Canonical Analysis based on Scatter Matrices

Sara Taskinen[*], Christophe Croux[+], Annaliisa Kankainen[*]
Esa Ollila[#], and Hannu Oja[*]

[*] *Dept. of Mathematics and Statistics, University of Jyväskylä*

[+] *Dept. of Applied Economics, K.U. Leuven*

[#] *Signal Processing Laboratory, Helsinki University of Technology*

Abstract

In this paper, the influence functions and limiting distributions of the canonical correlations and coefficients based on affine equivariant scatter matrices are developed for elliptically symmetric distributions. General formulas for limiting variances and covariances of the canonical correlations and canonical vectors based on scatter matrices are obtained. Also the use of the so called shape matrices in canonical analysis is investigated. The scatter and shape matrices based on the affine equivariant Sign Covariance Matrix as well as the Tyler's shape matrix serve as examples. Their finite sample and limiting efficiencies are compared to those of the Minimum Covariance Determinant estimator and S-estimates through theoretical and simulation studies. The theory is illustrated by an example.

Keywords: Canonical correlations, canonical variables, canonical vectors, shape matrix, sign covariance matrix, Tyler's estimate

[*]Corresponding author: Hannu Oja, Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35, Fin-40351 Jyväskylä, Finland; email: Hannu.Oja@maths.jyu.fi

1 Introduction

The purpose of canonical correlation analysis (CCA) is to describe the linear interrelations between two random multivariate vectors. New coordinate systems are found for both vectors in such a way that, in both systems, the marginals of the random variables are uncorrelated and have unit variances, and that the covariance matrix between the two random vectors is a diagonal matrix with descending positive diagonal elements. The new variables and their correlations are called canonical variates and canonical correlations, respectively. Moreover, the rows of the transformation matrix are called canonical vectors. Canonical analysis is one of the fundamental contributions to multivariate inference by Harold Hotelling (1936).

To be more specific, assume that $\boldsymbol{x}$ and $\boldsymbol{y}$ are p- and q-variate random vectors, $p \leq q$ and $k = p + q$. Let F be the cumulative distribution function of the k-variate variable $\boldsymbol{z} = (\boldsymbol{x}^T, \boldsymbol{y}^T)^T$. Decompose its covariance matrix (if it exists) as

$$\Sigma = \Sigma(F) = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}$$

where Σ_{xx} and Σ_{yy} are nonsingular. In canonical analysis, one thus finds a $p \times p$ matrix $A = A(F)$, a $q \times q$ matrix $B = B(F)$ and $p \times p$ diagonal matrix $R = R(F) = diag(\rho_1, \ldots, \rho_p)$, $\rho_1 \geq \ldots \geq \rho_p$, such that

$$\begin{pmatrix} A^T & 0 \\ 0 & B^T \end{pmatrix} \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \begin{pmatrix} I_p & (R, 0) \\ (R, 0)^T & I_q \end{pmatrix}. \quad (1)$$

The diagonal elements of R are called the **canonical correlations**, the columns of A and B the **canonical vectors** and the random vectors

$$\boldsymbol{x}' = A^T \boldsymbol{x} \quad \text{and} \quad \boldsymbol{y}' = B^T \boldsymbol{y}$$

give the **canonical variates**.

Simple calculations show that

$$\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} A = A(R, 0)(R, 0)^T$$

and

$$\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} B = B(R, 0)^T (R, 0).$$

Therefore A and (the first p columns of) B contain the eigenvectors of the matrices

$$M_A = \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \quad \text{and} \quad M_B = \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy}, \quad (2)$$

respectively. The eigenvalues of M_A and M_B are the same and are given by the diagonal elements of R^2, so by the squared canonical correlations. We will assume throughout the paper that $\rho_1 > \ldots > \rho_q$ to avoid multiplicity problems. From (1) we see that the eigenvectors need to be chosen such that

$$A^T \Sigma_{xx} A = I_p \text{ and } B^T \Sigma_{yy} B = I_q. \tag{3}$$

Alternatively, one can also find eigenvalues and orthonormal eigenvectors A_0 and B_0 of symmetric matrices as

$$\Sigma_{xx}^{-1/2} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1/2} A_0 = A_0 (R, 0)(R, 0)^T$$

and

$$\Sigma_{yy}^{-1/2} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1/2} B_0 = B_0 (R, 0)^T (R, 0),$$

with $A_0^T A_0 = I_p$ and $B_0^T B_0 = I_q$. The regular canonical vectors are then $A = \Sigma_{xx}^{-1/2} A_0$ and $B = \Sigma_{yy}^{-1/2} B_0$. For more information on the canonical analysis problem, see e.g. Johnson and Wichern (1998, chapter 10).

To estimate the population canonical correlations and vectors one typically estimates Σ by the sample covariance matrix, and computes afterwards the eigenvalues and eigenvectors of the sample counterparts of the matrices M_A and M_B given in (2). This procedure is optimal for a multivariate normal distribution F, but it turns out to be less efficient at heavier tailed model distributions. Moreover, the sample covariance matrix is highly sensible to outliers, and a canonical analysis based on this matrix will then give unreliable results. For these reasons, it can be appropriate to estimate Σ by other, more robust estimators. As such, Karnel (1991) proposed to use M-estimators and Croux and Dehon (2002) the Minimum Covariance Determinant estimator. However, no asymptotic theory has been developed yet for canonical analysis based on robust covariance matrix estimators.

Is was only quite recently that Anderson (1999) completed the asymptotic theory for canonical correlation analysis based on the sample covariance matrix. In this paper we study the asymptotic distribution of estimates of canonical correlations and canonical vectors based on more general estimators of the population covariance matrix, the so called scatter matrices. The results will not be restricted to the normal case, but are valid for the class of elliptically symmetric model distributions. Moreover, also the asymptotic distribution for canonical analysis based on shape matrices has been derived.

The plan of the paper is as follows. Section 2 reviews scatter matrices, and the general form of their influence function and limiting variance. We also treat shape matrices, which are estimating the form of the underlying elliptical distribution, but have no size information. The main contribution of the paper is in Section 3, where expressions for the influence function, the limiting distribution and the limiting efficiencies are derived for canonical correlations and vectors based on any regular scatter and shape matrix estimator. In Section 4, numerical values for the asymptotic efficiencies at normal distributions are presented for several scatter matrices: the Sign Covariance Matrix (Visuri et al, 2000), the Minimum Covariance Determinant estimator (Rousseeuw 1985), and S-estimators (Davies 1987). We also consider Tyler's shape matrix (Tyler 1987) estimator. By means of a simulation study, the finite sample efficiencies are compared with the limiting ones. Finally, a real data example will illustrate the methods. The Appendix collects all the proofs.

2 Scatter and shape matrices

2.1 Some definitions

A $k \times k$ matrix valued statistical functional $C = C(F)$ is a **scatter matrix** if it is positive definite and symmetric ($PDS(k)$) and affine equivariant. We can denote $C(F)$ alternatively as $C(\boldsymbol{z})$ if $\boldsymbol{z} \sim F$. Affine equivariance then means that $C(D^T\boldsymbol{z} + \boldsymbol{b}) = D^T C(\boldsymbol{z})D$ for all $k \times k$ matrices D and k-vectors $\boldsymbol{b}$. This implies that, for a spherically symmetric distribution F_0, $C(F_0) = c_0 I_k$ with some constant $c_0 > 0$. If F is the cdf of the elliptic random vector $\boldsymbol{z} = D^T\boldsymbol{z}_0 + \boldsymbol{b}$, where $\boldsymbol{z}_0 \sim F_0$ and D is a positive definite $k \times k$ matrix, then $C(F) = c_0 D^T D$. As c_0 depends on the functional C and the distribution F_0, a correction factor is needed for having Fisher consistency towards the regular covariance matrix $\Sigma(F)$. Introducing such a correction factor also allows comparisons between different scatter matrix estimates at a specific model.

A functional $V = V(F)$, or alternatively $V(\boldsymbol{z})$, is a **shape matrix** if it is $PDS(k)$ with $Det(V) = 1$ and affine equivariant in the sense that

$$V(D^T\boldsymbol{z} + \boldsymbol{b}) \;=\; \{Det[D^T V(\boldsymbol{z})D]\}^{-1/k}\; D^T V(\boldsymbol{z})D$$

The condition $Det(V) = 1$ is sometimes replaced by the condition $Tr(V) = k$ but the former one is more convenient here (Ollila et al., 2003a). If $C(F)$ is a scatter matrix then

$$V(F) \;=\; \{Det[C(F)]\}^{-1/k}\; C(F)$$

is the associated shape matrix. It can be seen as a standardized version of $C(F)$. A shape matrix can, however, be given without any reference to a scatter matrix; the Tyler's shape matrix serves as an example and will be discussed in detail later. For the above elliptical distribution F of $\boldsymbol{z} = D^T\boldsymbol{z}_0 + \boldsymbol{b}$, $V(F) = [Det(D^TD)]^{-1/k}D^TD$. This means that in the elliptic model, shape matrices estimate the same population quantity and are directly comparable without any modifications. Note that in several multivariate inference problems, the test and estimation procedures are based on the shape matrix only.

Finally note that if $C(F)$ is a scatter matrix, the functional $S(F) = Det(C(F))$ is a global scalar valued **scale measure**. The scale measure $Det(\Sigma(F))$ given by the regular covariance matrix is the well-known Wilks' generalized variance. In general, we will say that $S(F)$ is a scale measure if it is nonnegative and affine equivariant in the sense that $S(G\boldsymbol{z}) = Det(G)^2S(\boldsymbol{z})$ for all non singular $k \times k$ matrices G. Finally note, that if $V(F)$ is a shape matrix and $S(F)$ is a scale measure, then

$$C(F) = [S(F)]^{1/k}V(F)$$

yields a scatter matrix. Thus the shape and scale information may be combined to build a scatter matrix.

