^{r \times m} \times \mathbb{R}^r \times S_p \times S_q \times (0, \infty), \end{aligned} \tag{22}$$

The asymptotic distribution of SM-estimators and its influence curve, which is a useful tool to evaluate the behavior under infinitesimal contaminations, need to be established over some set of distributions. Let \mathcal{D} be the set of distributions on \mathbb{R}^m , and G_n stands for the empirical distribution function based on n points in \mathbb{R}^m . Call \mathcal{F}_n the subset of such distributions. Thus, let us define the following subsets of distributions,

$$\begin{aligned} \mathcal{E} &= \{F \in \mathcal{D} : F \text{ is elliptical}\}, \\ C_\varepsilon(\mathcal{E}, \mathcal{F}_1) &= \{G \in \mathcal{D} : G = (1 - \varepsilon)F + \varepsilon\delta_{\mathbf{z}}, F \in \mathcal{E}, \delta_{\mathbf{z}} \in \mathcal{F}_1, \mathbf{z} \in \mathbb{R}^m\}, \varepsilon \in [0, 1]. \end{aligned}$$

Let \mathcal{H} be a subset of \mathcal{D} such that $\mathcal{E} \cup \mathcal{F}_n \cup C_\varepsilon(\mathcal{E}, \mathcal{F}_1) \subset \mathcal{H}$. Then, the SM-functional $\tilde{\theta} : \mathcal{H} \rightarrow \mathcal{G}$ is defined as

$$\tilde{\theta}(H) = (\mathbf{A}_{SM}^o(H), \mathbf{B}_{SM}^o(H), \mathbf{a}_{SM}^o(H), \boldsymbol{\Sigma}_{xx}^{(R)}(H), \boldsymbol{\Sigma}_{yy}^{(R)}(H), \sigma_o(H)).$$

If $vec : \mathbb{R}^{m \times r} \rightarrow \mathbb{R}^{mr}$ stands for the operator which vectorizes a matrix by stacking the columns on top of each other, we establish some useful notation corresponding to parameters. Set $\tilde{\theta} = (\mathbf{A}, \mathbf{B}, \mathbf{a}, \boldsymbol{\Sigma}, \sigma)$ and $\theta = \left([vec(\mathbf{D}^t)]^t, \mathbf{a}^t \right)^t$. The SM-estimators turn out to be $\hat{\mathbf{D}}_{SM} = \frac{1}{\sqrt{2}} \left(\hat{\mathbf{A}}_{SM}^o - \hat{\mathbf{B}}_{SM}^o \right)$ and $\hat{\theta} = \tilde{\theta}(F_n) = \left(\hat{\mathbf{A}}_{SM}^o, \hat{\mathbf{B}}_{SM}^o, \hat{\mathbf{a}}_{SM}, \hat{\boldsymbol{\Sigma}}_{xx}^{(R)}, \hat{\boldsymbol{\Sigma}}_{yy}^{(R)}, \hat{\sigma} \right)$, with $F_n \in \mathcal{F}_n$. Let us now denote some functionals, $\mathbf{D}_{SM}(\cdot) = 1/\sqrt{2} \left(\mathbf{A}_{SM}^o(\cdot) - \mathbf{B}_{SM}^o(\cdot) \right)$, $\mathbf{D}_\varepsilon = \mathbf{D}_{SM}(F_\varepsilon)$, $\tilde{\theta}(F_\varepsilon) = \tilde{\theta}_\varepsilon$, with $F_\varepsilon \in C_\varepsilon(\mathcal{E}, \mathcal{F}_1)$, $\tilde{\theta}_o = (\mathbf{A}_o, \mathbf{B}_o, \mathbf{a}_o, \boldsymbol{\Sigma}_{xx}, \boldsymbol{\Sigma}_{yy}, \sigma_o)$ and $\mathbf{D}_o = \frac{1}{\sqrt{2}} (\mathbf{A}_o - \mathbf{B}_o)$. Take $\theta_o = \left([vec(\mathbf{D}_o^t)]^t, \mathbf{a}_o^t \right)^t$, $\theta_{SM} = \left([vec(\mathbf{D}_{SM}^t(\cdot))]^t, \mathbf{a}_{SM}^t(\cdot) \right)^t$.

If $\mathbf{D}^* = \mathbf{D}\tilde{\boldsymbol{\Sigma}}$, set $\theta^* = \left(\left\{ vec \left[(\mathbf{D}^*)^t \right] \right\}^t, \mathbf{a}^t \right)^t$. If $\hat{\mathbf{D}}_{SM}^* = \hat{\mathbf{D}}_{SM} \hat{\tilde{\boldsymbol{\Sigma}}}$, set

$$\hat{\theta}_{SM} = \left(\left[vec \left(\hat{\mathbf{D}}_{SM}^t \right) \right]^t, \hat{\mathbf{a}}_{SM}^t \right)^t \tag{23}$$

$$\hat{\theta}_{SM}^* = \left(\left\{ vec \left[\left(\hat{\mathbf{D}}_{SM}^* \right)^t \right] \right\}^t, \hat{\mathbf{a}}_{SM}^t \right)^t, \tag{24}$$

$$\hat{\mathbf{v}}_{SM,k} = \left(\hat{\mathbf{A}}_{SM} \right)^t \mathbf{f}_k^{(r)} \text{ and } \hat{\mathbf{w}}_{SM,k} = \left(\hat{\mathbf{B}}_{SM} \right)^t \mathbf{f}_k^{(r)}, \text{ for } k = 1, \dots, r. \tag{25}$$

If $\mathbf{D}_o^* = \mathbf{D}_o \tilde{\Sigma}_o$, put $\boldsymbol{\theta}_o^* = \left(\left\{ \text{vec} \left[(\mathbf{D}_o^*)^t \right] \right\}^t, \mathbf{a}_o^t \right)^t$, $\mathbf{v}_k = \left(\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1/2} \mathbf{A}_o^t \right) \mathbf{f}_k^{(r)}$, and $\mathbf{w}_k = \left(\boldsymbol{\Sigma}_{\mathbf{yy}}^{-1/2} \mathbf{B}_o^t \right) \mathbf{f}_k^{(r)}$, $k = 1, \dots, r$. Then, we can rewrite (21) in the following way,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \chi' \left(\frac{\|\mathbf{D}^* \mathbf{z}_i - \mathbf{a}\|^2}{\sigma} \right) (\mathbf{D}^* \mathbf{z}_i - \mathbf{a}) = \mathbf{0}_r, \\ & \widehat{\Sigma}^{-1} \left[\widehat{\Sigma}^2 - (\mathbf{D}^*)^t \mathbf{D}^* \right] \frac{1}{n} \sum_{i=1}^n \chi' \left(\frac{\|\mathbf{D}^* \mathbf{z}_i - \mathbf{a}\|^2}{\sigma} \right) \mathbf{z}_i (\mathbf{D}^* \mathbf{z}_i - \mathbf{a})^t = \mathbf{0}_{m \times r}. \end{aligned}$$

Then, let us take the functions $\phi_1 : \mathbb{R}^m \times \mathcal{G} \rightarrow \mathbb{R}^{rm}$, $\phi_2 : \mathbb{R}^m \times \mathcal{G} \rightarrow \mathbb{R}^r$, $\bar{\phi}_1 : \mathbb{R}^m \times \mathcal{G} \rightarrow \mathbb{R}^{rm}$, $\bar{\phi}_2 : \mathbb{R}^m \times \mathcal{G} \rightarrow \mathbb{R}^r$ and $\bar{\phi} : \mathbb{R}^m \times \mathcal{G} \rightarrow \mathbb{R}^{r(m+1)}$ as follows,

$$\begin{aligned} \phi_1(\mathbf{z}, \tilde{\boldsymbol{\theta}}) &= \chi' \left(\frac{\|\mathbf{D} \tilde{\Sigma} \mathbf{z} - \mathbf{a}\|^2}{\sigma} \right) \text{vec} \left[(\mathbf{I}_m - \mathbf{D}^t \mathbf{D}) \tilde{\Sigma} \mathbf{z} (\mathbf{D} \tilde{\Sigma} \mathbf{z} - \mathbf{a})^t \right] \\ \phi_2(\mathbf{z}, \tilde{\boldsymbol{\theta}}) &= \chi' \left(\frac{\|\mathbf{D} \tilde{\Sigma} \mathbf{z} - \mathbf{a}\|^2}{\sigma} \right) (\mathbf{D} \tilde{\Sigma} \mathbf{z} - \mathbf{a}), \quad \phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \end{aligned} \tag{26}$$

$$\begin{aligned} \bar{\phi}_1(\mathbf{z}, \mathbf{D}^*, \mathbf{a}, \sigma) &= \chi' \left(\frac{\|\mathbf{D}^* \mathbf{z} - \mathbf{a}\|^2}{\sigma} \right) \text{vec} \left[\tilde{\Sigma}^{-1} \left(\tilde{\Sigma}^2 - (\mathbf{D}^*)^t \mathbf{D}^* \right) \mathbf{z} (\mathbf{D}^* \mathbf{z} - \mathbf{a})^t \right] \\ \bar{\phi}_2(\mathbf{z}, \mathbf{D}^*, \mathbf{a}, \sigma) &= \chi' \left(\frac{\|\mathbf{D}^* \mathbf{z} - \mathbf{a}\|^2}{\sigma} \right) (\mathbf{D}^* \mathbf{z} - \mathbf{a}), \quad \bar{\phi} = \begin{pmatrix} \bar{\phi}_1 \\ \bar{\phi}_2 \end{pmatrix} \end{aligned}$$

The corresponding expected values and covariance matrices are denoted by

$$\begin{aligned} & \Phi_1 : \mathcal{G} \rightarrow \mathbb{R}^r, \Phi_2 : \mathcal{G} \rightarrow \mathbb{R}^{rm} \text{ and } \Phi : \mathcal{G} \rightarrow \mathbb{R}^{r(m+1)} \\ & \bar{\Phi}_1 : \bar{\mathcal{G}} \rightarrow \mathbb{R}^r, \bar{\Phi}_2 : \bar{\mathcal{G}} \rightarrow \mathbb{R}^{rm} \text{ and } \bar{\Phi} : \bar{\mathcal{G}} \rightarrow \mathbb{R}^{r(m+1)} \\ & \Phi_h(\tilde{\boldsymbol{\theta}}) = E_F \left(\phi_h(\mathbf{z}, \tilde{\boldsymbol{\theta}}) \right), h = 1, 2; \Phi(\tilde{\boldsymbol{\theta}}) = E_F \phi(\mathbf{z}, \tilde{\boldsymbol{\theta}}) = \left(\Phi_1^t(\tilde{\boldsymbol{\theta}}), \Phi_2^t(\tilde{\boldsymbol{\theta}}) \right)^t \\ & \bar{\Phi}_h(\tilde{\boldsymbol{\theta}}) = E_F \left(\bar{\phi}_h(\mathbf{z}, \tilde{\boldsymbol{\theta}}) \right), h = 1, 2; \bar{\Phi}(\tilde{\boldsymbol{\theta}}) = E_F \bar{\phi}(\mathbf{z}, \tilde{\boldsymbol{\theta}}) = \left(\bar{\Phi}_1^t(\tilde{\boldsymbol{\theta}}), \bar{\Phi}_2^t(\tilde{\boldsymbol{\theta}}) \right)^t \\ & V(\tilde{\boldsymbol{\theta}}) = E_F \phi(\mathbf{z}, \tilde{\boldsymbol{\theta}}) \left(\phi(\mathbf{z}, \tilde{\boldsymbol{\theta}}) \right)^t \text{ and } V_o = V(\tilde{\boldsymbol{\theta}}_o). \end{aligned}$$

3.2 Conditions, asymptotic normality and variances

To deal with the asymptotic behavior of SM-estimators, let us introduce some extra conditions.

