Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT

Journal of Multivariate Analysis 87 (2003) 328–355

Journal of Multivariate Analysis

http://www.elsevier.com/locate/jmva

# The affine equivariant sign covariance matrix: asymptotic behavior and efficiencies

Esa Ollila,[a,*] Hannu Oja,[b] and Christophe Croux[c]

[a] *Signal Processing Laboratory, Helsinki University of Technology, P.O. Box 3000, Espoo FIN-02015, Finland*

[b] *University of Jyväskylä, Jyväskylä, Finland*

[c] *University of Leuven, Leuven, Belgium*

Received 24 September 2001

**Abstract**

We consider the affine equivariant sign covariance matrix (SCM) introduced by Visuri et al. (J. Statist. Plann. Inference 91 (2000) 557). The population SCM is shown to be proportional to the inverse of the regular covariance matrix. The eigenvectors and standardized eigenvalues of the covariance matrix can thus be derived from the SCM. We also construct an estimate of the covariance and correlation matrix based on the SCM. The influence functions and limiting distributions of the SCM and its eigenvectors and eigenvalues are found. Limiting efficiencies are given in multivariate normal and $t$-distribution cases. The estimates are highly efficient in the multivariate normal case and perform better than estimates based on the sample covariance matrix for heavy-tailed distributions. Simulations confirmed these findings for finite-sample efficiencies.



*AMS 2000 subject classifications:* 62G35; 62G05; 62H12

*Keywords:* Affine equivariance; Covariance and correlation matrices; Efficiency; Eigenvectors and eigenvalues; Influence function; Multivariate median; Multivariate sign; Robustness

*Corresponding author.

*E-mail addresses:* esollila@wooster.hut.fi (E. Ollila), ojahannu@maths.jyu.fi (H. Oja), christophe.croux@econ.kuleuven.ac.be (C. Croux).

0047-259X/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S0047-259X(03)00045-9

## 1. Introduction

Let $\boldsymbol{X}$ be a $k$-dimensional random vector with finite second-order moments. Denote $\Sigma$ its covariance matrix, which we suppose to be non-singular. The spectral decomposition of the covariance matrix is given by $\Sigma = P\Lambda P^T$, where $P$ is the matrix with the eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ of $\Sigma$ in its columns and $\Lambda$ is a diagonal matrix with the corresponding eigenvalues $\lambda_1, \ldots, \lambda_k$ as diagonal elements. We may also state the spectral decomposition in the form $\Sigma = \lambda P\Lambda^* P^T$, where $\lambda = (\lambda_1 \cdots \lambda_k)^{1/k}$ is the geometrical mean of the eigenvalues and $\Lambda = \lambda\Lambda^*$. The matrix $\Lambda^*$ is then a diagonal matrix of *standardized eigenvalues*

$$\lambda_j^* = \frac{\lambda_j}{(\lambda_1 \cdots \lambda_k)^{1/k}}. \tag{1}$$

Bensmail and Celeux [2] use the terms *scale*, *shape* and *orientation* for the items $\lambda$, $\Lambda^*$ and $P$.

In this article we consider the affine equivariant *sign covariance matrix* (SCM) which can be used to estimate the shape $\Lambda^*$ and orientation $P$ of the covariance matrix. Under a specified elliptical model distribution a consistent estimate of $\Sigma$ can be obtained. The SCM estimator has been proposed by Visuri et al. [26], but its asymptotic properties have not yet been considered. The SCM estimator is based on the concept of *affine equivariant signs*, which have been applied for hypothesis testing in the multivariate one sample case [12] and for MANOVA [13]. For a review of multivariate signs and ranks, see [20].