2.2 Influence functions

Influence functions are often used for robustness considerations. The influence function measures the robustness of a functional T against a single outlier, that is, the effect of an infinitesimal contamination located at a single point $\boldsymbol{z}$ on the estimator (see Hampel et al., 1986). Consider herefore the contaminated distribution

$$F_\epsilon = (1-\epsilon)F + \epsilon\Delta_{\boldsymbol{z}},$$

where $\Delta_{\boldsymbol{z}}$ is the cdf of a distribution with probability mass one at a singular point $\boldsymbol{z}$. Then the influence function of T is defined as

$$IF(\boldsymbol{z}; T, F) = \lim_{\epsilon \to 0} \frac{T(F_\epsilon) - T(F)}{\epsilon}.$$

Lemma 1 in Croux and Haesbroeck (2000) states that, for any scatter functional $C(F)$, there exist two real valued functions $\gamma_C(r)$ and $\delta_C(r)$ such that the influence function of C at a spherical F_0, symmetric around the origin and with $C(F_0) = I_k$, is given by

$$IF(\boldsymbol{z}; C, F_0) = \gamma_C(r)\boldsymbol{u}\boldsymbol{u}^T - \delta_C(r)I_k \tag{4}$$

where $r = ||\boldsymbol{z}||$ and $\boldsymbol{u} = ||\boldsymbol{z}||^{-1}\boldsymbol{z}$. Then one easily finds for the associated shape functional

$$IF(\boldsymbol{z}; V, F_0) = IF(\boldsymbol{z}; [Det(C)]^{-1/k}C, F_0) = \gamma_C(r)\left[\boldsymbol{u}\boldsymbol{u}^T - \frac{1}{k}I_k\right].$$

and for the associated size functional

$$IF(\boldsymbol{z}; Det(C), F_0) = IF(\boldsymbol{z}; Tr(C), F_0) = \gamma_C(r) - k\delta_C(r). \tag{5}$$

Vice versa, for any shape matrix V and any scale measure S as defined above, we have that the influence functions at spherical F_0 should be of the form

$$IF(\boldsymbol{z}; V, F_0) = \gamma_V(r)\left[\boldsymbol{u}\boldsymbol{u}^T - \frac{1}{k}I_k\right]$$

and

$$IF(\boldsymbol{z}; S, F_0) = \delta_S(r),$$

respectively. Then the resulting combination scatter matrix $C = [Det(S)]^{1/k}V$ has influence function

$$IF(\boldsymbol{z}; C, F_0) = \gamma_V(r)\boldsymbol{u}\boldsymbol{u}^T - \frac{1}{k}[\gamma_V(r) - \delta_S(r)]I_k. \tag{6}$$

Due to equivaraince properties, we readily find that at the elliptic distribution F of $\boldsymbol{z} = D^T\boldsymbol{z}_0 + \boldsymbol{b}$, with $C(F) = D^TD$,

$$IF(\boldsymbol{z}; C, F) = D^T[\gamma_C(r)\boldsymbol{u}\boldsymbol{u}^T - \delta_C(r)I_k]D,$$

$$IF(\boldsymbol{z}; Det(C), F) = Det(D)[\gamma_C(r) - k\delta_C(r)],$$

and finally for the associated shape matrix:

$$IF(\boldsymbol{z}; V, F) = \frac{\gamma_C(r)}{Det(D)^2}D^T\left[\boldsymbol{u}\boldsymbol{u}^T - \frac{1}{k}I_k\right]D,$$

where

$$r^2 = (\boldsymbol{z} - \boldsymbol{b})^T(D^TD)^{-1}(\boldsymbol{z} - \boldsymbol{b}) \quad \text{and} \quad \boldsymbol{u} = \frac{1}{r}(D^TD)^{-1/2}(\boldsymbol{z} - \boldsymbol{b})$$

are the squared Mahalanobis distance and the Mahalanobis angle of $\boldsymbol{z}$.

2.3 Limiting variances

Assume next that $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ is a random sample from a spherical distribution with cdf F_0 and covariance matrix I_k. Let $\widehat{C}_n$ be the estimator associated to the functional C, that is $\widehat{C}_n = C(F_n)$ where F_n is the empirical distribution function computed from the sample. Assume that a correction factor is used to adjust the estimate so that $C(F_0) = I_k$. It is then often true that the limiting distribution of $\sqrt{n}(\widehat{C}_n - I_k)$ is multivariate normal with zero mean matrix and variance $E_{F_0}[IF(\boldsymbol{z}; C, F_0) \otimes IF(\boldsymbol{z}; C, F_0)^T]$. The limiting variances of the diagonal and the off-diagonal elements of $\sqrt{n}(\widehat{C}_n - I_k)$ are then given as

$$ASV(C_{11}; F_0) = \frac{2(k-1)}{k^2(k+2)} E[\gamma_C^2(r)] + \frac{1}{k^2} E[(\gamma_C(r) - k\delta_C(r))^2]$$

and

$$ASV(C_{12}; F_0) = \frac{E[\gamma_C^2(r)]}{k(k+2)},$$

respectively, and the limiting covariances between the diagonal elements are

$$ASC(C_{11}, C_{22}; F_0) = ASV(C_{11}; F_0) - 2ASV(C_{12}; F_0).$$

All other limiting covariances vanish. More formally, the limiting distribution of $\sqrt{n}\text{vec}(\widehat{C}_n - I_k)$ is k^2-variate normal with zero mean vector and covariance matrix

$$ASV(C_{12}; F_0)(I_{k^2} + I_{k,k}) + ASC(C_{11}, C_{22}; F_0)\text{vec}(I_k)\text{vec}(I_k)^T,$$

where vec vectorizes a matrix and $I_{k,k}$ is a $k^2 \times k^2$ matrix with (i,j)-block being equal to a $k \times k$ matrix that is 1 at entry (j,i) and zero elsewhere.

Similarly, the limiting distribution of $\sqrt{n}\text{vec}(\widehat{V}_n - I_k)$ is k^2-variate normal with zero mean vector and covariance matrix

$$ASV(V_{12}; F_0)\left(I_{k^2} + I_{k,k} - \frac{2}{k}\text{vec}(I_k)\text{vec}(I_k)^T\right),$$

with

$$ASV(V_{12}; F_0) = \frac{E[\gamma_V^2(r)]}{k(k+2)}.$$

The limit distribution of the shape matrix estimator is thus characterized by one single number, where the limiting distribution of a scatter matrix estimator is completely determined by 2 numbers. The latter follows already from Tyler (1982).

3 Canonical correlations and canonical vectors

3.1 Definitions

Assume that the k-variate distribution of $\boldsymbol{z} = (\boldsymbol{x}^T, \boldsymbol{y}^T)^T$ is elliptic with cumulative distribution function F and that $p \le q$. Consider the scatter matrix

$$C = C(F) = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix}.$$

with nonsingular C_{xx} and C_{yy}. The matrices $A = A(F)$, $B = B(F)$ and $R = R(F)$ chosen such that

$$C\left(\begin{pmatrix} A^T\boldsymbol{x} \\ B^T\boldsymbol{y} \end{pmatrix}\right) = C\left(\begin{pmatrix} A^T\boldsymbol{x} \\ (B_1, B_2)^T\boldsymbol{y} \end{pmatrix}\right) = \begin{pmatrix} I_p & (R,0) \\ (R,0)^T & I_q \end{pmatrix}$$

then yield the canonical vectors and correlations. The canonical correlations in R are the same for all scatter matrices C. Note that, if the p canonical correlations are distinct, then the $p \times p$ matrix A and $q \times p$ matrix B_1 are unique up to a sign and the $q \times (q-p)$ matrix B_2 is unique up to multiplication on the right by an orthogonal $(q-p) \times (q-p)$ matrix. The values of the canonical vectors A and B will depend on the used scatter functional C via the constant c_0. If, however, the scatter functional is such that $C(F) = \Sigma$, then the canonical vectors become comparable over different scatter matrix estimators used.

Now let $A(F)$, $B(F)$ and $R(F)$ be determined by a shape matrix functional $V = V(F)$ such that

$$V\left(\begin{pmatrix} A^T\boldsymbol{x} \\ B^T\boldsymbol{y} \end{pmatrix}\right) = Det\left(\begin{pmatrix} I_p & (R,0) \\ (R,0)^T & I_q \end{pmatrix}\right)^{-1/k} \begin{pmatrix} I_p & (R,0) \\ (R,0)^T & I_q \end{pmatrix}.$$

One has now that the same canonical correlations R are obtained again, but the canonical vectors are only unique up to a constant. We therefore make the choice to take A^* and B^* such that $A^{*T}V_{xx}A^* = I_p$ and $B^{*T}V_{yy}B^* = I_q$. If the shape functional V is associated to a scatter functional C as described in Section 2.1, then

$$A^* = [Det(C)]^{1/2k}A \quad \text{and} \quad B^* = [Det(C)]^{1/2k}B.$$

We call A^* and B^* the **standardized canonical vectors.** These standardized canonical vectors are comparable between any two scatter or shape matrix functional used, whether a correction factor has been used or not.

3.2 Influence functions

Let $\boldsymbol{z} = (\boldsymbol{x}^T, \boldsymbol{y}^T)^T$ follow the $k = (p+q)$ dimensional model distribution F, and denote F' the cdf of the canonical variates

$$\boldsymbol{z}' = \begin{pmatrix} A^T\boldsymbol{x} \\ B^T\boldsymbol{y} \end{pmatrix}.$$

As the canonical correlation and vector estimates are affine equivariant, for computation of the influence function it is enough to consider the distribution F' where

$$R(F') = diag(\rho_1, \ldots, \rho_p), \quad A(F') = I_p \text{ and } B(F') = \begin{pmatrix} I_p & 0 \\ 0 & B_{22} \end{pmatrix},$$

and B_{22} is an orthogonal $(q-p) \times (q-p)$ matrix. Then $C_{xx}(F') = I_p$, $C_{yy}(F') = I_q$ and $C_{xy}(F') = C_{yx}^T(F') = (R, 0)$. The influence functions of A, B and R at F' are obtained as follows.