C12 $\chi : [0, \infty) \rightarrow [0, 1]$ is twice continuously differentiable.

C13 $E_F [\|\mathbf{z}\|^8] < \infty$.

C14 $\Phi : \mathbb{R}^{r(m+1)} \rightarrow \mathbb{R}^{r(m+1)}$ and $\bar{\Phi} : \mathbb{R}^{r(m+1)} \rightarrow \mathbb{R}^{r(m+1)}$ are continuously differentiable and the matrices

$$\mathbf{\Omega} = \frac{\partial \Phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_o) \in \mathbb{R}^{r(m+1) \times r(m+1)} \text{ and } \mathbf{\Omega}^* = \frac{\partial \bar{\Phi}(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}^*}(\tilde{\boldsymbol{\theta}}_o) \in \mathbb{R}^{r(m+1) \times r(m+1)}$$

are nonsingular.

The usual way to derive the asymptotic normality of an estimator obtained as zero of a equation, $\mathbf{0} = \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{z}_i, \hat{\boldsymbol{\theta}})$ is to consider a Taylor expansion of the form

$$\Phi(\tilde{\boldsymbol{\theta}}) = \Phi(\tilde{\boldsymbol{\theta}}_o) + \frac{\partial \Phi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}(\tilde{\boldsymbol{\theta}}_o)(\boldsymbol{\theta} - \boldsymbol{\theta}_o) + R(\tilde{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\theta}_o), \tag{27}$$

where $R(\tilde{\boldsymbol{\theta}}) \rightarrow 0$ as $\tilde{\boldsymbol{\theta}} \rightarrow \tilde{\boldsymbol{\theta}}_o$. After some manipulations, by summing up and subtracting some terms, we come up with an expression of the form

$$\mathbf{0} = \Phi(\hat{\boldsymbol{\theta}}) - \Phi(\tilde{\boldsymbol{\theta}}_o) + \mathbf{Z}_n + \mathbf{W}_n, \tag{28}$$

where $\sqrt{n}\mathbf{Z}_n$ converges in distribution to a multivariate normal distribution with $\mathbf{0}$ mean and covariance matrix $E\phi\phi^t$ and $\mathbf{W}_n = o_P(1/\sqrt{n})$ by arguments from the empirical processes theory. C13 is a restrictive condition on the model to obtain that the functions given in (26) belong to a Euclidean class (Pakes and Pollard 1989) and obtain the behavior of \mathbf{W}_n . By joining (27) and (28), we get the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}_{SM} - \boldsymbol{\theta}_o)$. This brief account for the asymptotic normality makes clearer the need for conditions C12 and C14. Kudraszow and Maronna (2010, 2011) used a similar approach to derive the asymptotic distribution of MM-estimators for the multivariate regression model.

If F is elliptically countered with density (18), let us call $\mathbf{r}_o(\mathbf{z}) = \mathbf{D}_o^* \mathbf{z} - \mathbf{a}_o$, $\mathbf{C} = \sum_{1 \leq i, k \leq r} [\mathbf{f}_i^{(r)}(\mathbf{f}_k^{(r)})^t \otimes \mathbf{t}_i \mathbf{t}_k^t]$, $\mathbf{A}_o = \text{diag}(\gamma_1^0, \dots, \gamma_r^0)$, $\mathbf{P}_r = \mathbf{I}_m - \mathbf{D}_o' \mathbf{D}_o$. Then, under C12 and C14, $\mathbf{\Omega}$ and $\mathbf{\Omega}^*$ turn out to be

$$\mathbf{\Omega} = \begin{pmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{pmatrix} \text{ and } \mathbf{\Omega}_o^* = \begin{pmatrix} \mathbf{\Omega}_{11} \left(\mathbf{I}_r \otimes (\tilde{\boldsymbol{\Sigma}}_o)^{-1} \right) & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} \left(\mathbf{I}_r \otimes (\tilde{\boldsymbol{\Sigma}}_o)^{-1} \right) & \mathbf{\Omega}_{22} \end{pmatrix}, \tag{29}$$

with $\mathbf{\Omega}_{11} \in \mathbb{R}^{rm \times rm}$, $\mathbf{\Omega}_{12} \in \mathbb{R}^{rm \times r}$, $\mathbf{\Omega}_{21} \in \mathbb{R}^{r \times rm}$ and $\mathbf{\Omega}_{22} \in \mathbb{R}^{r \times r}$. Take the matrix

$$\boldsymbol{\Psi}(\mathbf{z}) = \frac{2}{\sigma_o} \chi'' \left(\frac{\|\mathbf{r}_o(\mathbf{z})\|^2}{\sigma_o} \right) \mathbf{r}_o(\mathbf{z}) \mathbf{r}_o^t(\mathbf{z}) + \chi' \left(\frac{\|\mathbf{r}_o(\mathbf{z})\|^2}{\sigma_o} \right) \mathbf{I}_r.$$

Thus, we get

$$\begin{aligned} \mathbf{\Omega}_{11} &= (\mathbf{I}_r \otimes \mathbf{P}_r) E_F \left[\boldsymbol{\Psi}(\mathbf{z}) \otimes \left(\tilde{\boldsymbol{\Sigma}}_o \mathbf{z} \right) \left(\tilde{\boldsymbol{\Sigma}}_o \mathbf{z} \right)^t \right] - (\mathbf{A}_o \otimes \mathbf{I}_m) (\mathbf{I}_{rm} + \mathbf{C}), \\ \mathbf{\Omega}_{22} &= -E_F \boldsymbol{\Psi}(\mathbf{z}), \\ \mathbf{\Omega}_{12} &= -(\mathbf{I}_r \otimes \mathbf{P}_r) E_F \left[\boldsymbol{\Psi}(\mathbf{z}) \otimes \left(\tilde{\boldsymbol{\Sigma}}_o \mathbf{z} \right) \right], \\ \mathbf{\Omega}_{21} &= E_F \left[\boldsymbol{\Psi}(\mathbf{z}) \otimes \left(\tilde{\boldsymbol{\Sigma}}_o \mathbf{z} \right)^t \right]. \end{aligned}$$

The matrix $\mathbf{V}_o = E_F \phi(\mathbf{z}, \tilde{\boldsymbol{\theta}}_o) \left(\phi(\mathbf{z}, \tilde{\boldsymbol{\theta}}_o) \right)^t$ turns out to be $\mathbf{V}_o = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}$, with $\mathbf{V}_{11} \in \mathbb{R}^{rm \times rm}$, $\mathbf{V}_{12} \in \mathbb{R}^{rm \times r}$ and $\mathbf{V}_{22} \in \mathbb{R}^{r \times r}$, that is,

$$\begin{aligned} \mathbf{V}_{11} &= E_F \left\{ \left[\chi' \left(\frac{\|\mathbf{r}_o(\mathbf{z})\|^2}{\sigma_o} \right) \mathbf{r}_o(\mathbf{z}) \mathbf{r}_o^t(\mathbf{z}) \right] \otimes \left[\mathbf{P}_r \left(\tilde{\boldsymbol{\Sigma}}_o \mathbf{z} \right) \left(\tilde{\boldsymbol{\Sigma}}_o \mathbf{z} \right)^t \mathbf{P}_r \right] \right\}, \\ \mathbf{V}_{22} &= E_F \left[\chi' \left(\frac{\|\mathbf{r}_o(\mathbf{z})\|^2}{\sigma_o} \right) \mathbf{r}_o(\mathbf{z}) \mathbf{r}_o^t(\mathbf{z}) \right], \\ \mathbf{V}_{12} &= E_F \left\{ \left[\chi' \left(\frac{\|\mathbf{r}_o(\mathbf{z})\|^2}{\sigma_o} \right) \mathbf{r}_o(\mathbf{z}) \mathbf{r}_o^t(\mathbf{z}) \right] \otimes \left[\mathbf{P}_r \left(\tilde{\boldsymbol{\Sigma}}_o \mathbf{z} \right) \right] \right\}. \end{aligned}$$

The derivation of the asymptotic behavior of the SM-estimator is based on the empirical processes theory (see Pakes and Pollard 1989) and the following theorem establishes the convergence in distribution for the SM-estimators.