The eigenvectors of the SCM can serve for a more robust version of classical *principal components analysis* (PCA). Using robust covariance matrix estimators for performing robust PCA has first been considered by Devlin et al. [10] by means of $M$-estimators. More recently, Croux and Haesbroeck [6] computed influence functions and efficiencies for eigenvectors and eigenvalues of high breakdown estimators of covariance. A PCA based on the SCM will not have a positive breakdown point, but is shown to be highly efficient at normal and heavier tailed distributions. Moreover, by using multivariate signs, the approach gets a non-parametric flavor.

Section 2 introduces the sample SCM matrix and its population counterpart, whereas Section 3 explicits the relation between the population covariance matrix and the SCM at location-scale families. The main contribution of the paper is the derivation of the influence function and limiting distribution of the SCM, treated in Section 4. Asymptotic behavior of the eigenvectors and standardized eigenvalues of the SCM are derived in the next section. Section 6 shows how one can easily obtain estimates for the population covariance and correlation matrix. Finally, by means of a modest simulation study the asymptotic efficiencies are compared with finite sample ones.

Throughout the paper, we will use lowercase boldface letter to denote a vector and an uppercase boldface letter instead if an underlying distribution is assumed. By

$\mathbb{E}_F(\boldsymbol{X})$ we will mean the expectation of a random vector $\boldsymbol{X}$ from a distribution $F$. All the proofs are reserved for the Appendix.

## 2. Affine equivariant sign covariance matrix

In the univariate case the sign of $x$ with respect to $\theta$ is the derivative of

$$V(x;\theta) = \mathrm{abs}\left\{\det\begin{pmatrix} 1 & 1 \\ \theta & x \end{pmatrix}\right\} = \mathrm{abs}\{x-\theta\}$$

with respect to $x$, that is $S(x;\theta) = \mathrm{sign}\{x-\theta\}$. The sample median is known to minimize the sum of the volumes (lengths of line segments or univariate simplices) $V(x_i;\theta)$, where $x_i$ are the data points. The empirical signs are then taken with respect to the sample median $\hat{\theta}$ and denoted by $\hat{S}_i = S(x_i;\hat{\theta})$ for $i = 1,\dots,n$. They are centered since $\sum \hat{S}_i = 0$.

Next, we extend this definition to the multivariate setting. Let $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ be $k$-variate ($k>1$) data set. The multivariate Oja [19] median $\hat{\boldsymbol{\theta}}$ minimizes the criterion function $\sum_{i_1<\dots<i_k} V(\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_k},\boldsymbol{\theta})$, where

$$V(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k,\boldsymbol{x}_{k+1}) = \frac{1}{k!}\mathrm{abs}\left\{\det\begin{pmatrix} 1 & \cdots & 1 & 1 \\ \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_k & \boldsymbol{x}_{k+1} \end{pmatrix}\right\}$$

is the volume of the $k$-variate simplex determined by the vertices $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k,\boldsymbol{x}_{k+1}$. To shorten the notations, write $I = (i_1,\dots,i_{k-1})$ with $1 \leqslant i_1 < \dots < i_{k-1} \leqslant n$, for an ordered set of indices. This new index $I$ then refers to a $k-1$ subset of observations with indices listed in $I$. The multivariate empirical sign vector of $k$-variate vector $\boldsymbol{x}$ with respect to $\boldsymbol{\theta}$ is the gradient of

$$\begin{aligned} V_n(\boldsymbol{x};\boldsymbol{\theta}) &= \frac{1}{\binom{n}{k-1}}\sum_I \mathrm{abs}\left\{\det\begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \\ \boldsymbol{\theta} & \boldsymbol{x}_{i_1} & \cdots & \boldsymbol{x}_{i_{k-1}} & \boldsymbol{x} \end{pmatrix}\right\} \\ &= \mathrm{ave}_I\{\mathrm{abs}[\det(\boldsymbol{x}_{i_1}-\boldsymbol{\theta} \quad \cdots \quad \boldsymbol{x}_{i_{k-1}}-\boldsymbol{\theta} \quad \boldsymbol{x}-\boldsymbol{\theta})]\} \qquad (2) \\ &= \mathrm{ave}_I\{\mathrm{abs}[\boldsymbol{e}^T(I;\boldsymbol{\theta})(\boldsymbol{x}-\boldsymbol{\theta})]\} \end{aligned}$$