From the conditions $A^T C_{xx} A = I_p$ and $B^T C_{yy} B = I_q$ we directly have

$$IF(\boldsymbol{z}'; A^T, F') + IF(\boldsymbol{z}'; C_{xx}, F') + IF(\boldsymbol{z}'; A, F') = 0 \tag{7}$$

and

$$\begin{aligned} &IF(\boldsymbol{z}'; B^T, F') \begin{pmatrix} I_p & 0 \\ 0 & B_{22} \end{pmatrix} + \begin{pmatrix} I_p & 0 \\ 0 & B_{22}^T \end{pmatrix} IF(\boldsymbol{z}'; C_{yy}, F') \begin{pmatrix} I_p & 0 \\ 0 & B_{22} \end{pmatrix} \\ &+ \begin{pmatrix} I_p & 0 \\ 0 & B_{22}^T \end{pmatrix} IF(\boldsymbol{z}'; B, F') = 0. \end{aligned} \tag{8}$$

Further, the conditions $A^T C_{xy} B = (R, 0)$ and $B^T C_{yx} A = (R, 0)^T$ yield

$$\begin{aligned} &IF(\boldsymbol{z}'; A^T, F')(R, 0) + IF(\boldsymbol{z}'; C_{xy}, F') \begin{pmatrix} I_p & 0 \\ 0 & B_{22} \end{pmatrix} \\ &+ (R, 0) IF(\boldsymbol{z}'; B, F') = IF(\boldsymbol{z}'; (R, 0), F') \end{aligned} \tag{9}$$

and

$$\begin{aligned} &IF(\boldsymbol{z}'; B^T, F')(R, 0)^T + \begin{pmatrix} I_p & 0 \\ 0 & B_{22}^T \end{pmatrix} IF(\boldsymbol{z}'; C_{yx}, F') \\ &+ (R, 0)^T IF(\boldsymbol{z}'; A, F') = IF(\boldsymbol{z}'; (R, 0)^T, F'). \end{aligned} \tag{10}$$

The diagonal elements of (7) and (8), for $i = 1, \ldots, p$, then give

$$IF(\boldsymbol{z}'; A_{ii}, F') = -\frac{1}{2} IF(\boldsymbol{z}'; [C_{xx}]_{ii}, F')$$

and

$$IF(\boldsymbol{z}'; B_{ii}, F') = -\frac{1}{2} IF(\boldsymbol{z}'; [C_{yy}]_{ii}, F').$$

From the diagonal elements of (9) and (10) one gets

$$IF(\boldsymbol{z}'; R_{ii}, F') = \rho_i IF(\boldsymbol{z}'; A_{ii}, F') + IF(\boldsymbol{z}'; [C_{xy}]_{ii}, F') + \rho_i IF(\boldsymbol{z}'; B_{ii}, F'),$$

$i = 1, \ldots, p$. Combining the 4 equations (7), (8), (9), and (10), one obtains for the off-diagonal elements of A ($i, j = 1, \ldots, p$, $i \neq j$) that

$$\begin{aligned}(\rho_j^2 - \rho_i^2) IF(\boldsymbol{z}'; A_{ij}, F') = & -IF(\boldsymbol{z}'; [C_{xx}]_{ij}, F')\rho_j^2 + IF(\boldsymbol{z}'; [C_{xy}]_{ij}, F')\rho_j \\ & + \rho_i IF(\boldsymbol{z}'; [C_{yx}]_{ij}, F') - \rho_i IF(\boldsymbol{z}'; [C_{yy}]_{ij}, F')\rho_j.\end{aligned}$$

For off-diagonal elements of B, $i = 1, \ldots, q$, $j = 1, \ldots, p$, $i \neq j$ one has

$$\begin{aligned}(\rho_j^2 - \rho_i^2) IF(\boldsymbol{z}'; B_{ij}, F') = & -IF(\boldsymbol{z}'; [C_{yy}]_{ij}, F')\rho_j^2 + IF(\boldsymbol{z}'; [C_{yx}]_{ij}, F')\rho_j \\ & + \rho_i IF(\boldsymbol{z}'; [C_{xy}]_{ij}, F') - \rho_i IF(\boldsymbol{z}'; [C_{xx}]_{ij}, F')\rho_j,\end{aligned}$$

where $\rho_i = 0$ as $q \geq i > p$.

Write now the canonical variates $\boldsymbol{z}'$ as

$$\boldsymbol{z}' = r \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix},$$

where r stands for the Mahalanobis distance of the canonical variates $\boldsymbol{z}'$, which equals the Mahalanobis distance of the untransformed variable $\boldsymbol{z}$. The influence functions at the elliptical F are now obtained using the equivariance and invariance properties of $A(F)$, $B(F)$ and $R(F)$ and are as follows (all proofs are in the Appendix):

Theorem 1. *Let C be the affine equivariant scatter matrix functional used to obtain the canonical correlations R and the canonical vectors A and B_1. Then the influence functions of the functionals R, A, and B_1 at the k-variate elliptic distribution F are*

$$IF(\boldsymbol{z}; R, F) = \gamma_C(r) H_1(\boldsymbol{u}, \boldsymbol{v}; R)$$

and

$$IF(\boldsymbol{z}; A, F) = A(F)\left[\gamma_C(r)H_2(\boldsymbol{u}, \boldsymbol{v}; R) + \frac{1}{2}\delta_C(r)I_p\right]$$

and

$$IF(\boldsymbol{z}; B_1, F) = B(F)\left[\gamma_C(r)H_3(\boldsymbol{u}, \boldsymbol{v}; R) + \frac{1}{2}\delta_C(r)\begin{pmatrix} I_p \\ 0 \end{pmatrix}\right].$$

Here H_1 is a diagonal matrix with diagonal elements

$$[H_1(\boldsymbol{u}, \boldsymbol{v}; R)]_{jj} = u_j v_j - \frac{1}{2}\rho_j u_j^2 - \frac{1}{2}\rho_j v_j^2, \quad j = 1, \ldots, p.$$

The elements of H_3 are

$$[H_3(\boldsymbol{u}, \boldsymbol{v}; R)]_{ij} = \frac{\rho_j(u_j - \rho_j v_j)v_i + \rho_i(v_j - \rho_j u_j)u_i}{\rho_j^2 - \rho_i^2},$$

for $i = 1, \ldots, q$, $j = 1, \ldots, p$, $i \neq j$ and $\rho_i = 0$ as $i > p$, and

$$[H_3(\boldsymbol{u}, \boldsymbol{v}; R)]_{jj} = -\frac{1}{2}v_j^2, \quad j = 1, \ldots, p.$$

Finally, the elements of H_2 are

$$[H_2(\boldsymbol{u}, \boldsymbol{v}; R)]_{ij} = [H_3(\boldsymbol{v}, \boldsymbol{u}; R)]_{ij}, \quad i, j = 1, \ldots, p.$$

The influence functions of the canonical correlations R, and the standardized canonical vectors A^* and B_1^* based on a shape matrix functional V are obtained using the fact that

$$A^* = [Det(C)]^{1/2k}A \quad \text{and} \quad B_1^* = [Det(C)]^{1/2k}B_1,$$

where C is a related scatter matrix constructed as $C(F) = S(F)^{1/k}V(F)$ for a given scale measure S, as described in Section 2.1.

Theorem 2. *Let V be the affine equivariant shape matrix functional used to obtain the canonical correlations R and the standardized canonical vectors A^* and B_1^*. Then the influence functions of the functionals R, A^*, and B_1^* at the k-variate elliptic distribution F are*

$$IF(\boldsymbol{z}; R, F) = \gamma_V(r)H_1(\boldsymbol{u}, \boldsymbol{v}; R)$$

and

$$IF(\boldsymbol{z}; A^*, F) = A^*(F)\gamma_V(r)\left[H_2(\boldsymbol{u}, \boldsymbol{v}; R) + \frac{1}{2k}I_p\right]$$

and

$$IF(\boldsymbol{z}; B_1^*, F) = B^*(F)\gamma_V(r)\left[H_3(\boldsymbol{u},\boldsymbol{v};R) + \frac{1}{2k}\begin{pmatrix} I_p \\ 0 \end{pmatrix}\right],$$

with H_1, H_2 and H_3 as in Theorem 1, and with r and $(\boldsymbol{u}^T, \boldsymbol{v}^T)^T$ as the Mahalanobis distance and Mahalanobis angle of the canonical variates $\boldsymbol{z}' = (A^T\boldsymbol{x}, B^T\boldsymbol{y})^T$, respectively.

Note that the above influence functions factorize in a product of a function of r and a function of $(\boldsymbol{u},\boldsymbol{v})$, where we know that the distribution of r and $(\boldsymbol{u},\boldsymbol{v})$ are statistically independent. Since $H_1(\boldsymbol{u},\boldsymbol{v},R)$, $H_2(\boldsymbol{u},\boldsymbol{v},R)$ and $H_3(\boldsymbol{u},\boldsymbol{v},R)$ are continuous functions on the periphery of an ellipsoid, it follows that the influence functions for the canonical correlations and standardized canonical vectors are bounded as soon as the associated γ_V is bounded.

3.3 Limiting distributions

Write now $\widehat{R}$ and $\widehat{A}$ and $\widehat{B}_1$ for the canonical correlation and vector estimates based on $\widehat{C}$. Assume that $\rho_1 > \ldots > \rho_p > 0$ and that the limiting distribution of $\sqrt{n}\,vec(\widehat{C} - C)$ is multivariate normal with zero mean vector and covariance matrix

$$E[vec\{IF(\boldsymbol{z};C,F)\}vec\{IF(\boldsymbol{z};C,F)\}^T].$$

Then the limiting distributions of $\widehat{R}$, $\widehat{A}$ and $\widehat{B}_1$ are as follows.

Theorem 3. *At an elliptical distribution F, we have that the limiting distribution of $\sqrt{n}\,vec(\widehat{R} - R)$ is multivariate normal with zero mean matrix and covariance matrix*

$$\begin{aligned} ASV(\widehat{R};F) &= E[vec\{IF(\boldsymbol{z};R,F)\}vec\{IF(\boldsymbol{z};R,F)\}^T] \\ &= E[\gamma_C^2(r)]E[vec\{H_1(\boldsymbol{u},\boldsymbol{v};R)\}vec\{H_1(\boldsymbol{u},\boldsymbol{v};R)\}^T]. \end{aligned}$$

Furthermore, the limiting distribution of $\sqrt{n}\,vec(\widehat{A} - A)$ is multivariate normal with zero mean matrix and covariance matrix

$$\begin{aligned} ASV(\widehat{A};F) &= E[vec\{IF(\boldsymbol{z};A,F)\}vec\{IF(\boldsymbol{z};A,F)\}^T] \\ &= (I_p \otimes A)ASV(\widehat{A};F')(I_p \otimes A^T). \end{aligned}$$

and the limiting distribution of $\sqrt{n}\,vec(\widehat{B}_1 - B_1)$ is multivariate normal with zero mean matrix and covariance matrix

$$\begin{aligned} ASV(\widehat{B}_1;F) &= E[vec\{IF(\boldsymbol{z};B_1,F)\}vec\{IF(\boldsymbol{z};B_1,F)\}^T] \\ &= (I_p \otimes B)ASV(\widehat{B}_1;F')(I_p \otimes B^T). \end{aligned}$$

Now let $\widehat{R}$, $\widehat{A}^*$ and $\widehat{B}_1^*$ be the canonical correlation and standardized canonical vector estimates based on a shape matrix $\widehat{V}$. If $\rho_1 > \ldots > \rho_p > 0$ and the limiting distribution of $\sqrt{n}\, vec(\widehat{V} - V)$ is multivariate normal with zero mean matrix and covariance matrix

$$E[vec\{IF(\boldsymbol{z}; V, F)\}vec\{IF(\boldsymbol{z}; V, F)\}^T],$$

then the limiting distributions of $\widehat{R}$, $\widehat{A}^*$ and $\widehat{B}_1^*$ are given by similar expressions as in Theorem 3.