Theorem 2 *Let $\mathbf{z}_1, \dots, \mathbf{z}_n$ be a random sample in \mathbb{R}^m from an elliptical distribution F with density (18), location parameter $\boldsymbol{\mu}_o$ and dispersion parameter $\boldsymbol{\Sigma}_o$. Suppose that conditions C0-C14 hold. If $\hat{\boldsymbol{\theta}}_{SM}$ as in (23), $\hat{\boldsymbol{\theta}}_{SM}^*$ as in (24), $\hat{\mathbf{v}}_{SM,k}$ and $\hat{\mathbf{w}}_{SM,k}$ as in (25) are sequences of SM-estimators and $\xrightarrow{\mathcal{D}}$ stands for convergence in distribution, then it holds that*

- (i) $\sqrt{n} \left(\hat{\boldsymbol{\theta}}_{SM} - \boldsymbol{\theta}_o \right) \xrightarrow{\mathcal{D}} N_{r(m+1)} \left(\mathbf{0}, \mathbf{V} \right)$, where the asymptotic covariance matrix is given by $\mathbf{V} = \boldsymbol{\Omega}^{-1} \mathbf{V}_o \boldsymbol{\Omega}^{-t}$,
- (ii) $\sqrt{n} \left(\hat{\boldsymbol{\theta}}_{SM}^* - \boldsymbol{\theta}_o^* \right) \xrightarrow{\mathcal{D}} N_{r(m+1)} \left(\mathbf{0}, \mathbf{V}^* \right)$, with the asymptotic covariance matrix given by $\mathbf{V}^* = \left(\boldsymbol{\Omega}^* \right)^{-1} \mathbf{V}_o \left(\boldsymbol{\Omega}^* \right)^{-t}$,
- (iii) $\sqrt{n} \left(\hat{\mathbf{v}}_{SM,k} - \mathbf{v}_k \right) \xrightarrow{\mathcal{D}} N_p \left(\mathbf{0}, \mathbf{V}_{\alpha,k}^* \right)$ and $\sqrt{n} \left(\hat{\mathbf{w}}_{SM,k} - \mathbf{w}_k \right) \xrightarrow{\mathcal{D}} N_q \left(\mathbf{0}, \mathbf{V}_{\beta,k}^* \right)$, for $k = 1, \dots, r$, with

$$\mathbf{V}_{\alpha,k}^* = \sum_{i,j=1}^p \mathbf{f}_j^{(p)} \mathbf{f}_{(k-1)m+j}^{(m)t} \mathbf{V}^* \mathbf{f}_{(k-1)m+i}^{(m)} \left(\mathbf{f}_i^{(p)} \right)^t,$$

$$\mathbf{V}_{\beta,k}^* = \sum_{i,j=1}^q \mathbf{f}_j^{(q)} \mathbf{f}_{km-q+j}^{(m)t} \mathbf{V}^* \mathbf{f}_{km-q+i}^{(m)} \left(\mathbf{f}_i^{(q)}\right)^t.$$

(iv) If $\mathbf{\Omega}_{22}$ is a nonsingular matrix, then

$$\mathbf{V}_{\alpha,k}^* = \mathbf{\Sigma}_{xx}^{-1/2} \mathbf{V}_{\alpha,k} \mathbf{\Sigma}_{xx}^{-1/2}, \text{ with } \mathbf{V}_{\alpha,k} = \sum_{i,j=1}^p \mathbf{f}_j^{(p)} \left(\mathbf{f}_{(k-1)m+j}^{(m)}\right)^t \mathbf{V} \mathbf{f}_{(k-1)m+i}^{(m)} \left(\mathbf{f}_i^{(p)}\right)^t$$

$$\mathbf{V}_{\beta,k}^* = \mathbf{\Sigma}_{yy}^{-1/2} \mathbf{V}_{\beta,k} \mathbf{\Sigma}_{yy}^{-1/2}, \text{ with } \mathbf{V}_{\beta,k} = \sum_{i,j=1}^q \mathbf{f}_j^{(q)} \left(\mathbf{f}_{km-q+j}^{(m)}\right)^t \mathbf{V} \mathbf{f}_{km-q+i}^{(m)} \left(\mathbf{f}_i^{(q)}\right)^t.$$

Proof It is deferred to the ‘‘Appendix’’. □

Remark 1 The asymptotics for the SM-estimators in the PCA case given in (5) derive in a similar manner to that of (i) in Theorem 2.

3.3 Asymptotic relative efficiency in the Gaussian case

In order to assess the loss of efficiency under normality, the asymptotic covariance of the SM-estimators is compared with that of the maximum likelihood estimator under normality. Let $\hat{\mathbf{v}}_{SM,k}$ and $\hat{\mathbf{w}}_{SM,k}$ be the k -th canonical vectors as in (25) and $\hat{\mathbf{v}}_{C,k}$ and $\hat{\mathbf{w}}_{C,k}$ the classical estimators obtained by solving (7) and (8) with the population covariances replaced by the empirical ones. Let $\mathbf{V}_{\alpha,k}^*(F)$ and $\mathbf{V}_{\alpha,k}^{*,C}(F)$ be the asymptotic variances of $\hat{\mathbf{v}}_{SM,k}$ and $\hat{\mathbf{v}}_{C,k}$ when the underlying distribution is F ($\mathbf{V}_{\beta,k}^*(F)$ and $\mathbf{V}_{\beta,k}^{*,C}(F)$ for $\hat{\mathbf{w}}_{SM,k}$ and $\hat{\mathbf{w}}_{C,k}$, respectively). $\mathbf{V}_{\alpha,k}^{*,C}(F)$ can be derived from the asymptotic distribution for SM-estimators stated above by taking χ as the identity function. Then, a measure of asymptotic relative efficiency is given by

$$ARE(\hat{\mathbf{v}}_{SM,k}, F) = \frac{\text{trace}\left(\mathbf{V}_{\alpha,k}^{*,C}(F)\right)}{\text{trace}\left(\mathbf{V}_{\alpha,k}^*(F)\right)}, \quad ARE(\hat{\mathbf{w}}_{SM,k}, F) = \frac{\text{trace}\left(\mathbf{V}_{\beta,k}^{*,C}(F)\right)}{\text{trace}\left(\mathbf{V}_{\beta,k}^*(F)\right)}.$$

To illustrate the behavior of the SM-estimators with respect to the classical estimator, we take $F = N_4(\mathbf{0}, \mathbf{\Sigma})$, with the partitioned matrix $\mathbf{\Sigma}$ given in (6) and $\mathbf{\Sigma}_{xx} = \mathbf{\Sigma}_{yy} = \mathbf{I}_2$, $\mathbf{\Sigma}_{xy} = \text{diag}(0.9, 0.5)$. Since $\mathbf{\Sigma}_{xx} = \mathbf{\Sigma}_{yy} = \mathbf{I}_2$, we have that $\mathbf{\Omega} = \mathbf{\Omega}^*$ and $\mathbf{V}^* = \mathbf{V}$. Thus, the efficiency for SM-estimators for canonical vectors coincides with that of the standardized SM-estimators.

Table 1 displays the asymptotic relative efficiency of $\hat{\mathbf{v}}_{SM,k}$ and $\hat{\mathbf{w}}_{SM,k}$ for different values of δ in the grid $G = \{0.5, 0.45, 0.4, 0.35, 0.3\}$. The larger the parameter δ , the smaller the relative efficiency, since the parameter δ is related to the robustness of the estimator whose breakdown point increases with δ (Adrover and Donato 2015). The second canonical vector seems to be much more affected by the trade-off between high breakdown point and efficiency.

Table 1 Asymptotic relative efficiency for the SM-estimators of canonical vectors as function of the parameter δ

δ	0.50	0.45	0.40	0.35	0.30
$ARE(\hat{v}_{SM,1}, F)$	0.734	0.807	0.871	0.926	0.973
$ARE(\hat{w}_{SM,1}, F)$	0.734	0.807	0.870	0.925	0.970
$ARE(\hat{v}_{SM,2}, F)$	0.236	0.398	0.580	0.772	0.965
$ARE(\hat{w}_{SM,2}, F)$	0.236	0.398	0.579	0.771	0.963

4 Robustness measures: qualitative robustness and influence function

4.1 Qualitative robustness

The qualitative robustness was introduced by Hampel (1971) and the concept captures the desirable fact that a robust estimator based on random samples from close distributions should induce close distributions either for any sample size. More precisely, a sequence of estimators $\{T_n\}_{n=1}^\infty$ is qualitatively robust in the probability measure F if and only if $\forall \varepsilon > 0 \exists \delta > 0$ and $\forall G, \forall n$, it holds that $d_{PR}(F, G) < \delta \implies d_{PR}(\mathcal{L}_F(T_n), \mathcal{L}_G(T_n)) < \varepsilon$, where d_{PR} stands for the Prohorov distance between probability measures and $\mathcal{L}_F(T_n)$ is the distribution induced by T_n as the random sample comes from F .

A related concept is the continuity of estimators. Hampel (1971) proves that estimators $\{T_n\}_{n=1}^\infty$ obtained from a continuous functional on the class of empirical distributions \mathcal{F}_n are qualitatively robust. A sequence of estimators $\{T_n\}_{n=1}^\infty$ is *continuous in the probability measure F* if and only if for all $\varepsilon > 0$ there exists $\delta > 0$ such that there exists n_o and for all $n, m \geq n_o$ it holds that $F_n \in \mathcal{F}_n, F_m \in \mathcal{F}_m, d_{PR}(F, F_n) < \delta, d_{PR}(F, F_m) < \delta \implies |T(F_n) - T(F_m)| < \varepsilon$.

Then, we proceed similarly to the proof for Theorem 1 to derive the continuity of the SM-functionals.

Theorem 3 Let $z = (x^t, y^t)^t \in \mathbb{R}^m$ be from an elliptical distribution with density (18). Suppose that conditions C0-C10 hold. Let us take any sequence of distributions $\{F_n\}_{n=1}^\infty$ such that F_n weakly converges to F . Assume that the dispersion functionals $\Sigma_{xx}^{(R)}$ and $\Sigma_{yy}^{(R)}$ are continuous in F . Let $(A_{SM}^o, B_{SM}^o, \alpha_{SM}^o)$ be the SM functional for CCA defined in (15). Therefore, it holds that $\lim_{n \rightarrow \infty} (A_{SM}^o(F_n), B_{SM}^o(F_n), \alpha_{SM}^o(F_n)) = (A_{SM}^o(F), B_{SM}^o(F), \alpha_{SM}^o(F))$.

4.2 Influence functions

To quantify the effect of contamination in the estimators and make comparisons among different proposals, the IF measures the rate of change because of infinitesimal effect provided by point masses, that is, for every $z_0 \in \mathbb{R}^m, m \geq 1$, we have

$$IF(z_0, T, F) = \lim_{\varepsilon \rightarrow 0^+} \frac{T((1 - \varepsilon)F + \varepsilon\delta_{z_0}) - T(F)}{\varepsilon} = \left. \frac{\partial}{\partial \varepsilon} T(F_\varepsilon) \right|_{\varepsilon=0},$$

with $F_\varepsilon = (1 - \varepsilon)F + \varepsilon\delta_{\mathbf{z}_0}$ and $\delta_{\mathbf{z}_0}$ is a point mass distribution at \mathbf{z}_0 .