with respect to $\boldsymbol{x}$. Here $\boldsymbol{e}(I;\boldsymbol{\theta})$ is the vector of cofactors corresponding to the last column of the matrix in (2) and the sums and the average go over all possible $k-1$ subsets $I$. Then the sign vector of $\boldsymbol{x}$ with respect to $\boldsymbol{\theta}$ is simply

$$\boldsymbol{S}_n(\boldsymbol{x};\boldsymbol{\theta}) = \mathrm{ave}_I\{\mathrm{sign}[\boldsymbol{e}^T(I;\boldsymbol{\theta})(\boldsymbol{x}-\boldsymbol{\theta})]\boldsymbol{e}(I;\boldsymbol{\theta})\}.$$

The empirical *multivariate signs* with respect to $\hat{\boldsymbol{\theta}}$ are then defined, as in the univariate case, by $\hat{\boldsymbol{S}}_i = \boldsymbol{S}_n(\boldsymbol{x}_i;\hat{\boldsymbol{\theta}})$, $i = 1,\dots,n$, where $\hat{\boldsymbol{\theta}}$ is the multivariate Oja median. These multivariate signs are thus centered, so $\sum \hat{\boldsymbol{S}}_i = \boldsymbol{0}$. The SCM is now simply defined as the usual covariance matrix computed from the empirical multivariate signs:

$$\hat{D} = \text{ave}_i\{\hat{\boldsymbol{S}}_i\hat{\boldsymbol{S}}_i^T\}.$$

The signs and the SCM enjoy the following affine equivariance property:

**Lemma 1.** *Let the sign vectors* $\hat{\boldsymbol{S}}_i^*$ *and the SCM* $\hat{D}^*$ *be calculated from the transformed observations* $\boldsymbol{x}_i^* = A\boldsymbol{x}_i + \boldsymbol{b}$ *with a non-singular matrix A and a k-dimensional vector* $\boldsymbol{b}$. *Then* $\hat{\boldsymbol{S}}_i^* = \text{abs}\{\det(A)\}(A^{-1})^T\hat{\boldsymbol{S}}_i$ *and* $\hat{D}^* = \det(A)^2(A^{-1})^T\hat{D}A^{-1}$.

In Fig. 1, a bivariate data set is pictured (left panel) together with the corresponding sign vectors $\hat{\boldsymbol{S}}_i$ (right panel). We see that the signs move the data points towards the periphery of an ellipse. The sign vector points roughly in the direction of the observation, while its magnitude depends on the dispersion of the data in the space orthogonal to the observation vector. The form of this ellipse is therefore merely determined by the *inverse* of the covariance structure of the data. Fig. 1 indicates that this structure has not been influenced by the outlier (marked by $\times$) present in the data cloud.

Next, we define the population counterparts of the multivariate median, signs and SCM. For an underlying distribution $F$, the theoretical Oja median $T(F)$ minimizes $\mathbb{E}_F[V(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, \boldsymbol{\theta})]$ or solves $\nabla\mathbb{E}_F[V(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k, T(F))] = \boldsymbol{0}$. The population multivariate sign of $\boldsymbol{x}$ with respect to $\boldsymbol{\theta}$ is given by

$$\boldsymbol{S}_F(\boldsymbol{x}; \boldsymbol{\theta}) = \mathbb{E}_F[\text{sign}\{\boldsymbol{E}^T(I; \boldsymbol{\theta})(\boldsymbol{x} - \boldsymbol{\theta})\}\boldsymbol{E}(I; \boldsymbol{\theta})],$$

where $\boldsymbol{E}(I; \boldsymbol{\theta})$ is the vector of cofactors of $\boldsymbol{x}$ in $(\boldsymbol{X}_1 - \boldsymbol{\theta} \;\cdots\; \boldsymbol{X}_{k-1} - \boldsymbol{\theta} \;\; \boldsymbol{x})$. Note that the population sign of $\boldsymbol{X}$ with respect to Oja median, $\boldsymbol{S}_F(\boldsymbol{X}; T(F))$, has expected