Theorem 4. *At an elliptical distribution F, we have that the limiting distribution of $\sqrt{n}\, vec(\widehat{R} - R)$ is multivariate normal with zero mean matrix and covariance matrix*

$$\begin{aligned} ASV(\widehat{R}; F) &= E[vec\{IF(\boldsymbol{z}; R, F)\}vec\{IF(\boldsymbol{z}; R, F)\}^T] \\ &= E[\gamma_C^2(r)]E[vec\{H_1(\boldsymbol{u}, \boldsymbol{v}; R)\}vec\{H_1(\boldsymbol{u}, \boldsymbol{v}; R)\}^T]. \end{aligned}$$

Furthermore, the limiting distribution of $\sqrt{n}\, vec(\widehat{A}^* - A^*)$ is multivariate normal with zero mean matrix and covariance matrix

$$\begin{aligned} ASV(\widehat{A}^*; F) &= E[vec\{IF(\boldsymbol{z}; A^*, F)\}vec\{IF(\boldsymbol{z}; A^*, F)\}^T] \\ &= c_R^2\,(I_p \otimes A^*)ASV(\widehat{A}^*; F')(I_p \otimes A^{*T}) \end{aligned}$$

and the limiting distribution of $\sqrt{n}\, vec(\widehat{B}_1^* - B_1^*)$ is multivariate normal with zero mean matrix and covariance matrix

$$\begin{aligned} ASV(\widehat{B}_1^*; F) &= E[vec\{IF(\boldsymbol{z}; B_1^*, F)\}vec\{IF(\boldsymbol{z}; B_1^*, F)\}^T] \\ &= c_R^2\,(I_p \otimes B^*)ASV(\widehat{B}_1^*; F')(I_p \otimes B^{*T}), \end{aligned}$$

where $c_R = |I_p - R^2|^{-1/2k}$.

In the next subsection we explicit further the limiting variances of canonical correlation and vector estimates.

3.4 Limiting covariances and efficiencies

Let F be again an elliptical model distribution and consider first the canonical distribution F' of the canonical variates $\boldsymbol{z}'$. As before, the spherical version of F will be denoted by F_0. The limiting covariances of the elements of $\widehat{R}$, $\widehat{A}$ and $\widehat{B}_1$ based on scatter matrix $\widehat{C}$ at F' are listed in the following theorem.

Theorem 5. *Let C_{12} be any off-diagonal and C_{11} any diagonal element of the scatter matrix C. At the canonical distribution F' we have that:*
(i) For $1 \leq i \leq p$, the asymptotic covariance matrix of $[\widehat{r}_i,\ \widehat{a}_{ii},\ \widehat{b}_{ii}]^T$ is

$$\frac{1}{4}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} ASV(\widehat{C}_{11}; F_0) + (1-\rho_i^2)\begin{bmatrix} (1-\rho_i^2) & -\frac{1}{2}\rho_i & -\frac{1}{2}\rho_i \\ -\frac{1}{2}\rho_i & 0 & -\frac{1}{2} \\ -\frac{1}{2}\rho_i & -\frac{1}{2} & 0 \end{bmatrix} ASV(\widehat{C}_{12}; F_0).$$

(ii) For $1 \leq i \neq j \leq p$, the asymptotic covariance matrix between $[\widehat{a}_{ii}, \widehat{b}_{ii}]^T$ and $[\widehat{a}_{jj}, \widehat{b}_{jj}]^T$ is

$$\frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} ASV(\widehat{C}_{11}; F_0) - \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} ASV(\widehat{C}_{12}; F_0).$$

(iii) For $1 \leq i \neq j \leq p$, the asymptotic covariance matrix of $[(\rho_j^2-\rho_i^2)\,\widehat{a}_{ij}, (\rho_i^2-\rho_j^2)\,\widehat{a}_{ji}]^T$ and also of $[(\rho_j^2-\rho_i^2)\,\widehat{b}_{ij}, (\rho_i^2-\rho_j^2)\,\widehat{b}_{ji}]^T$, is given by

$$\begin{bmatrix} (1-\rho_j^2)(\rho_i^2+\rho_j^2-2\rho_i^2\rho_j^2) & (1-\rho_i^2)(1-\rho_j^2)(\rho_i^2+\rho_j^2) \\ (1-\rho_i^2)(1-\rho_j^2)(\rho_i^2+\rho_j^2) & (1-\rho_i^2)(\rho_i^2+\rho_j^2-2\rho_i^2\rho_j^2) \end{bmatrix} ASV(\widehat{C}_{12}; F_0)$$

(iv) For $1 \leq i \neq j \leq p$, the asymptotic covariance matrix between $[(\rho_j^2-\rho_i^2)\,\widehat{a}_{ij}, (\rho_i^2-\rho_j^2)\,\widehat{a}_{ji}]^T$ and $[(\rho_j^2-\rho_i^2)\,\widehat{b}_{ij}, (\rho_i^2-\rho_j^2)\,\widehat{b}_{ji}]^T$ is given by

$$\begin{bmatrix} \rho_i\rho_j(2-\rho_i^2-3\rho_j^2+\rho_i^2\rho_j^2+\rho_j^4) & 2\rho_i\rho_j(1-\rho_i^2)(1-\rho_j^2) \\ 2\rho_i\rho_j(1-\rho_i^2)(1-\rho_j^2) & \rho_i\rho_j(2-\rho_j^2-3\rho_i^2+\rho_i^2\rho_j^2+\rho_i^4) \end{bmatrix} ASV(\widehat{C}_{12}; F_0),$$

(v) For $j = 1, \ldots, p$, and with $q \geq i > p$, the asymptotic variance of $\widehat{b}_{ij}$ is given by

$$(\rho_j^{-2} - 1) ASV(\widehat{C}_{12}; F_0),$$

All the other limiting covariances between elements of $\widehat{R}$, $\widehat{A}$ or $\widehat{B}_1$ are equal to zero.

The special case of the the sample covariance matrix $\widehat{Cov}$ at normal distribution give limiting covariances obtained earlier by Anderson (1999). In this special case $ASV(\widehat{Cov}_{11}; F_0) = 2$ and $ASV(\widehat{Cov}_{12}; F_0) = 1$, and expressions (i), (iii) and (iv) correspond with those of Anderson (1999). Note that the second statement of Theorem 5 gives, for the special case of the normal distribution and the sample covariance matrix, a zero asymptotic covariance matrix between $[\widehat{a}_{ii}, \widehat{b}_{ii}]^T$ and $[\widehat{a}_{jj}, \widehat{b}_{jj}]^T$. Anderson (1999) also assumed $p = q$, and therefore did not reported the last statement of Theorem 5 for $\widehat{Cov}$.

From Theorems 3 and 5 the marginal distributions of the canonical correlation and vector estimates at elliptical F can readily be obtained. Denote $\widehat{\boldsymbol{a}}_1, \ldots, \widehat{\boldsymbol{a}}_p$ the columns of $\widehat{A}$ and $\widehat{\boldsymbol{b}}_1, \ldots, \widehat{\boldsymbol{b}}_p$ the columns of $\widehat{B}_1$.

Corollary 1. *Let F be an elliptical distribution, then $\sqrt{N}(\widehat{r}_j - \rho_j)$, $\sqrt{N}(\widehat{\boldsymbol{a}}_j - \boldsymbol{a}_j)$ and $\sqrt{N}(\widehat{\boldsymbol{b}}_j - \boldsymbol{b}_j)$ have limiting normal distribution with zero mean and asymptotic variances*

$$ASV(\widehat{r}_j; F) = (1-\rho_j^2)^2 ASV(\widehat{C}_{12}; F_0),$$

$$\begin{aligned} ASV(\widehat{\boldsymbol{a}}_j; F) &= \frac{1}{4} ASV(\widehat{C}_{11}; F_0) \boldsymbol{a}_j \boldsymbol{a}_j^T \\ &+ ASV(\widehat{C}_{12}; F_0) \sum_{\substack{k=1 \\ k \neq j}}^{p} \frac{(\rho_k^2 + \rho_j^2 - 2\rho_k^2 \rho_j^2)(1-\rho_j^2)}{(\rho_j^2 - \rho_k^2)^2} \boldsymbol{a}_k \boldsymbol{a}_k^T \end{aligned}$$

and

$$\begin{aligned} ASV(\widehat{\boldsymbol{b}}_j; F) &= \frac{1}{4} ASV(\widehat{C}_{11}; F_0) \boldsymbol{b}_j \boldsymbol{b}_j^T \\ &+ ASV(\widehat{C}_{12}; F_0) \sum_{\substack{k=1 \\ k \neq j}}^{q} \frac{(\rho_k^2 + \rho_j^2 - 2\rho_k^2 \rho_j^2)(1-\rho_j^2)}{(\rho_j^2 - \rho_k^2)^2} \boldsymbol{b}_k \boldsymbol{b}_k^T, \end{aligned}$$

for every $1 \leq j \leq p$. For $q \geq k > p$, we put $\rho_k = 0$.

Note that the multiplication of $B_2 = (\boldsymbol{b}_{p+1}, \ldots, \boldsymbol{b}_q)$ by an orthogonal matrix does not affect the value of the asymptotic variances $ASV(\widehat{\boldsymbol{b}}_j; F)$ of the first p canonical vectors. Moreover, corollary 1 implies that the asymptotic relative efficiency at elliptical F of the estimate $\widehat{r}_{j,C}$ based on a scatter matrix $\widehat{C}$ with respect to $\widehat{r}_{j,C'}$ based on a scatter matrix $\widehat{C}'$ is simply

$$ARE(r_{j,C}, r_{j,C'}; F) = \frac{ASV(\widehat{C}'_{12}; F_0)}{ASV(\widehat{C}_{12}; F_0)},$$

and the asymptotic relative efficiencies of two canonical vector estimates $\widehat{\boldsymbol{a}}_{j,C}$ and $\widehat{\boldsymbol{a}}_{j,C'}$ are determined by the following ratios

$$ARE(a_{jj,C}, a_{jj,C'}; F) = \frac{ASV(\widehat{C}'_{11}; F_0)}{ASV(\widehat{C}_{11}; F_0)}$$

and

$$ARE(a_{ij,C}, a_{ij,C'}; F) = \frac{ASV(\widehat{C}'_{12}; F_0)}{ASV(\widehat{C}_{12}; F_0)}.$$

The above relative efficiencies equal thus relative efficiencies of on- and off-diagonal elements of the scatter matrices at F_0.

At F' all asymptotic covariances of estimates $\widehat{R}$, $\widehat{A}^*$ and $\widehat{B}_1^*$ based on shape matrix $\widehat{V}$ are as follows.