Adrover and Donato (2015) treated two different versions for a robust canonical correlation. The matrix $\mathbf{M} = \sum_{i=1}^{p+q} \gamma_i^0 \mathbf{t}_i^0 (\mathbf{t}_i^0)^t$ in (13) has its eigenvalues in decreasing order. The $k - th$ canonical correlation is given by $\rho_k^2 = (\gamma_{m-k+1}^0 - 1)^2$. To construct the SM-functional for the square of the $k - th$ canonical correlation, we take

$$r(\mathbf{z}, H) = \frac{\|\mathbf{D}_{SM}(H) \tilde{\Sigma}(H) \mathbf{z} - \mathbf{a}_{SM}\|^2}{\sigma(H)}, \quad \tilde{\mu}(H) = \frac{E_H [\chi'(r(\mathbf{z}, H)) \tilde{\Sigma}(H) \mathbf{z}]}{E_H [\chi'(r(\mathbf{z}, H))]}.$$

Given $R = E_H \left[\chi'(r(\mathbf{z}, H)) (\tilde{\Sigma}(H) \mathbf{z} - \tilde{\mu}(H)) (\tilde{\Sigma}(H) \mathbf{z} - \tilde{\mu}(H))^t \right]$, we call \mathbf{M}_{xy}^0 the matrix obtained after pre- and post-multiplying R by $(\mathbf{I}_p \ \mathbf{0}_{p \times q})$ and $(\mathbf{0}_{q \times p} \ \mathbf{I}_q)^t$. Set

$$\mathbf{M}^0(H) = \begin{pmatrix} \mathbf{I}_p & \mathbf{M}_{xy}^0(H) \\ (\mathbf{M}_{xy}^0(H))^t & \mathbf{I}_q \end{pmatrix},$$

and take $\{\gamma_j^0(H), \mathbf{t}_{SM,j}^0(H)\}_{j=1}^m$ the eigenvalues and eigenvectors of the functional $\mathbf{M}^0(H)$, that is, $\mathbf{M}^0(H) \mathbf{t}_{SM,m-k+1}^0(H) = \gamma_{m-k+1}^0(H) \mathbf{t}_{SM,m-k+1}^0(H)$. The SM- $k - th$ canonical correlation is given by $\rho_{SM,k}^2(H) = (\gamma_{m-k+1}^0(H) - 1)^2, k = 1, \dots, r$.

Another concept for measuring the association between canonical variates was given by Branco et al. (2005). Given $(\mathbf{x}^t, \mathbf{y}^t) \sim H$ and $(\mathbf{v}_k^t(H)\mathbf{x}, \mathbf{w}_k^t(H)\mathbf{y}) \sim H_k$, let us take a robust bivariate dispersion functional $\Sigma^{(R)}(H_k)$ whose i, j entry is $\sigma_{ij}^{(R)}(H_k), 1 \leq i, j \leq 2$. Thus, a robust correlation functional is easily obtained from $\Sigma^{(R)}(H_k)$,

$$\rho_{C,k}^2(H_k) = (\sigma_{12}^{(R)})^2 / \left[(\sigma_{11}^{(R)})^2(H_k) (\sigma_{22}^{(R)})^2(H_k) \right]. \tag{30}$$

Let us define some quantities related to the derivation of the IF. Given $\mathbf{z}_0 = (\mathbf{x}_0^t, \mathbf{y}_0^t)^t \in \mathbb{R}^{p+q}$ and $\|\mathbf{r}_o(\mathbf{z})\| \sim \tilde{F}$ if $\mathbf{z} \sim F$, let σ be an M-scale functional based on $\|\mathbf{r}_o(\mathbf{z})\|$ defined in (5) and (14) with influence function $IFS = IF(\|\mathbf{r}_o(\mathbf{z}_0)\|, \sigma, \tilde{F})$. Let $\tilde{\Sigma} \in \mathbb{R}^{m \times m}$ be a functional with $IFD = IF(\mathbf{z}_0, \text{vec}(\tilde{\Sigma}), F)$. Put

$$O(\tilde{\Sigma}_o, \mathbf{D}_o, \mathbf{a}_o, \mathbf{z}) = \left((\text{vec}(\tilde{\Sigma}_o))^t (\mathbf{D}_o^t \mathbf{D}_o \otimes \mathbf{z}\mathbf{z}^t) - (\text{vec}(\mathbf{z}\mathbf{a}_o^t \mathbf{D}_o))^t \right),$$

$$\xi_o = E_F \left[\chi'(\|\mathbf{r}_o(\mathbf{z})\|^2 / \sigma_o) \|\mathbf{r}_o(\mathbf{z})\|^2 \right],$$

and

$$\begin{aligned}
 \mathbf{T}_1 &= E_F \left[\chi'' \left(\|\mathbf{r}_o(\mathbf{z})\|^2 / \sigma_o \right) \|\mathbf{r}_o(\mathbf{z})\|^2 \left(\mathbf{r}_o(\mathbf{z}) \otimes P_r \tilde{\Sigma}_o \mathbf{z} \right) \right] (IFS), \\
 \mathbf{T}_2 &= 2E_F \left[\chi'' \left(\|\mathbf{r}_o(\mathbf{z})\|^2 / \sigma_o \right) \left(\mathbf{r}_o(\mathbf{z}) \otimes P_r \tilde{\Sigma}_o \mathbf{z} \right) O \left(\tilde{\Sigma}_o, \mathbf{D}_o, \mathbf{a}_o, \mathbf{z} \right) \right] (IFD), \\
 \mathbf{T}_3 &= E_F \left[\chi' \left(\|\mathbf{r}_o(\mathbf{z})\|^2 / \sigma_o \right) \left[\left(\mathbf{z}^t \otimes \mathbf{D}_o \right) (IFD) \right] \otimes \left(\mathbf{P}_r \tilde{\Sigma}_o \mathbf{z} \right) \right], \\
 \mathbf{T}_4 &= E_F \left[\chi' \left(\|\mathbf{r}_o(\mathbf{z})\|^2 / \sigma_o \right) \left(\mathbf{r}_o(\mathbf{z}) \otimes \left(\mathbf{z}^t \otimes \mathbf{I}_m \right) \right) (IFD) \right], \\
 \mathbf{T}_5 &= \chi' \left(\|\mathbf{r}_o(\mathbf{z}_0)\|^2 / \sigma_o \right) \left(\mathbf{r}_o(\mathbf{z}_0) \otimes \mathbf{P}_r \tilde{\Sigma}_o \mathbf{z}_0 \right), \\
 \mathbf{T}_{1,0} &= E_F \left[\chi'' \left(\|\mathbf{r}_o(\mathbf{z})\|^2 / \sigma_o \right) \|\mathbf{r}_o(\mathbf{z})\|^2 \mathbf{r}_o(\mathbf{z}) \right], \\
 \mathbf{T}_{2,0} &= 2E_F \left[\chi'' \left(\|\mathbf{r}_o(\mathbf{z})\|^2 / \sigma_o \right) \mathbf{r}_o(\mathbf{z}) O \left(\tilde{\Sigma}_o, \mathbf{D}_o, \mathbf{a}_o, \mathbf{z} \right) \right], \\
 \mathbf{T}_{3,0} &= E_F \left[\chi' \left(\|\mathbf{r}_o(\mathbf{z})\|^2 / \sigma_o \right) \left(\mathbf{z}^t \otimes \mathbf{D}_o \right) \right], \\
 \mathbf{T}_{4,0} &= \chi' \left(\|\mathbf{r}_o(\mathbf{z}_0)\|^2 / \sigma_o \right) \mathbf{r}_o(\mathbf{z}_0).
 \end{aligned}$$

Then, we can establish the following result for the influence function.

Theorem 4 *Let $\mathbf{z}_1, \dots, \mathbf{z}_n$ be a random sample in \mathbb{R}^m , $m = p + q$, from an elliptical distribution F with density (18), location parameter $\boldsymbol{\mu}_0$ and dispersion parameter Σ_0 . Suppose that conditions C0-C14 hold and the functionals $\Sigma_{xx}^{(R)}$ and $\Sigma_{yy}^{(R)}$ are Fisher-consistent. Let $\boldsymbol{\Omega} \in \mathbb{R}^{r(m+1) \times r(m+1)}$ be the invertible matrix defined in (29). Then,*

- (i) *If $\mathbf{T}_v = -\frac{1}{\sigma_o} \mathbf{T}_1 + \frac{2}{\xi_o} (\mathbf{T}_2 + \sigma_o (\mathbf{T}_3 + \mathbf{T}_4)) + \mathbf{T}_5$ and $\mathbf{T}_a = \frac{1}{\sigma_o} \mathbf{T}_{1,0} IFS - \frac{1}{\xi_o} (\mathbf{T}_{2,0} + \sigma_o \mathbf{T}_{3,0}) (IFD) - \mathbf{T}_{4,0}$, we get $IF(\mathbf{z}_0, \boldsymbol{\theta}_{SM}, F) = \boldsymbol{\Omega}^{-1} \mathbf{T}(\mathbf{z}_0, F)$, with $\mathbf{T}(\mathbf{z}_0, F) = (\mathbf{T}'_v, \mathbf{T}'_a)^t$.*
- (ii) *The IF for $\rho_{SM,k}^2$ is $IF(\mathbf{z}_0, \rho_{SM,k}^2, F) = 2(\gamma_k^0 - 1)(t_k^0)^t IF(\mathbf{z}_0, \mathbf{M}^0, F) t_k^0$, $k = 1, \dots, r$.*
- (iii) *Let $(x_0, y_0) = ((v_k^0)^t \mathbf{x}_0, (\mathbf{w}_k^0)^t \mathbf{y}_0)$, $((v_k^0)^t \mathbf{x}, (\mathbf{w}_k^0)^t \mathbf{y}) \sim \tilde{F}_k$ for $(\mathbf{x}^t, \mathbf{y}^t)^t \sim F$. Take the functionals \mathbf{v}_k and \mathbf{w}_k corresponding to the k -canonical vector such that there exist their IF, $IFV_k = IF((\mathbf{x}_0, \mathbf{y}_0), \mathbf{v}_k, F)$ and $IFW_k = IF((\mathbf{x}_0, \mathbf{y}_0), \mathbf{w}_k, F)$, respectively. Let $g_{i+j-1}(\mathcal{L}(\mathbf{x}, \mathbf{y}), \mathbf{v}, \mathbf{w}) = \left(\sigma_{ij}^{(R)}\right)^2 (\mathcal{L}(\mathbf{v}^t \mathbf{x}, \mathbf{w}^t \mathbf{y}))$, $1 \leq i \leq j \leq 2$ such that there exist $g_{l,k,v} = \frac{\partial g_l}{\partial v}(\tilde{F}_k, \mathbf{v}_k^0, \mathbf{w}_k^0)$ and $g_{l,k,w} = \frac{\partial g_l}{\partial w}(\tilde{F}_k, \mathbf{v}_k^0, \mathbf{w}_k^0)$. Then, the IF of $\rho_{C,k}^2$ is given by*