Theorem 6. *Let V_{12} be any off-diagonal and V_{11} any diagonal element of the shape matrix V. Denote $c_R = |I_p - R^2|^{-1/2k}$. At the canonical distribution F', where we have that:*
(i) For $1 \le i \le p$, the asymptotic covariance matrix of $[\widehat{r}_i,\ c_R\,\widehat{a^*}_{ii},\ c_R\,\widehat{b^*}_{ii}]^T$ is

$$\begin{bmatrix} (1-\rho_i^2)^2 & -\frac{1}{2}\rho_i(1-\rho_i^2) & -\frac{1}{2}\rho_i(1-\rho_i^2) \\ -\frac{1}{2}\rho_i(1-\rho_i^2) & \frac{1}{2}-\frac{1}{2k} & -\frac{1}{2}(\frac{1}{k}-\rho_i^2) \\ -\frac{1}{2}\rho_i(1-\rho_i^2) & -\frac{1}{2}(\frac{1}{k}-\rho_i^2) & \frac{1}{2}-\frac{1}{2k} \end{bmatrix} ASV(\widehat{V}_{12};F_0)$$

(ii) For $1 \le i \ne j \le p$, the asymptotic covariance matrix between $c_R\,[\widehat{a^*}_{ii}, \widehat{b^*}_{ii}]^T$ and $c_R\,[\widehat{a^*}_{jj}, \widehat{b^*}_{jj}]^T$ is

$$-\frac{1}{2k}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} ASV(\widehat{V}_{12};F_0).$$

(iii) For $1 \le i \ne j \le p$, the asymptotic covariance matrix of $c_R[(\rho_j^2-\rho_i^2)\,\widehat{a^*}_{ij}, (\rho_i^2-\rho_j^2)\,\widehat{a^*}_{ji}]^T$ and also of $c_R\,[(\rho_j^2-\rho_i^2)\,\widehat{b^*}_{ij}, (\rho_i^2-\rho_j^2)\,\widehat{b^*}_{ji}]^T$, is given by

$$\begin{bmatrix} (1-\rho_j^2)(\rho_i^2+\rho_j^2-2\rho_i^2\rho_j^2) & (1-\rho_i^2)(1-\rho_j^2)(\rho_i^2+\rho_j^2) \\ (1-\rho_i^2)(1-\rho_j^2)(\rho_i^2+\rho_j^2) & (1-\rho_i^2)(\rho_i^2+\rho_j^2-2\rho_i^2\rho_j^2) \end{bmatrix} ASV(\widehat{C}_{12};F_0)$$

(iv) For $1 \le i \ne j \le p$, the asymptotic covariance matrix between $c_R\,[(\rho_j^2-\rho_i^2)\,\widehat{a^*}_{ij}, (\rho_i^2-\rho_j^2)\,c_R\,\widehat{a^*}_{ji}]^T$ and $[(\rho_j^2-\rho_i^2)\,\widehat{b^*}_{ij}, (\rho_i^2-\rho_j^2)\,\widehat{b^*}_{ji}]^T$ is given by

$$\begin{bmatrix} \rho_i\rho_j(2-\rho_i^2-3\rho_j^2+\rho_i^2\rho_j^2+\rho_j^4) & 2\rho_i\rho_j(1-\rho_i^2)(1-\rho_j^2) \\ 2\rho_i\rho_j(1-\rho_i^2)(1-\rho_j^2) & \rho_i\rho_j(2-\rho_j^2-3\rho_i^2+\rho_i^2\rho_j^2+\rho_i^4) \end{bmatrix} ASV(\widehat{V}_{12};F_0),$$

(v) For $j = 1, \ldots, p$, and with $q \ge i > p$, the asymptotic variance of $\widehat{b^*}_{ij}$ is given by

$$(\rho_j^{-2}-1)ASV(\widehat{V}_{12};F_0),$$

All the other limiting covariances between elements of $\widehat{R}$, $\widehat{A^*}$ or $\widehat{B^*}_1$ are equal to zero.

Combining Theorems 4 and 6 it is again immediate to obtain the marginal distributions of the canonical correlations and standardized canonical vectors based on a shape matrix estimator.

Corollary 2. *Let F be an elliptical distribution, then $\sqrt{N}(\widehat{r}_j - \rho_j)$, $\sqrt{N}(\widehat{\boldsymbol{a}}_j^* - \boldsymbol{a}_j^*)$ and $\sqrt{N}(\widehat{\boldsymbol{b}}_j^* - \boldsymbol{b}_j^*)$ have limiting normal distribution with zero mean and asymptotic variances*

$$ASV(\widehat{r}_j; F) = (1-\rho_j^2)^2 ASV(\widehat{V}_{12}; F_0),$$

$$\begin{aligned} ASV(\widehat{\boldsymbol{a}}_j; F) = \Bigg(& \left(\frac{1}{2} - \frac{1}{2k}\right) \boldsymbol{a}_j \boldsymbol{a}_j^T \\ & + \sum_{\substack{k=1 \\ k \neq j}}^{p} \frac{(\rho_k^2 + \rho_j^2 - 2\rho_k^2\rho_j^2)(1-\rho_j^2)}{(\rho_j^2 - \rho_k^2)^2} \boldsymbol{a}_k \boldsymbol{a}_k^T \Bigg) ASV(\widehat{V}_{12}; F_0). \end{aligned}$$

and

$$\begin{aligned} ASV(\widehat{\boldsymbol{b}}_j; F) = \Bigg(& \left(\frac{1}{2} - \frac{1}{2k}\right) \boldsymbol{b}_j \boldsymbol{b}_j^T \\ & + \sum_{\substack{k=1 \\ k \neq j}}^{q} \frac{(\rho_k^2 + \rho_j^2 - 2\rho_k^2\rho_j^2)(1-\rho_j^2)}{(\rho_j^2 - \rho_k^2)^2} \boldsymbol{b}_k \boldsymbol{b}_k^T \Bigg) ASV(\widehat{V}_{12}; F_0), \end{aligned}$$

where $\rho_k = 0$, as $k > p$.

Note that now all the asymptotic efficiencies of canonical correlation and vector estimates based on $\widehat{V}$ relative to estimates based on $\widehat{V}'$ are given by

$$\frac{ASV(\widehat{V}'_{12}; F_0)}{ASV(\widehat{V}_{12}; F_0)}.$$

Table 1 lists these asymptotic relative efficiencies of canonical correlation and vector estimates based on robust shape matrices with respect to the estimates based on classical shape matrix at k-variate normal distribution. Considered robust shape matrices are based on affine equivariant sign covariance matrix (SCM), a 25% breakdown S-estimator with biweight loss-functions, a 25% breakdown Reweighted Minimum Covariance Determinant (RMCD), Tyler's M-estimate and the 25% breakdown MCD-estimator. Asymptotic variances for the SCM were obtained by Ollila et al (2003b), for S-estimators results of Lopuhäa (1989) have been used, for the MCD and RMCD scatter estimators asymptotic variances have been computed by Croux and Haesbroeck (1999), and finally Tyler (1987) showed that the asymptotic variance of Tyler's M-estimate equals $k/(k+2)$. The estimators appearing in Table 1 have been sorted in decreasing order of efficiency. The SCM estimator, being a covariance matrix build from affine equivariant spatial sign vector, has a very high efficiency at the normal model. S-estimators have a slightly lower efficiency, but in contrast to the SCM they have a high breakdown point. The other high breakdown point estimators RMCD and MCD suffer from larger losses in efficiency. Tyler's M estimate has a low breakdown point, but is very fast to compute (see Hettmansperger and Randles, 2002), and has good efficiency in larger dimensions.

Table 1: Asymptotic Relative Efficiencies the canonical correlation and vector estimates based on several robust shape matrices relative to the estimates based on the classical sample covariance matrix at a k-variate normal distribution.

k	SCM	S	RMCD	Tyler	MCD
4	0.982	0.953	0.786	0.667	0.284
6	0.991	0.975	0.837	0.750	0.356
8	0.994	0.984	0.864	0.800	0.403
10	0.996	0.988	0.881	0.833	0.438
20	0.999	0.995	0.917	0.909	0.529

4 Small sample studies

4.1 Finite-sample efficiencies

In this Section we compare by means of a modest simulation study finite-sample efficiencies of canonical correlation and vector estimates based on the robust shape matrices with corresponding estimates based on the classical shape matrix. A number of $M = 1000$ samples of sizes $n = 20, 50, 100, 300$ were generated from three different $2p$-variate normal distributions with fixed covariance matrices

$$\Sigma = \begin{pmatrix} I_p & R \\ R & I_p \end{pmatrix},$$

where $R = diag(\rho_1, \ldots, \rho_p)$. Our choices for canonical correlations were (a) $\rho_1 = 0.8$, $\rho_2 = 0.2$ (b) $\rho_1 = 0.6$, $\rho_2 = 0.4$ and (c) $\rho_1 = 0.9$, $\rho_2 = 0.6$, $\rho_3 = 0.3$. The estimated quantities were the canonical correlations and the standardized canonical vectors. The estimated values were compared with the theoretical ones by the following mean squared errors (MSE). The MSE of the jth canonical correlation is given by

$$\mathrm{MSE}(\widehat{r}_j) = \frac{1}{M} \sum_{m=1}^{M} (\widehat{r}_j^{(m)} - \rho_j)^2,$$

where ρ_j is the true canonical correlation and $\widehat{r}_j^{(m)}$ the corresponding estimate computed from the mth generated sample. Further, the MSE of the jth canonical vector is measured by

$$\mathrm{MSE}(\widehat{\boldsymbol{a}}_j^*) = \frac{1}{M} \sum_{m=1}^{M} \left(\cos^{-1} \left(\frac{|\boldsymbol{a}_j^{*T} \widehat{\boldsymbol{a}}_j^{*(m)}|}{\|\boldsymbol{a}_j^*\| \cdot \|\widehat{\boldsymbol{a}}_j^{*(m)}\|} \right) \right)^2,$$

where $\boldsymbol{a}_j^*$ is the theoretical vector and $\widehat{\boldsymbol{a}}_j^{*(m)}$ the estimate obtained from the mth generated sample. Thus, this MSE is the average squared angle between the estimated and the true standardized canonical vectors. Working with the angle has the advantage that the same MSE are obtained, whether one works with the standardized or unstandardized canonical vectors. The estimated efficiencies were then computed as ratios of the simulated MSE's and are listed in Tables 2-4.

As seen in Table 2, the finite-sample efficiencies convergence to the asymptotic ones listed in the previous Section. For the SCM and the S-estimator the finite-sample efficiencies are very stable over the different sample sizes, but the results for the MCD- and RMCD-estimators appear to be unstable at smaller samples sizes ($n = 20$, $n = 50$). For small samples, Tyler's estimator seems to be more efficient than RMCD. Note that the MCD is more efficient and the RMCD less efficient at small sample sizes than one would expect from the asymptotic results.