$$\begin{aligned}
 &IF((\mathbf{x}_0, \mathbf{y}_0), \rho_{C,k}^2, F) \\
 &= \frac{1}{g_1^2(\tilde{F}_k, \mathbf{v}_k^0, \mathbf{w}_k^0) g_3^2(\tilde{F}_k, \mathbf{v}_k^0, \mathbf{w}_k^0)}
 \end{aligned}$$

$$\begin{aligned}
 & \times \frac{1}{2} \left\{ \left[\sum_{\substack{1 \leq i, j, l \leq 3 \\ i \neq j, i \neq l, j \neq l}} (-1)^{i+j} g_i(\tilde{F}_k, \mathbf{v}_k^0, \mathbf{w}_k^0) g_j(\tilde{F}_k, \mathbf{v}_k^0, \mathbf{w}_k^0) IF((x_0, y_0), g_l, \tilde{F}_k) \right] \right. \\
 & + \left[\sum_{\substack{1 \leq i, j, l \leq 3 \\ i \neq j, i \neq l, j \neq l}} (-1)^{i+j} g_i(\tilde{F}_k, \mathbf{v}_k^0, \mathbf{w}_k^0) g_j(\tilde{F}_k, \mathbf{v}_k^0, \mathbf{w}_k^0) g_{l,k,v} \right] IFV_k \\
 & \left. + \left[\sum_{\substack{1 \leq i, j, l \leq 3 \\ i \neq j, i \neq l, j \neq l}} (-1)^{i+j} g_i(\tilde{F}_k, \mathbf{v}_k^0, \mathbf{w}_k^0) g_j(\tilde{F}_k, \mathbf{v}_k^0, \mathbf{w}_k^0) g_{l,k,w} \right] IFW_k \right\}.
 \end{aligned}$$

Proof It is deferred to the “Appendix”. □

4.3 Discussion

The term \mathbf{T}_5 through the vector $\mathbf{P}_r \tilde{\Sigma}_o \mathbf{z}_0$ is the only one affected by the contaminated point $\mathbf{z}_0 = (x_0, y_0)$, yielding unbounded IF which coincides with the usual behavior for S -estimation. This term displays that the worst scenario for the IF yielding unboundedness is given by outliers living in the subspace orthogonal to the subspace generated by the eigenvectors of \mathbf{M} under the elliptical model. It is worth noting that even $\Sigma_{xx}^{(R)}$ and $\Sigma_{yy}^{(R)}$ with bounded IF do not guarantee bounded IF for the CCA functionals. We next display some pictures to show the behavior of the IF under two point mass contaminations. Let us consider a vector $\mathbf{z} = (\mathbf{x}^t, \mathbf{y}^t)^t \sim N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = (0.5, 1, 1.5, 2)^t$, $\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Sigma}_{yy} = \mathbf{I}_2$, $\boldsymbol{\Sigma}_{xy} = \text{diag}(0.9, 0.5)$. To ease the computation, the covariance matrices $\boldsymbol{\Sigma}_{xx}$ and $\boldsymbol{\Sigma}_{yy}$ are assumed to be known and therefore they are not estimated from the data. The eigenvectors associated with the largest eigenvalues are $\mathbf{t}_1^0 = (1/\sqrt{2}, 0, -1/\sqrt{2}, 0)^t$ and $\mathbf{t}_2^0 = (0, -1/\sqrt{2}, 0, 1/\sqrt{2})^t$, and the point mass is taken either $\mathbf{z}_0 = (x_{01}, x_{02}, 0, 0)$ or $\mathbf{z}_0 = (x_{01}, 0, y_{01}, 0)$. The plots display unstandardized influence functions to depict the behavior under point masses contamination, (x_{01}, x_{02}) vs $\|IF(\mathbf{z}_0, \mathbf{w}_{SM,1}, F)\|$ and (x_{01}, x_{02}) vs $\|IF(\mathbf{z}_0, \mathbf{w}_{C,1}, F)\|$, with $\mathbf{w}_{C,1}$ and $\mathbf{w}_{SM,1}$ the classical estimator and the SM-estimator, respectively, for the first canonical vector associated with the random vector \mathbf{y} .

5 Robust proposals for measuring association

The classical scenario for CCA involves the correlation as a measure for the maximal association between the linear combinations of two random vectors, and the correlation between those linear combinations is called the canonical correlation. (30) is a robust correlation functional to measure association obtained from a bivariate robust dispersion estimator. A special case arises from the S-estimators defined by Davies

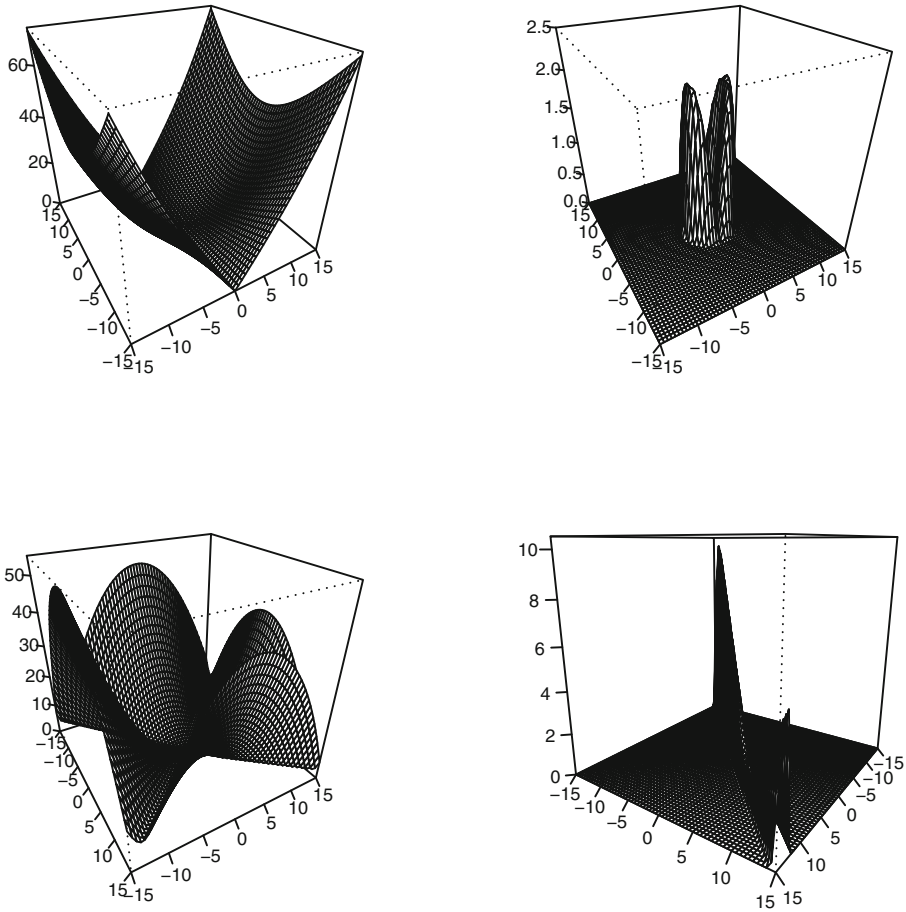


Fig. 1 The plot (x_{01}, x_{02}) vs $\|IF(\mathbf{z}_0, \mathbf{w}_{C,1}, F)\|$ (above on the left) and the plot (x_{01}, x_{02}) vs $\|IF(\mathbf{z}_0, \mathbf{w}_{SM,1}, F)\|$ (above on the right) correspond to the IF for the classical and SM estimators for the first canonical vector associated with the vector \mathbf{y} and $\mathbf{z}_0 = (x_{01}, x_{02}, 0, 0)$. The plot (x_{01}, y_{01}) vs $\|IF(\mathbf{z}_0, \mathbf{w}_{C,1}, F)\|$ (below on the left) and the plot (x_{01}, y_{01}) vs $\|IF(\mathbf{z}_0, \mathbf{w}_{SM,1}, F)\|$ (below on the right) correspond to the IF for the classical and SM estimators for the first canonical vector associated with the vector \mathbf{y} and $\mathbf{z}_0 = (x_{01}, 0, x_{02}, 0)$

(1987) for multivariate location and scatter $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathbb{R}^p \times \mathcal{O}_p$. Given the set

$$\mathcal{D} = \left\{ (\boldsymbol{\mu}, \boldsymbol{\Sigma}) \in \mathbb{R}^p \times \mathcal{O}_p : E\chi \left[(\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \leq \delta \right\},$$

S-estimators are defined as the solutions $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \arg \min_{\mathcal{D}} \det(\boldsymbol{\Sigma})$. This formulation is equivalent to take the M-scale $s(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \min \{s > 0 : E\chi[(\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})/s] \leq \delta, \det(\boldsymbol{\Sigma}) = 1\}$, $(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) = \arg \min s(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = [s(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)]^{1/p} \boldsymbol{\Sigma}^*$. If $p = 2$, let $\hat{\boldsymbol{\mu}}, \hat{\sigma}_{11}$ and $\hat{\sigma}_{22}$ be preliminary estimates for the location and dispersion

parameters μ , σ_{11} and σ_{22} . Set $\widehat{\Sigma} = \begin{pmatrix} \hat{\sigma}_{11} & 0 \\ 0 & \hat{\sigma}_{22} \end{pmatrix}$ and take the standardized vector $\tilde{x} = \widehat{\Sigma}^{-1/2}(\mathbf{x} - \hat{\mu})$. Then, we might estimate the correlation parameter as

$$b^* = \arg \min s(b), \tag{31}$$

with $s(b)$ defined as

$$s(b) = \min \left\{ s > 0 : E\chi \left(\frac{\tilde{x}^t \Sigma^{-1} \tilde{x}}{s} \right) \leq \delta, \Sigma = \frac{1}{(1-b^2)^{1/2}} \begin{pmatrix} 1 & b \\ b & 1 \end{pmatrix} \right\}.$$

We next motivate this robust measure of association b^* by giving another point of view closely related to that of closeness between two random variables, which is behind the proposals (5) and (15). We also discuss another robust procedure for robust correlation. We later analyze the usual properties that a measure of association is usually required for both procedures.

5.1 Motivating the robust correlation b^* and association properties

Let us suppose that we have two random variables X and Y with distribution functions F_X and F_Y , respectively. Let $T(\cdot)$ and $S(\cdot)$ be two equivariant estimators for location and dispersion which allow us to take the standardized random variables $U = (X - T(F_X))/S(F_X)$ and $V = (Y - T(F_Y))/S(F_Y)$. In case of having finite second moments, and taking T as the expected value and S as the standard deviation, when considering either the objective function $E(U - \lambda V)^2$ or $E(V - \lambda U)^2$ one obtains $\hat{\lambda} = E(UV) = \rho$, the Pearson correlation between U and V , as a global minimizer. A measure of association ν must be symmetric, that is, it should verify that $\nu(U, V) = \nu(V, U)$, which entails to consider $E[(U - \lambda V)^2 + (V - \lambda U)^2]$ as objective function. The argument in the expected value is a quadratic form with eigenvalues $(1 - \lambda)^2$ and $(1 + \lambda)^2$ associated with the eigenvectors $2^{-1/2}(1, 1)^t$ and $2^{-1/2}(1, -1)^t$, respectively, that is,

$$(U - \lambda V)^2 + (V - \lambda U)^2 = (1 - \lambda)^2 \left(\frac{U + V}{\sqrt{2}} \right)^2 + (1 + \lambda)^2 \left(\frac{U - V}{\sqrt{2}} \right)^2 \tag{32}$$

The variables $Z = (U + V)/\sqrt{2}$ and $W = (U - V)/\sqrt{2}$ are uncorrelated, $Var(Z) = 1 + \rho$, $Var(W) = 1 - \rho$, with $\rho = E(UV)$. The minimization of the objective function $E[(1 - \lambda)^2 Z^2 + (1 + \lambda)^2 W^2]$ suggests to consider a more general expression, by taking any pair of uncorrelated zero mean random variables Z and W such that their joint distribution is parametrized by $\eta \in C \subset \mathbb{R}$, that is, $E_\eta Z = 0 = E_\eta W$, $E_\eta(ZW) = 0$, $Var_\eta(Z) = z(\eta)$, $Var_\eta(W) = w(\eta)$. Given functions $a : C \rightarrow \mathbb{R}$ and $b : C \rightarrow \mathbb{R}$, we proceed similarly to (10) and (11), coming up with a new pair of variables αZ and βW “as close as possible”, that is, we look for

$\hat{\lambda}$ such that

$$E_{\eta}(a(\hat{\lambda})Z - b(\hat{\lambda})W)^2 \leq E_{\eta}(a(\lambda)Z - b(\lambda)W)^2 \text{ for all } \lambda, \eta \in C \subset \mathbb{R} \quad (33)$$

subject to

$$0 < \text{Var}_{\eta}(a(\eta)Z) = \text{Var}_{\eta}(b(\eta)W). \quad (34)$$

We want to derive which smooth functions a and b should be chosen to make the underlying parameter η the value in which the minimum is obtained. The objective function subject to the constraint (34) to be minimized is $E_{\eta}(a(\lambda)Z - b(\lambda)W)^2 = a^2(\lambda)z(\eta) + b^2(\lambda)w(\eta)$. If $h(\lambda) = a^2(\lambda)$ and $g(\lambda) = b^2(\lambda)$ are differentiable, the critical points solve the equation $h'(\lambda)z(\eta) + g'(\lambda)w(\eta) = 0$. In order to make the true parameter η a critical point of the function $E_{\eta}(a(\lambda)Z - b(\lambda)W)^2$, it must hold that $h'(\eta)z(\eta) = -g'(\eta)w(\eta)$, which let us say that $g(\eta) = C'(z(\eta)/w(\eta))^{1/2}$ and $h(\eta) = C'(w(\eta)/z(\eta))^{1/2}$ with $C' > 0$. Observe that the function $(a(\lambda)Z - b(\lambda)W)^2$ has one term which is independent of λ which entails to consider the quadratic form

$$\begin{aligned} q(\lambda) &= h(\lambda)Z^2 + g(\lambda)W^2 \\ &= \left(\frac{w(\lambda)}{z(\lambda)}\right)^{1/2} \left(\frac{U+V}{\sqrt{2}}\right)^2 + \left(\frac{z(\lambda)}{w(\lambda)}\right)^{1/2} \left(\frac{U-V}{\sqrt{2}}\right)^2. \end{aligned} \quad (35)$$

Let χ be verifying the conditions C1, C2, C3 and C5. If $0 < \delta < 1$ and $\phi(\lambda) = (w(\lambda)/z(\lambda))^{1/2} = \theta$, we can define an M-scale to evaluate the largeness of $q(\lambda)$ through the equation

$$s(\theta) = \min \left\{ s > 0 : E\chi \left(\frac{\theta \left(\frac{U+V}{\sqrt{2}}\right)^2 + \theta^{-1} \left(\frac{U-V}{\sqrt{2}}\right)^2}{s} \right) \leq \delta \right\}, \quad (36)$$

and the estimator is defined to be $\hat{\theta} = \arg \min_{\theta \in \phi(C)} s(\theta)$. If the function ϕ is invertible, the estimator for λ is given by $\hat{\lambda} = \phi^{-1}(\hat{\theta})$. Then, going back to Z and W , with $z(\eta) = 1 + \eta$ and $w(\eta) = 1 - \eta$, we have that the correlation parameter η is estimated through the quadratic form $q(\lambda) = 2^{-1} [\lambda_1 (U + V)^2 + \lambda_2 (U - V)^2]$ with eigenvalues $\lambda_1 = ((1 - \lambda)/(1 + \lambda))^{1/2}$ and $\lambda_2 = ((1 + \lambda)/(1 - \lambda))^{1/2}$. Then, the M-scale in (36) can be defined as

$$s(\lambda) = \min \left\{ s > 0 : E\chi \left(\frac{q(\lambda)}{s} \right) \leq \delta \right\}, \lambda \in (-1, 1). \quad (37)$$

Put $s(1) = \liminf_{\lambda \uparrow 1^-} s(\lambda)$ and $s(-1) = \liminf_{\lambda \downarrow -1^+} s(\lambda)$. Then, we define the SM-estimator for correlation as

$$\hat{\lambda} = \arg \min_{\lambda \in [-1, 1]} s(\lambda). \quad (38)$$

If we use an M-scale based on the quadratic form given in (32), similarly to (37) and (38), we obtain that the solution is $\hat{\lambda} = \frac{1-\sqrt{1-\rho^2}}{\rho} 1_{(0,1)}(\rho) + \frac{1+\sqrt{1-\rho^2}}{\rho} 1_{(-1,0)}(\rho)$ when (X, Y) is elliptically distributed with correlation ρ , and ρ is obtained as $\rho = 2\hat{\lambda}/(1 + \hat{\lambda}^2)$.

It is easily seen that (38) and (31) yield the same estimation.

Lemma 3 *Let $\hat{\lambda}$ be as in (38). Then, $\hat{\lambda} = b^*$.*

Proof It is deferred to the ‘‘Appendix’’. \square

Since the eigenvalues of the quadratic form (32) do not follow (35), ρ is obtained after transforming $\hat{\lambda}$. Next Lemma will show that the estimator (38) has the usual properties required for an association estimator, except for the cases $\hat{\lambda} = -1$ or $\hat{\lambda} = 1$ which entail that the random variables X and Y are linearly related with a probability greater than or equal to $1 - \delta$.

Lemma 4 *Let $\hat{\lambda}$ be as in (38). Then, (i) the estimator is location and scale invariant. (ii) $\hat{\lambda} \in [-1, 1]$. (iii) If $\hat{\lambda} = 1$, then $P(X = aY + b) \geq 1 - \delta$, with $a = S(F_X)/S(F_Y)$ and $b = -(S(F_X)T(F_Y))/S(F_Y) + T(F_X)$. (respectively, if $\hat{\lambda} = -1$, then $P(X = cY + d) \geq 1 - \delta$, with $c = -S(F_X)/S(F_Y)$ and $d = S(F_X)T(F_Y)/S(F_Y) + T(F_X)$. (iv) If (X, Y) is elliptically distributed with correlation ρ , then $\hat{\lambda} = \rho$.*

Proof It is deferred to the ‘‘Appendix’’. \square

5.2 A ‘‘depth-based’’ correlation measure and association properties

The quadratic form (32) can be used to derive another robust procedure to detect ρ consistently which is reminiscent of depth-based procedures (see Adrover et al. 2002). We can pin a residual against adversaries as follows. Let us take a measure of how badly λ performs compared with another fit using θ with lower residuals, that is, $p(\lambda) = \max_{\theta} P((U - \theta V)^2 + (V - \theta U)^2 < (U - \lambda V)^2 + (V - \lambda U)^2)$. Finally, the parameter with the best worst performance is chosen as, $\hat{\lambda} = \arg \min_{\lambda} p(\lambda)$. It is easy to see that the maximum in $p(\lambda)$ is obtained when $\theta = \lambda$ and

$$p(\lambda) = \max \left(P \left((2UV)/(U^2 + V^2) < \lambda \right), P \left((2UV)/(U^2 + V^2) > \lambda \right) \right).$$

Then, the minimum occurs if $\hat{\lambda} = \text{med}((2UV)/(U^2 + V^2))$.

The following lemma shows the properties that $\hat{\lambda}$ possesses as an association measure.

Lemma 5 *Let $\hat{\lambda} = \text{med} \left(\frac{2UV}{U^2 + V^2} \right)$. Then, (i) The estimator is location and scale invariant. (ii) $\hat{\lambda} \in [-1, 1]$. (iii) If $\hat{\lambda} = 1$, then $P(X = aY + b) \geq 0.5$, with $a = S(F_X)/S(F_Y)$ and $b = -aT(F_Y) + T(F_X)$ (respectively, if $\hat{\lambda} = -1$, then $P(X = cY + d) \geq 0.5$, with $c = -S(F_X)/S(F_Y)$ and $d = (-cT(F_Y)) + T(F_X)$. (iv) If (X, Y) is elliptically distributed with correlation ρ , then $\hat{\lambda} = \rho$.*

Proof It is deferred to the ‘‘Appendix’’. \square

6 Concluding remarks

Maronna (2005) shows the remarkable prediction behavior of SM-estimation for PCA compared to some other robust procedures. Adrover and Donato (2015) also conclude in a similar way in the CCA context, by considering either mean squared error or relative prediction error as performance measures under contamination scenarios. By these means, SM-estimation deserved to be thoroughly analyzed, studying its asymptotic properties as consistency and asymptotic normality. Adrover and Donato (2015) specially highlighted the relationship between PCA and CCA (see also ten Berge 1979). Therefore, the derivation of these asymptotic properties of the estimators for both procedures as well as the IF is totally similar. We also considered robust properties such as qualitative robustness and influence function and the unbounded influence, which is a usual behavior of the S-estimation for other models, is also shown. In the end, we turned our attention to a basic problem, the robust association or correlation. We reasoned similarly to that of SM-estimation in the CCA, and we come up with a robust correlation measure, which is totally related to S-estimation for bivariate dispersion. We also discuss another consistent robust correlation by using an approach for regression depth from Adrover et al. (2002). The usual properties for correlation are discussed.