In the second case samples were generated from a 4-variate normal distribution, such that the true canonical correlations were closer to each other than in the previous case. Corresponding finite-sample efficiencies are given in Table 3. As compared to the earlier case, now the differences between the finite-sample and asymptotic efficiencies ar more pronounced in particular for small sample sizes. Even in the case $n = 300$, the efficiencies are still quite different from the asymptotical ones for some estimators. The SCM- and S-estimators seem to be the most stable, whereas RMCD- and MCD-estimators behave as unsteadily as in the previous case. This simulation experiments suggest that convergence to the limit distribution

Table 2: Finite-sample efficiencies of the canonical correlation and vector estimates based on five robust shape matrices relative to estimates based on the classical shape matrix. Samples were generated from a 4-variate normal distribution. The quantities to be estimated were $\rho_1 = 0.8$, $\rho_2 = 0.2$, $\boldsymbol{a}_1^{*T} = (1,0)^T$ and $\boldsymbol{a}_2^{*T} = (0,1)^T$.

		SCM	S	RMCD	Tyler	MCD
$\widehat{r}_1$:	$n = 20$	1.008	0.950	0.614	0.747	0.512
	$n = 50$	0.985	0.955	0.606	0.633	0.345
	$n = 100$	0.946	0.975	0.753	0.698	0.323
	$n = 300$	0.973	0.959	0.746	0.660	0.308
$\widehat{r}_2$:	$n = 20$	1.077	0.960	0.641	0.767	0.523
	$n = 50$	1.044	0.972	0.741	0.726	0.482
	$n = 100$	0.983	0.936	0.741	0.668	0.420
	$n = 300$	0.965	0.947	0.758	0.675	0.313
$\widehat{\boldsymbol{a}}_1^*$:	$n = 20$	1.102	0.942	0.381	0.592	0.283
	$n = 50$	1.032	0.957	0.495	0.637	0.226
	$n = 100$	0.988	0.948	0.685	0.651	0.265
	$n = 300$	1.072	0.955	0.757	0.694	0.289
$\widehat{\boldsymbol{a}}_2^*$:	$n = 20$	1.088	0.946	0.523	0.696	0.405
	$n = 50$	0.995	0.944	0.562	0.650	0.290
	$n = 100$	0.987	0.936	0.720	0.661	0.313
	$n = 300$	1.098	0.969	0.766	0.692	0.312
	$n = \infty$	0.982	0.953	0.786	0.667	0.284

is slower when the canonical correlations are closer to each other, and hence the canonical vectors of different orders harder to distinguish.

In the third case samples were generated from a 6-variate normal distribution, so $p = q = 3$. Efficiencies of the first canonical correlation and vector estimates are reported in Table 4. Again, as n increases, the efficiencies seem to converge to the asymptotic ones. Similar conclusions as for the first simulation scheme hold; again the convergence of RMCD- and MCD-estimators is slower than the convergence of the others.

To compute the estimators, the FAST-MCD algorithm of Rousseeuw and Van Driessen (1999) was used for computation of the 25% breakdown point MCD and RMCD estimators.

Table 3: Finite-sample efficiencies of the canonical correlation and vector estimates. Samples were generated from a 4-variate normal distribution. The quantities to be estimated were $\rho_1 = 0.6$, $\rho_2 = 0.4$, $\boldsymbol{a}_1^{*T} = (1, 0)^T$ and $\boldsymbol{a}_2^{*T} = (0, 1)^T$.

		SCM	S	RMCD	Tyler	MCD
$\widehat{r}_1$:	$n = 20$	1.081	0.942	0.513	0.655	0.403
	$n = 50$	0.933	0.934	0.582	0.628	0.294
	$n = 100$	1.016	0.928	0.701	0.693	0.302
	$n = 300$	1.034	0.977	0.782	0.672	0.291
$\widehat{r}_2$:	$n = 20$	0.975	0.986	0.738	0.786	0.688
	$n = 50$	1.001	0.956	0.645	0.717	0.399
	$n = 100$	0.996	0.956	0.715	0.642	0.324
	$n = 300$	0.936	0.972	0.759	0.647	0.287
$\widehat{\boldsymbol{a}}_1^*$:	$n = 20$	1.054	0.952	0.775	0.860	0.716
	$n = 50$	0.962	0.915	0.646	0.704	0.471
	$n = 100$	1.088	0.984	0.677	0.658	0.339
	$n = 300$	1.004	0.965	0.696	0.635	0.202
$\widehat{\boldsymbol{a}}_2^*$:	$n = 20$	1.075	0.960	0.812	0.859	0.745
	$n = 50$	0.959	0.905	0.681	0.708	0.506
	$n = 100$	1.093	0.979	0.693	0.672	0.381
	$n = 300$	1.022	0.961	0.719	0.652	0.222
	$n = \infty$	0.982	0.953	0.786	0.667	0.284

Table 4: Finite-sample efficiencies of the first canonical correlation and vector estimates. Samples were generated from a 6-variate normal distribution. The quantities to be estimated were $\rho_1 = 0.9$ and $\boldsymbol{a}_1^{*T} = (1,0)^T$.

		SCM	S	RMCD	Tyler	MCD
$\widehat{r}_1$:	$n = 20$	1.025	0.996	0.493	0.729	0.454
	$n = 50$	0.987	0.961	0.675	0.724	0.422
	$n = 100$	0.970	0.964	0.758	0.706	0.394
	$n = 300$	0.991	0.940	0.793	0.688	0.350
$\widehat{\boldsymbol{a}}_1^*$:	$n = 20$	1.043	0.922	0.281	0.691	0.270
	$n = 50$	0.955	0.931	0.477	0.690	0.301
	$n = 100$	0.966	0.969	0.656	0.698	0.316
	$n = 300$	0.972	0.952	0.783	0.701	0.345
	$n = \infty$	0.991	0.975	0.837	0.750	0.356

The S-estimator has been computed with the surreal algorithm of Ruppert (1992). For the computation of the SCM, the same approximations as in Ollila et al (2003b, section 7) were used.

4.2 An example

In this Section we apply the proposed methods through a simple example. We consider the Linnerud data (Tenenhaus, p. 15) consisting of 20 observations and wish to describe the relationships between two sets of variables, namely x_1=weight, x_2=waist measurement, x_3=pulse and y_1=pull-ups, y_2=bendings, y_3=jumps. In order to compare the methods proposed above, we consider canonical correlation and vector estimates obtained from different shape matrices. Estimates as well as corresponding standard deviations, obtained using the asymptotic results given in Corollary 2, are listed in Table 5. The coefficients of the different canonical vectors are often used to interpret the canonical variates, since they give the weight of every variable. By reporting the standard error around these coefficients, one can quickly see whether these coefficients are significantly different from zero or not. Although reporting these standard errors is no common practice in canonical analysis (probably also because the asymptotic distribution of the canonical vectors has only been established recently, even in the classical case), it helps to detect non-significant coefficients and it helps avoiding overinterpretation. For example, one sees that for all shape matrices considered $\boldsymbol{a}_1^*$ is mainly determined by x_2, and to a lesser extend by x_1. On the other hand, $\boldsymbol{b}_1^*$ is mainly determined by y_2, and to a lesser extend by y_3. Note that standard errors are larger for the less efficient estimators, like the MCD. Differences between the different estimation procedures do not seem to be substantial. A more detailed look is revealed by the plot of the the first canonical variates (x_1', y_1') in the Figure 1. The fitted lines are resulting from the canonical analysis, having as equation $y_1' = \hat{\rho}_1 x_1'$. We see that the Classical and the SCM approach, both having a zero breakdown point, have been attracted by the outliers in the upper right and lower left corner of the plot. The MCD and RMCD have been more resistant with respect to these outliers, and the data cloud is more concentrated around the linear fit, as is also witnessed by the higher values for the first correlation coefficent of these estimators.

5 Conclusion

Results concerning the asymptotic distribution for the canonical correlations only have received much attention in the literature (e.g. Hsu 1941, Eaton and Tyler 1994), but much less attention has been given to the limiting distribution of canonical vectors. Anderson (1999) reviews previous work on the asymptotics of canonical analysis, and clearly states the asymptotic variances and covariances of both canonical correlations and vectors derived from the sample covariance matrix. It is not without interest to have information on the asymptotic variance of the canonical vectors since it allows, for example, to compute (asymptotic) standard errors around the coefficients of the canonical vectors. Since these coefficients are often interpreted as the contributions of the original marginal variables to the canonical vectors, it is useful to check on their significance.

In this paper a full treatment of the asymptotic distribution of the canonical correlations and canonical vectors derived from any regular affine equivariant scatter matrix estimator is given. Results do not only hold at the normal, but at any elliptical distribution where the scatter matrix being used is well defined and asymptotically normal. Moreover, we allow for a different dimension of the two multivariate variables $\boldsymbol{x}$ and $\boldsymbol{y}$, a situation often occuring in practice. The advantage of working with shape matrices, yielding standardized canonical vectors, has also been pointed out. Also here, a full treatment of the asymptotic distribution of the canonical correlations and standardized canonical vectors derived from any regular affine equivariant shape matrix estimator has been presented.

Table 5: Canonical correlation and vector estimates for the Linnerud data given by the classical shape matrix, the SCM-, the S-, the RMCD- based, Tyler's, and the MCD-based shape matrix. The standard deviations are reported between parentheses.