Acknowledgements We would like to thank two anonymous referees and the Associate Editor for their comments and suggestions that have resulted in a much improved paper.

Appendix

Proof of Theorem 1 Let \mathcal{G} be as in (22), $\Xi_o = \Sigma_{xx}$, $\Gamma_o = \Sigma_{yy}$ and the sets

$$E_\eta = \left\{ (\mathbf{A}, \mathbf{B}, \mathbf{a}, \Xi, \Gamma) \in \mathcal{O}_{r,p} \times \mathcal{O}_{r,q} \times \mathbb{R}^r \times S_p \times S_q : \right. \\ \left. \|\mathbf{A} - \mathbf{A}_o\| < \eta, \|\mathbf{B} - \mathbf{B}_o\| < \eta, \|\Xi - \Xi_o\| < \eta, \|\Gamma - \Gamma_o\| < \eta \right\}$$

$$\tilde{E}_\eta = \left\{ (\mathbf{A}, \mathbf{B}, \mathbf{a}, \Xi, \Gamma) \in \mathcal{O}_{r,p} \times \mathcal{O}_{r,q} \times \mathbb{R}^r \times S_p \times S_q : \right. \\ \left. \|\Xi - \Xi_o\| < \eta, \|\Gamma - \Gamma_o\| < \eta \right\}.$$

The function $g_P(\mathbf{A}, \mathbf{B}, \mathbf{a}, \Xi, \Gamma, \sigma) = E_P \left(\frac{\mathbf{A}\Xi^{-1/2}\mathbf{x} - \mathbf{B}\Gamma^{-1/2}\mathbf{y}}{\sigma} \right)$ is continuous in \mathcal{G} for $P = F$. Given $\epsilon > 0$, by using Theorem 1 in Adrover and Donato (2015), we get that there exist $\eta > 0$ and $0 < \tilde{\eta} < \delta$

$$\inf_{\tilde{E}_\eta, \sigma < \sigma_o - \epsilon} g_F(\mathbf{A}, \mathbf{B}, \mathbf{a}, \Xi, \Gamma, \sigma) > \delta + \tilde{\eta}/2$$

$$\sup_{E_\eta, \sigma > \sigma_o + \epsilon} g_F(\mathbf{A}, \mathbf{B}, \mathbf{a}, \Xi, \Gamma, \sigma) < \delta - \tilde{\eta}/2.$$

Then, if $(\mathbf{A}, \mathbf{B}, \mathbf{a}, \Xi, \Gamma) \in E_\eta$, there exists $n_0(\epsilon)$ such that the M-scale $\hat{\sigma} \in [\sigma_o - \epsilon, \sigma_o + \epsilon]$ for all $n > n_0(\epsilon)$. If $(\mathbf{A}, \mathbf{B}, \mathbf{a}, \Xi, \Gamma) \notin E_\eta$, there exists $\eta > 0$ and $M > 0$ such that for all $n > n_1(\epsilon)$ and $F_n \in \mathcal{F}_n$, it holds that

$$\inf_{\mathbf{a} \in [-M, M]^c, E_\eta^c, \sigma \in [\sigma_o - \epsilon, \sigma_o + \epsilon]} g_{F_n}(\mathbf{A}, \mathbf{B}, \mathbf{a}, \Xi, \Gamma, \sigma) > \delta + \tilde{\eta}/2.$$

Consequently, we can conclude that the SM- estimators in (15) belong to a closed bounded set. Moreover, $\lim_{n \rightarrow \infty} \hat{\sigma} = \sigma_o$. Therefore, the Fisher consistency given in Theorem 1 in Adrover and Donato (2015) let us conclude that any convergent subsequent should converge to $(\mathbf{A}_o, \mathbf{B}_o, \mathbf{a}_o)$ and the consistency follows.

The following technical lemma is needed to prove the asymptotic normality.

Lemma A.1 *Let z_1, \dots, z_n be a random sample in \mathbb{R}^m from an elliptical distribution F with density (18), location parameter μ_o and dispersion parameter Σ_o . Suppose that conditions C0-C14 hold. Let \mathcal{G} be as in (22). Let ϕ_1 and ϕ_2 be defined as in (26). Then, there exists a function $\tilde{\theta} : \mathcal{H} \rightarrow \mathcal{G}$ and a bounded set $\mathcal{C} \subset \mathcal{G}$ such that $\tilde{\theta}_o$ is an interior point of $\tilde{\theta}(\mathcal{C})$ and the sets $\mathcal{F}_{1i} = \{\phi_{1i}(\mathbf{z}, \tilde{\theta}(\xi)) : \xi \in \mathcal{C}\}$, $i = 1, \dots, mr$ and $\mathcal{F}_{2k} = \{\phi_{2k}(\mathbf{z}, \tilde{\theta}(\xi)) : \xi \in \mathcal{C}\}$, $k = 1, \dots, r$ are Euclidean classes with envelopes F_{1i} , $i = 1, \dots, m$ and F_{2k} , $k = 1, \dots, r$, such that $E_F(F_{1i})^2 < \infty$ and $E_F(F_{2k})^2 < \infty$.*

Proof of Theorem 2(i) Given $\tilde{\theta}_o$, Lemma A.1 says that $\phi_{1i}(\mathbf{z}, \tilde{\theta}_o) \in \mathcal{F}_{1i}$, $i = 1, \dots, mr$ and $\phi_{2k}(\mathbf{z}, \tilde{\theta}_o) \in \mathcal{F}_{2k}$, $k = 1, \dots, r$. Moreover, since Theorem 1 ensures the consistency of $\hat{\theta} = (\hat{\mathbf{A}}_{SM}^o, \hat{\mathbf{B}}_{SM}^o, \hat{\mathbf{a}}_{SM}, \hat{\Sigma}_{xx}^{(R)}, \hat{\Sigma}_{yy}^{(R)}, \hat{\sigma})$ to $\tilde{\theta}_o$, given $\varepsilon_0 > 0$ we can find n_0 such that for any $n \geq n_0$, it holds that $P(\phi_{1i}(\mathbf{z}, \hat{\theta}) \in \mathcal{F}_{1i}, \phi_{2k}(\mathbf{z}, \hat{\theta}) \in \mathcal{F}_{2k}) > 1 - \varepsilon_0$ for all $i \in \{1, \dots, mr\}$ and $k \in \{1, \dots, r\}$. Given \mathcal{F} a Euclidean class and $\delta > 0$, set $[\delta] = \{(f_1, f_2) \in \mathcal{F} \times \mathcal{F} : \int (f_1 - f_2)^2 dP < \delta^2\}$. Given a sequence of independent identically distributed random variables ξ_1, \dots, ξ_n such that $\xi_1 \sim P$, set

$$v_n(f) = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^n \left(f(\xi_i) - \int f dP \right) \right], f \in \mathcal{F}.$$

Given $\varepsilon > 0$ and $\eta > 0$, Lemma 2.16 of Pakes and Pollard (1989), C12 and C13 say that there exist $\delta > 0$ and $n_1 \in \mathbb{N}$ such that, for all $n \geq n_1$, $(\phi_{1j}(\mathbf{z}, \hat{\theta}), \phi_{1j}(\mathbf{z}, \tilde{\theta}_o)) \in [\delta]$ and $\limsup_{n \rightarrow \infty} P \left\{ \sup_{[\delta]} |v_n(\phi_{1j}(\cdot, \hat{\theta})) - v_n(\phi_{1j}(\cdot, \tilde{\theta}_o))| > \eta \right\} < \varepsilon$. Then, we can conclude that

$$\begin{aligned} v_n(\phi_j(\cdot, \hat{\theta})) - v_n(\phi_j(\cdot, \tilde{\theta}_o)) &= o_P(1), j = 1, 2, \\ v_n(\phi(\cdot, \hat{\theta})) - v_n(\phi(\cdot, \tilde{\theta}_o)) &= o_P(1). \end{aligned}$$

Since $\Phi(\tilde{\theta}_o) = \mathbf{0}$, by summing up and subtracting some terms we have

$$\mathbf{0} = \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{z}_i, \hat{\theta}) - \Phi(\tilde{\theta}_o) = \Phi(\hat{\theta}) - \Phi(\tilde{\theta}_o) + \frac{1}{\sqrt{n}} v_n(\phi(\cdot, \hat{\theta}))$$

$$= \Phi(\hat{\theta}) - \Phi(\tilde{\theta}_o) + \frac{1}{\sqrt{n}} v_n(\phi(\cdot, \tilde{\theta}_o)) + \frac{1}{\sqrt{n}} \left[v_n(\phi(\cdot, \hat{\theta})) - v_n(\phi(\cdot, \tilde{\theta}_o)) \right].$$

By C14 and (27), it holds that $-\frac{1}{\sqrt{n}} v_n(\phi(\cdot, \tilde{\theta}_o)) = [\Omega + o_P(1)](\hat{\theta}_{SM} - \theta_o) + o_P(1/\sqrt{n})$, and by the Central Limit Theorem we have that $v_n(\phi(\cdot, \tilde{\theta}_o)) \xrightarrow{\mathcal{D}} N_{r(m+1)}(\mathbf{0}, \mathbf{V})$. Since Ω is invertible, we obtain that $\sqrt{n}(\hat{\theta}_{SM} - \theta_o) \xrightarrow{\mathcal{D}} N_{r(m+1)}(\mathbf{0}, \Omega^{-1} \mathbf{V}_o (\Omega^{-1})^t)$. Straightforward computation let us conclude the explicit form of the asymptotic dispersion matrix and the proof follows. (ii) follows closely from (i). (iii) follows from the fact that $\hat{\mathbf{v}}_{SM,k}$ and $\hat{\mathbf{w}}_{SM,k}$ are subvectors of $\hat{\boldsymbol{\theta}}_{SM}^*$ given in (ii). (iv) gives a simpler form of the asymptotic covariance matrix obtained in (iii) in case of having a non-singular matrix of derivatives $\frac{\partial \Phi_2(\boldsymbol{\theta})}{\partial \mathbf{a}}(\tilde{\boldsymbol{\theta}}_o)$. \square