	Classical			SCM			S		
$\widehat{\boldsymbol{r}}$	0.796	0.201	0.073	0.774	0.168	0.010	0.768	0.122	0.036
	(0.082)	(0.215)	(0.222)	(0.090)	(0.218)	(0.225)	(0.093)	(0.223)	(0.226)
$\widehat{\boldsymbol{a}}_1^*$	0.332	-5.213	0.087	0.336	-5.432	0.128	0.336	-5.552	0.112
	(0.154)	(1.069)	(0.271)	(0.163)	(1.141)	(0.287)	(0.176)	(1.228)	(0.318)
$\widehat{\boldsymbol{a}}_2^*$	-0.807	3.897	-0.339	-0.741	4.247	0.355	-0.562	3.518	0.813
	(0.186)	(2.750)	(2.050)	(0.475)	(1.988)	(2.004)	(1.444)	(7.155)	(3.086)
$\widehat{\boldsymbol{a}}_3^*$	0.082	-1.670	-1.540	-0.342	0.733	-1.516	-0.682	3.290	-1.462
	(1.097)	(5.485)	(0.510)	(0.998)	(5.906)	(0.526)	(1.202)	(7.664)	(1.741)
$\widehat{\boldsymbol{b}}_1^*$	0.699	0.178	-0.148	0.719	0.178	-0.150	0.584	0.182	-0.150
	(0.476)	(0.044)	(0.047)	(0.495)	(0.044)	(0.052)	(0.484)	(0.044)	(0.055)
$\widehat{\boldsymbol{b}}_2^*$	-0.751	0.021	0.219	-0.956	0.036	0.221	-1.415	0.052	0.213
	(3.457)	(0.283)	(0.127)	(3.228)	(0.260)	(0.164)	(4.361)	(0.373)	(0.327)
$\widehat{\boldsymbol{b}}_3^*$	2.592	-0.209	0.086	2.436	-0.192	0.117	2.063	-0.175	0.153
	(1.101)	(0.065)	(0.300)	(1.345)	(0.076)	(0.299)	(3.026)	(0.126)	(0.455)

	RMCD			Tyler			MCD		
$\widehat{\boldsymbol{r}}$	0.826	0.431	0.110	0.801	0.084	0.014	0.868	0.442	0.144
	(0.078)	(0.199)	(0.241)	(0.092)	(0.256)	(0.258)	(0.092)	(0.302)	(0.367)
$\widehat{\boldsymbol{a}}_1^*$	0.432	-7.402	0.275	0.271	-5.825	-0.006	0.328	-6.479	0.443
	(0.192)	(1.676)	(0.352)	(0.192)	(1.465)	(0.325)	(0.225)	(2.112)	(0.426)
$\widehat{\boldsymbol{a}}_2^*$	0.715	4.963	-1.581	0.741	-4.406	1.633	-0.552	3.668	-1.516
	(0.347)	(3.206)	(0.333)	(1.965)	(11.533)	(1.157)	(0.634)	(5.217)	(0.481)
$\widehat{\boldsymbol{a}}_3^*$	0.523	-3.114	-0.355	-0.611	3.542	0.351	-0.683	4.629	0.255
	(0.465)	(3.799)	(0.977)	(2.386)	(14.307)	(5.253)	(0.577)	(4.672)	(1.480)
$\widehat{\boldsymbol{b}}_1^*$	0.154	0.191	-0.195	0.304	0.187	-0.200	0.362	0.161	-0.157
	(0.343)	(0.036)	(0.077)	(0.382)	(0.045)	(0.076)	(0.430)	(0.049)	(0.093)
$\widehat{\boldsymbol{b}}_2^*$	0.675	0.030	-0.315	-1.650	0.080	0.179	0.341	0.054	-0.312
	(1.026)	(0.090)	(0.097)	(3.331)	(0.482)	(0.976)	(1.657)	(0.130)	(0.118)
$\widehat{\boldsymbol{b}}_3^*$	1.825	-0.109	0.087	1.035	-0.149	0.304	1.877	-0.117	0.055
	(0.507)	(0.062)	(0.203)	(5.310)	(0.266)	(0.583)	(0.583)	(0.092)	(0.309)

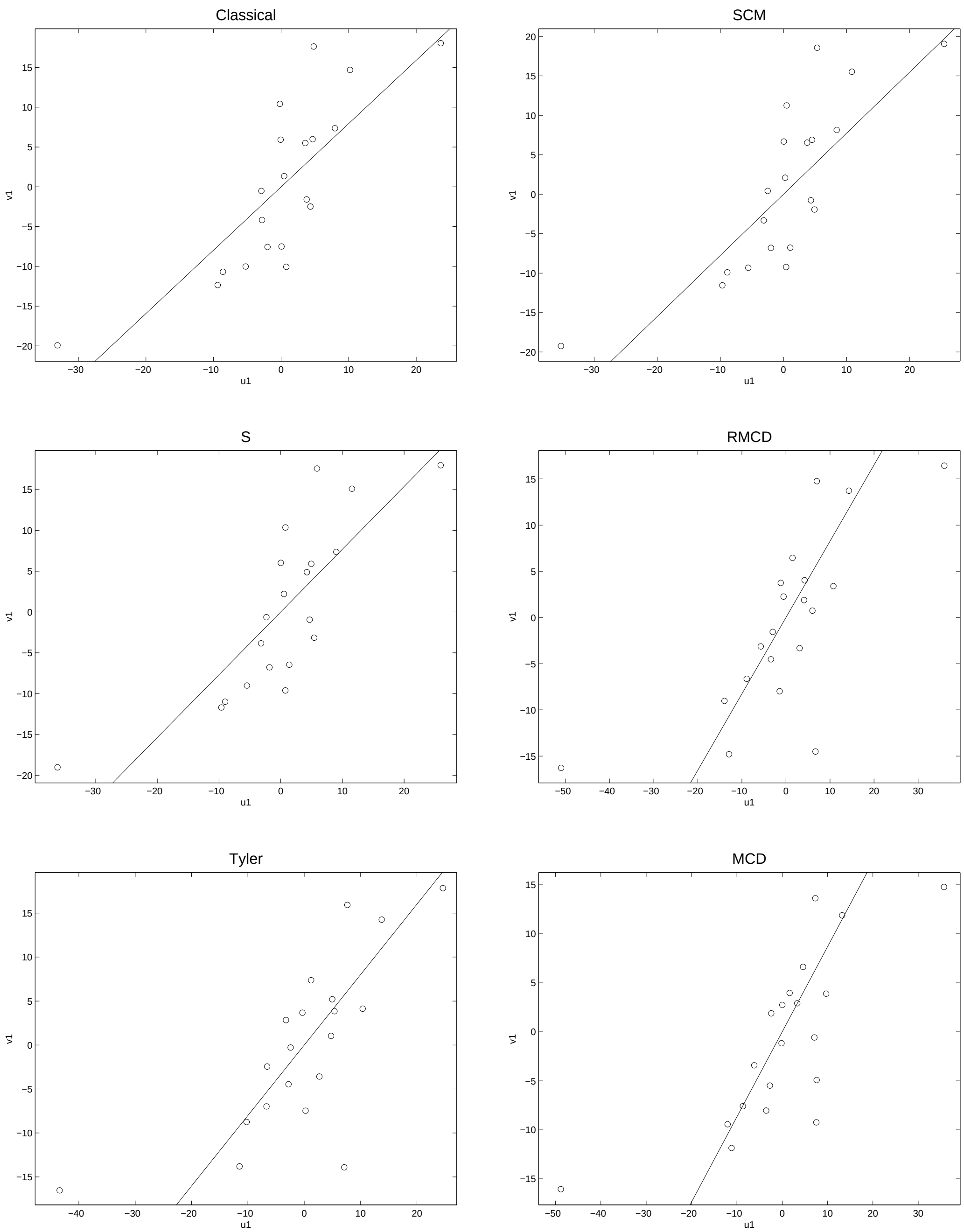

Figure 1: Scatterplot of the first canonical variates based on classical and robust shape matrices.

Appendix

Proof of Theorem 1 The canonical variates $\boldsymbol{z}'$ follow an elliptical distribution F' with $C(F')$ as described at the beginning of Section 3.2. Then there exists a symmetric positive definite matrix $H = C(F')^{-1/2}$ such that $\boldsymbol{z}_0 = H\boldsymbol{z}'$ follows a spherical distribution F_0. Write $r^2 = \|\boldsymbol{z}_0\|^2 = \boldsymbol{z}'^T C(F')^{-1}\boldsymbol{z}'$ and $\boldsymbol{z}_0/\|\boldsymbol{z}_0\| = (\boldsymbol{s}^T, \boldsymbol{t}^T)^T$. Then r and $(\boldsymbol{s}^T, \boldsymbol{t}^T)$ are independent and the latter variable is uniformly distributed at the periphery of the k-variate unit-sphere. It turns out to be convenient to write canonical variates as functions of spherical variables:

$$\boldsymbol{z}' = rH^{-1}(\boldsymbol{s}^T, \boldsymbol{t}^T)^T \tag{11}$$

where

$$H^{-1} = \begin{pmatrix} \sum_{j=1}^{p} H_j & 0 \\ 0 & I_{q-p} \end{pmatrix}$$

and H_j is a $2p \times 2p$ matrix with four non-zero elements namely $[H_j]_{i,i} = [H_i]_{p+i,p+i} = (1 + \Delta_i^2)^{-1/2}$ and $[H_j]_{i,p+i} = [H_i]_{p+i,i} = \Delta_i(1 + \Delta_i^2)^{-1/2}$, where $\rho_i = 2\Delta_i(1 + \Delta_i^2)^{-1}$ for $1 \le i \le p$. The Mahalanobis angle of $\boldsymbol{z}'$ equals then $(\boldsymbol{u}^T, \boldsymbol{v}^T)^T = H^{-1}(\boldsymbol{s}^T, \boldsymbol{t}^T)^T$.

Equation (4) gives

$$IF(\boldsymbol{z}_0; C, F_0) = \gamma_C(r) \begin{pmatrix} \boldsymbol{s} \\ \boldsymbol{t} \end{pmatrix} (\boldsymbol{s}^T, \boldsymbol{t}^T) - \delta_C(r) I_k$$

and affine equivariance of C yields

$$\begin{aligned} IF(\boldsymbol{z}'; C, F') &= H^{-1} IF(H\boldsymbol{z}'; C, F_0)(H^{-1})^T \\ &= \gamma_C(r) \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} (\boldsymbol{u}^T, \boldsymbol{v}^T) - \delta_C(r) \begin{pmatrix} I_p & (R, 0) \\ (R, 0)^T & I_q \end{pmatrix}. \end{aligned} \tag{12}$$

Combining (12) with the formulas derived before stating the theorem already yield the expressions for the influence functions at F'.

Write now $R(G)$, $A(G)$ and $B(G)$ as $R(\boldsymbol{x}^T, \boldsymbol{y}^T)^T$, $A(\boldsymbol{x}^T, \boldsymbol{y}^T)^T$ and $B(\boldsymbol{x}^T, \boldsymbol{y}^T)^T$ if the cdf of $(\boldsymbol{x}^T, \boldsymbol{y}^T)^T$ is G. The affine invariance and equivariance properties imply then

$$\begin{aligned} R(\boldsymbol{x}^T, \boldsymbol{y}^T)^T &= R(\tilde{A}\boldsymbol{x}^T, \tilde{B}\boldsymbol{y}^T)^T \\ A(\boldsymbol{x}^T, \boldsymbol{y}^T)^T &= \tilde{A}\, A(\tilde{A}\boldsymbol{x}^T, \tilde{B}\boldsymbol{y}^T)^T \\ B_1(\boldsymbol{x}^T, \boldsymbol{y}^T)^T &= \tilde{B}\, B_1(\tilde{A}\boldsymbol{x}^T, \tilde{B}\boldsymbol{y}^T)^T. \end{aligned}$$

for every $p \times p$ matrix $\tilde{A}$ and every $q \times q$ matrix $\tilde{B}$. Then by the definition of the influence function and equivariance and invariance properties we have

$$IF(\boldsymbol{z}; R, F) = IF(\boldsymbol{z'}; R, F'),$$

$$IF(\boldsymbol{z}; A, F) = \lim_{\epsilon \to 0} \frac{A(F_\epsilon) - A(F)}{\epsilon} = \lim_{\epsilon \to 0} \frac{A((1-\epsilon)F + \epsilon \Delta_{\boldsymbol{z}}) - A(F)}{\epsilon}$$
$$= A(F) \lim_{\epsilon \to 0} \frac{A((1-\epsilon)F' + \epsilon \Delta_{\boldsymbol{z'}}) - A(F')}{\epsilon} = A(F) IF(\boldsymbol{z'}; A, F')$$

and similarly

$$IF(\boldsymbol{z}; B_1, F) = B(F) IF(\boldsymbol{z'}; B_1, F').$$

From the above relations between the influence functions at F and F', the desired influence functions follow.