Proof of Theorem 4 Given $F_\varepsilon = (1 - \varepsilon) F + \varepsilon \delta_{z_0}$, we have to look for the SM-functionals defined as $g(\mathbf{D}, \tilde{\boldsymbol{\Sigma}}_\varepsilon, \mathbf{a}, \sigma_\varepsilon) = E_{F_\varepsilon} \chi \left(\frac{\|\mathbf{D} \tilde{\boldsymbol{\Sigma}}_\varepsilon \mathbf{z} - \mathbf{a}\|^2}{\sigma_\varepsilon(\mathbf{D}, \mathbf{a})} \right) = \delta$. Then, we look for a restricted minimum $\mathbf{D} \in \mathcal{O}_{r,m}$, $\mathbf{t}_1, \dots, \mathbf{t}_r$ are the rows of \mathbf{D} and the Lagrangian can be expressed as

$$L(\mathbf{t}_1, \dots, \mathbf{t}_r, \mathbf{a}, \tilde{\boldsymbol{\Sigma}}_\varepsilon, \eta_{11}, \dots, \eta_{rr}) = \sigma_\varepsilon(\mathbf{D}, \mathbf{a}) - \sum_{j=1}^r \eta_{jj} (\mathbf{t}_j^t \mathbf{t}_j - 1) - \sum_{1 \leq j < k \leq r} \eta_{jk} \mathbf{t}_k^t \mathbf{t}_j,$$

where $\mathbf{t}_{1,\varepsilon}, \dots, \mathbf{t}_{r,\varepsilon}, \mathbf{a}_\varepsilon$ are critical points for L . The proof follows closely to that of Theorem 1 in Croux and Ruiz-Gazen (2005). \square

Proof of Lemma 3 The eigenvectors $(1, 1)$ and $(1, -1)$ of $\tilde{\boldsymbol{\Sigma}}$ correspond to the eigenvalues $((1 - b)/(1 + b))^{1/2}$ and $((1 + b)/(1 - b))^{1/2}$, respectively. Then, the quadratic forms in both definitions coincide and $\hat{\lambda} = b^*$. \square

Proof of Lemma 4 (i) and (ii) are easily derived. (iii) $\hat{\lambda} = 1$ implies that $s(1) \leq s(\lambda)$ for all $\lambda \in [-1, 1]$. Let $q(\lambda)$ be as in (35). Since $0 = \lim_{s \rightarrow \infty} E \chi(q(\lambda)/s) \leq \delta$, this implies that $s(\lambda) < \infty$ and $s(1) < \infty$. Thus, $\delta \geq \lim_{\lambda \rightarrow 1^-} E \chi(q(\lambda)/s(\lambda))$ and $P(U = V) \geq 1 - \delta$ which says that $P(X = aY + b) \geq 1 - \delta$, with $a = S(F_X)/S(F_Y)$ and $b = -aT(F_Y) + T(F_X)$. In case that $\hat{\lambda} = -1$, we get $P(U = -V) \geq 1 - \delta$, and $P(X = cY + d) \geq 1 - \delta$, with $c = -S(F_X)/S(F_Y)$ and $d = S(F_X)T(F_Y)/S(F_Y) + T(F_X)$. Thus, (iii) is proved. (iv) In case of having (X, Y) elliptically distributed with correlation ρ , Lemma 3 let us affirm that $\hat{\lambda} = \rho$. \square

Proof of Lemma 5 (i) is easily derived and (iii) follows as in Lemma 4 (iii). (ii) Since $(U \pm V)^2 \geq 0$, then $-(U^2 + V^2) \leq UV \leq (U^2 + V^2)$ and $|2UV/(U^2 + V^2)| \leq 1$. Therefore, $\text{med}(2UV/(U^2 + V^2)) \in [-1, 1]$. (iv) In case of having (X, Y) elliptically distributed with correlation ρ , (U, V) is elliptically distributed with density $f(u, v) =$

$1/K f_0((u^2 + 2\rho uv + v^2)/\sqrt{1 - \rho^2}))$ and $K = \pi(1 - \rho^2)^{-1/2}(F_0(\infty) - F_0(0))$ with F_0 a primitive of f_0 . To see that $P_c = P(2UV/(U^2 + V^2) \leq \rho) = 0.5$, we perform some change of variables to get that

$$P_c = \frac{2^{-1}(1 - \rho^2)^{-1/2}}{K} \int I_{(-\infty, \rho)} \left(\frac{z^2 - w^2 + \rho(z^2 + w^2)}{z^2 + w^2 + \rho(z^2 - w^2)} \right) f_0(z^2 + w^2) dz dw.$$

Using spherical coordinates, $1 + \rho \cos 2\theta \geq 0$ and the fact that $\frac{\cos 2\theta + \rho}{1 + \rho \cos 2\theta} \leq \rho$ if and only if $\cos 2\theta \leq 0$, then we have

$$\begin{aligned} P_c &= \frac{2^{-1}(1 - \rho^2)^{-1/2}}{K} \int_0^{2\pi} \int_0^\infty I_{(-\infty, \rho)} \left(\frac{\cos 2\theta + \rho}{1 + \rho \cos 2\theta} \right) r f_0(r^2) dr d\theta \\ &= \frac{2^{-1}(1 - \rho^2)^{-1/2}}{K} \int_0^{2\pi} \int_0^\infty I_{[\pi/4, 3\pi/4] \cup [5\pi/4, 7\pi/4]}(\theta) r f_0(r^2) dr d\theta \\ &= \frac{2^{-1}\pi(1 - \rho^2)^{-1/2}}{K} \int_0^\infty r f_0(r^2) dr d\theta = \frac{\pi(1 - \rho^2)^{-1/2}}{2K} (F_0(\infty) - F_0(0)) = \frac{1}{2}, \end{aligned}$$

which shows that $\hat{\lambda} = \rho$ and the result follows. \square

References

- Adrover J, Donato SM (2015) A robust predictive approach for canonical correlation analysis. *J Multivar Anal* 133:356–376
- Adrover J, Maronna R, Yohai V (2002) Relationships between maximum depth and projection estimates. *J Stat Plan Inference* 105:363–375
- Alfons A, Croux C, Filzmoser P (2017) Robust maximum association estimators. *J Am Stat Assoc* 112(517):436–445
- Anderson TW (1999) Asymptotic theory for canonical correlation analysis. *J Multivar Anal* 70(1):1–29
- Boente G (1987) Asymptotic theory for robust principal components. *J Multivar Anal* 21(1):67–78
- Branco JA, Croux C, Filzmoser P, Oliveira MR (2005) Robust canonical correlations: a comparative study. *Comput Stat* 20(2):203–229
- Croux C, Haesbroeck G (2000) Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika* 87:603–618
- Croux C, Ruiz-Gazen A (2005) High breakdown estimators for principal components: the projection-pursuit approach revisited. *J Multivar Anal* 95:206–226
- Croux C, García-Escudero LA, Gordaliza A, Ruwet C, San Martín R (2017) Robust PCA based on trimming. *Stat Sin* 27:1437–1459
- Cui H, He X, Ng KW (2003) Asymptotic distributions of principal components based on robust dispersions. *Biometrika* 90:953–966
- Das S, Sen PK (1998) Canonical correlations. In: Armitage P, Colton T (eds) *Encyclopedia of biostatistics*. Wiley, New York, pp 468–482
- Davies PL (1987) Asymptotic behavior of S-estimators of multivariate location estimators and dispersion matrices. *Ann Stat* 15:1269–1292
- Drašković G, Breloy A, Pascal F (2019) On the asymptotics of Maronna's robust PCA. *IEEE Trans Signal Process* 67(19):4964–4975
- Furrer R, Genton MG (2011) Aggregation-cokriging for highly multivariate spatial data. *Biometrika* 98:615–631
- Hampel FR (1971) A general qualitative definition of robustness. *Ann Math Stat* 42(6):1887–1896
- Hotelling H (1936) Relations between two sets of variables. *Biometrika* 28:321–377
- Kudraszow N, Maronna RA (2010) Estimates of MM type for the multivariate linear model. Technical report. [arXiv:1004.4883](https://arxiv.org/abs/1004.4883)

- Kudraszow N, Maronna RA (2011) Estimates of MM type for the multivariate linear model. *J Multivar Anal* 102:1280–1292
- Li G, Chen Z (1985) Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and Monte Carlo. *J Am Stat Assoc* 80:759–766
- Maronna R (1976) Robust M-estimators of multivariate location and scatter. *Ann Stat* 4(1):51–67
- Maronna R (2005) Principal components and orthogonal regression based on robust scales. *Technometrics* 47(3):264–273
- Maronna R, Martin D, Yohai V, Salibián-Barrera (2019) Robust statistics, theory and methods, 2nd edn. Wiley Series in Probability and Statistics. Wiley, New York
- Pakes A, Pollard D (1989) Simulation and the asymptotics of optimization estimators. *Econometrica* 57(5):1027–1057
- Rao CR (1962) Relations between weak and uniform convergence of measures with applications. *Ann Math Stat* 33:659–680
- Seber GAF (2004) Multivariate observations, 2nd edn. Wiley, New York
- Taskinen S, Croux C, Kankainen A, Ollila E, Oja H (2006) Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *J Multivar Anal* 97:359–384
- ten Berge JMF (1979) On the equivalence of two oblique congruence rotation methods and orthogonal approximations. *Psychometrika* 44:359–364
- Tyler D (1981) Asymptotic inference for eigenvectors. *Ann Stat* 9(4):725–736
- Wilms I, Croux C (2015) Sparse canonical correlation analysis from a predictive point of view. *Biom J* 57(5):834–851
- Wilms I, Croux C (2016) Robust sparse canonical correlation analysis. *BMC Syst Biol* 10:72. <https://doi.org/10.1186/s12918-016-0317-9>
- Yohai V (1987) High breakdown-point and high efficiency robust estimates for regression. *Ann Stat* 15(2):642–658
- Yohai V, García Ben M (1981) Canonical variables as optimal predictors. *Ann Stat* 8(4):865–869
- Yohai V, Zamar R (1988) High breakdown-point estimates of regression by means of the minimization of an efficient scale. *J Am Stat Assoc* 83(402):406–413

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.