Proof of Theorem 2 First note that the canonical correlations derived from V or the associated scatter matrix C are the same. Therefore it follows form Theorem 1 and (6) that

$$IF(\boldsymbol{z}'; R, F') = \gamma_V(r) H_1(\boldsymbol{u}, \boldsymbol{v}; R).$$

By Theorem 1 the influence functions of $A^* = [Det(C)]^{1/2k} A$ and $B_1^* = [Det(C)]^{1/2k} B_1$ are

$$IF(\boldsymbol{z}'; A^*, F') = [Det(C(F'))]^{1/2k} IF(\boldsymbol{z}'; A, F') + A(F') IF(\boldsymbol{z}'; [Det(C)]^{1/2k}, F')$$
$$= \left|I_p - R^2\right|^{1/2k} \left[IF(\boldsymbol{z}'; A, F') + \frac{1}{2k} \left|I_p - R^2\right|^{-1} IF(\boldsymbol{z}'; Det(C), F') \right]$$
$$= \left|I_p - R^2\right|^{1/2k} \left[\gamma_C(R) H_2(\boldsymbol{u}, \boldsymbol{v}; R) + \frac{1}{2} \delta_C(r) I_p + \frac{1}{2k} \gamma_C(r) I_p - \frac{1}{2} \delta_C(r) I_p \right]$$
$$= \left|I_p - R^2\right|^{1/2k} \gamma_V(r) \left[H_2(\boldsymbol{u}, \boldsymbol{v}; R) + \frac{1}{2k} I_p \right],$$

where it was used that $IF(\boldsymbol{z}'; Det(C), F') = Det(C(F')) IF(\boldsymbol{z_0}; Det(C), F_0)$ together with (5). Similarly

$$IF(\boldsymbol{z}'; B_1^*, F') = \left|I_p - R^2\right|^{1/2k} \gamma_V(r) \left[H_3(\boldsymbol{u}, \boldsymbol{v}; R) + \frac{1}{2k} \begin{pmatrix} I_p \\ 0 \end{pmatrix} \right].$$

The affine invariance and equivariance properties of the functionals R, A^* and B^* yield

$$R(\boldsymbol{x}^T, \boldsymbol{y}^T)^T = R(\tilde{A}\boldsymbol{x}^T, \tilde{B}\boldsymbol{y}^T)^T$$
$$A^*(\boldsymbol{x}^T, \boldsymbol{y}^T)^T = \left|I_p - R^2\right|^{-1/2k} \tilde{A}\, A^*(\tilde{A}\boldsymbol{x}^T, \tilde{B}\boldsymbol{y}^T)^T$$
$$B_1^*(\boldsymbol{x}^T, \boldsymbol{y}^T)^T = \left|I_p - R^2\right|^{-1/2k} \tilde{B}\, B_1^*(\tilde{A}\boldsymbol{x}^T, \tilde{B}\boldsymbol{y}^T)^T.$$

for every $p \times p$ matrix $\tilde{A}$ and every $p \times q$ matrix $\tilde{B}$. So at elliptical F the influence functions become

$$IF(\boldsymbol{z}; R, F) = IF(\boldsymbol{z}'; R, F') = \gamma_V(r) H_1(\boldsymbol{u}, \boldsymbol{v}; R),$$

$$\begin{aligned} IF(\boldsymbol{z}; A^*, F) &= \left| I_p - R^2 \right|^{-1/2k} A^*(F)\, IF(\boldsymbol{z}'; A^*, F') \\ &= A^*(F)\gamma_V(r) \left[H_2(\boldsymbol{u}, \boldsymbol{v}; R) + \frac{1}{2k} I_p \right] \end{aligned}$$

and

$$IF(\boldsymbol{z}; B_1, F) = B^*(F)\gamma_V(r) \left[H_3(\boldsymbol{v}, \boldsymbol{v}; R) + \frac{1}{2k} \begin{pmatrix} I_p \\ 0 \end{pmatrix} \right].$$

Proof of Theorem 3 The asymptotic normality of $\widehat{R}$, $\widehat{A}$ and $\widehat{B}_1$ follows simply by the delta-method, see for example Anderson (1999). The asymptotic variances are obtained by using Theorem 1 and the following property of vec-operator: $vec(BCD) = (D^T \otimes B) vec(C)$. Consider for example the asymptotic variance of $A(F)$. Write

$$IF(\boldsymbol{z}; A, F) = A(F) \left[\gamma_C(R) H_2(\boldsymbol{u}, \boldsymbol{v}; R) + \frac{1}{2} \delta_C(r) I_p \right] := AJ.$$

Then

$$\begin{aligned} ASV(\widehat{A}; F) &= E\left[vec\{AJI_p\} vec\{AJI_p\}^T \right] \\ &= E\left[(I_p \otimes A) vec\{J\} \left[(I_p \otimes A) vec\{J\} \right]^T \right] \\ &= (I_p \otimes A) E\left[vec\{J\} vec\{J\}^T \right] (I_p \otimes A^T) \\ &= (I_p \otimes A) ASV(\widehat{A}; F') (I_p \otimes A^T). \end{aligned}$$

Proof of Theorem 4 As the proof of Theorem 3.

Proof of Theorem 5 Consider for example the limiting variance of $\widehat{r}_i$, for an $1 \le i \le p$. Theorem 4 gives

$$\begin{aligned} ASV(\widehat{r}_i; F') &= E[IF(\boldsymbol{z}'; R_{ii}, F')^2] = E\left[\gamma_C^2(r) H_1(\boldsymbol{u}, \boldsymbol{v}, R(F')_{ii}^2 \right] \\ &= E[\gamma_C^2(r) \left(u_i v_i - \frac{1}{2} \rho_i u_i^2 - \frac{1}{2} \rho_i v_i^2 \right)^2] \end{aligned}$$

Use now the transformation given in the first paragraph of the proof of Theorem 1:

$$u_i = \frac{s_i + \Delta_i t_i}{\sqrt{1+\Delta_i^2}} \quad \text{and} \quad v_i = \frac{\Delta_i s_i + t_i}{\sqrt{1+\Delta_i^2}},$$

where s_i and t_i are different marginals of a vector distributed uniformly on the periphery pf the k dimensional unit sphere, and also independent of r. Then, after some tedious calculations,

$$\begin{aligned} ASV(\widehat{r}_i; F') &= (1-\rho_i^2)^2 E[\gamma_C^2(r)] E[s_i^2 t_i^2] \\ &= (1-\rho_i^2)^2 \frac{E[\gamma_C^2(r)]}{k(k+2)} \\ &= (1-\rho_i^2)^2 ASV(C_{12}; F_0). \end{aligned}$$

When carrying out the calculations, symmetry properties of s_i and t_i can be used, together with $E[s_i^2] = 1/k, E[s_i^4] = 3/(k(k+2))$, and $E[s_i^2 t_i^2] = 1/(k(k+2))$ (see lemma 5 in Ollila et al., 2003b)

Other limiting variances and covariances are obtained in a more or less similar way, by carefully carrying out computations along the lines above.

Proof of Theorem 6 As the proof of Theorem 5.

References

Anderson, T.W. (1999). Asymptotic theory for canonical correlation analysis *Journal of multivariate analysis*, **70**, 1–29.

Croux, C. and Dehon, C. (2002). Analyse canonique basée sur des estimateurs robustes de la matrice de covariance. *La Revue de Statistique Appliquee*, **2**, 5-26.

Croux, C., and Haesboeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis,* **71**, 161-190.

Croux, C., and Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation Matrix: influence Functions and efficiencies. *Biometrika*, **87**,603-618.

Davies, P.L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Annals of Statistics*, **15**, 1269–1292.

Eaton, M.L. and Tyler, D. (1994). The asymptotic distribution of singular values with applications to canonical correlation analysis. *Journal of Multivariate Analysis*, **50**, 238–264.

Hampel, F.R., Ronchetti, E.M, Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust statistics: the approach based on influence functions.* Wiley, New York.

Hettmansperger, T.P. and Randles, R. (2002). A practical affine equivariant multivariate median. *Biometrika*, **89**, 851–860.

Hotelling, H. (1936). Relations between two sets of variables. *Biometrika*, **28**, 321–377.

Hsu, P.L. (1941). On the Limiting Distribution of the Canonical Correlations. *Biometrika*, **32**, 38-45.

Johnson, R.A., and Wichern, D.W. (1998). *Applied Multivariate Statistical Analysis, 4th edition.* Prentice-Hall, New York.

Karnel, G. (1991). Robust canonical Ccrrelation and correspondance analysis. *The Frontiers of Statistical Scientific and Industrial Applications*, Volume II of the Proceedings of ICOSCO-I, The First International Conference on Statistical Computing, 335-354.

Lopuhaä, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and scatter. *Annals of Statistics* **17**, 1662–1683.

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

Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: A Review. *Scandinavian Journal of Statistics*, **26**, 319–343.

Ollila, E., Hettmansperger, T.P. and Oja, H. (2003a). Affine equivariant multivariate sign methods. Under revision.

Ollila, E., Oja, H. and Croux, C. (2003b). The Affine equivariant sign covariance matrix: Asymptotic behavior and efficiencies. *Journal of Multivariate Analysis*, in print.

Rousseeuw, P.J. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications, Vol. B,* eds. W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Dordrecht: Reidel, 283-297.

Rousseeuw, P.J., and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41, 212–223.

Ruppert, D. (1992). Computing S-estimators for regression and multivariate location/dispersion. *Journal of Computational and Graphical Statistics*, **1**, 253–270.

Tyler, D.E. (1982). Radial estimates and the test for sphericity. *Biometrika*, **69**, 429–436.

Tyler, D.E. (1987). A distribution-free M-estimator of multivariate scatter. *The Annals of Statistics*, **15**, 234–251.

Visuri, S., Koivunen, V. and Oja, H. (2000). Sign and rank covariance matrices. *Journal of Statistical Planning and Inference*, **91**, 557–575.