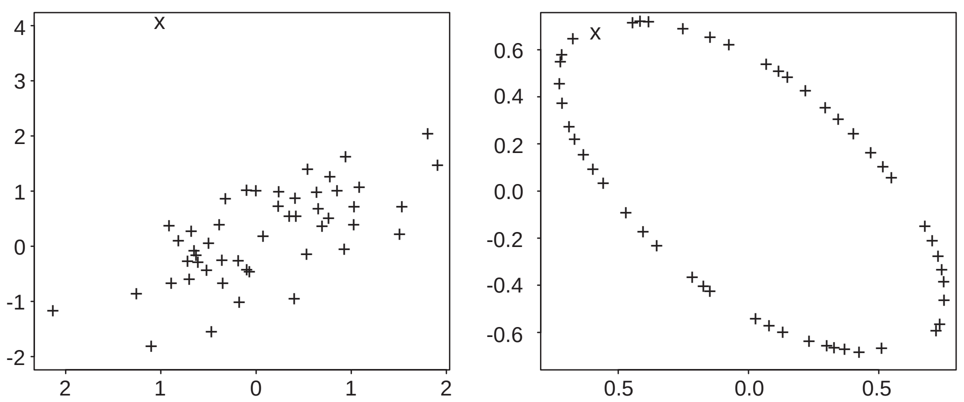

Fig. 1. Representation of a bivariate data cloud (left panel) together with the corresponding bivariate sign vectors (right panel) calculated from the data.

value zero. Finally, the population SCM is

$$D = D(F) = \mathbb{E}_F[\boldsymbol{S}_F(\boldsymbol{X}; T(F))\boldsymbol{S}_F^T(\boldsymbol{X}; T(F))].$$

Note that $D(F)$ exists if the first-order moments of $F$ are finite.

## 3. The relation between $\Sigma$ and the SCM

### 3.1. Elliptical model

Consider first the spherical case and let $F_0$ be the cdf of a $k$-variate spherical distribution with mean vector $\mathbf{0}$ and covariance matrix $I_k$ ( the $k \times k$ identity matrix), which we call a *standardized* spherical distribution. A spherically distributed random variable $\boldsymbol{X} \sim F_0$ can be decomposed as $\boldsymbol{X} = R\boldsymbol{U}$ where $R = ||\boldsymbol{X}||$ and $\boldsymbol{U} = \boldsymbol{X}/R$ are independent with $\boldsymbol{U}$ being uniformly distributed on the unit sphere. The Oja median $T(F_0)$ is then a zero vector and the population sign of $\boldsymbol{x}$ with respect to the Oja median equals (cf. Lemma A.2 in the appendix)

$$\boldsymbol{S}_{F_0}(\boldsymbol{x}; \mathbf{0}) = \frac{\Gamma^k(\frac{k}{2})\mathbb{E}_{F_0}^{k-1}(R)}{\sqrt{\pi}\Gamma^{k-1}(\frac{k+1}{2})}\boldsymbol{u} = c_{F_0}\boldsymbol{u}, \tag{3}$$

which is a constant times the direction vector $\boldsymbol{u} = \boldsymbol{x}/||\boldsymbol{x}||$. Thus, population signs at $F_0$ are on the sphere with radius $c_{F_0}$. For the $k$-variate standard normal distribution $F_0 = \Phi$, for example, $c_\Phi = 2^{(k-1)/2}\Gamma(k/2)\pi^{1/2}$. From (3) it readily follows that the theoretical SCM of $F_0$ equals

$$D(F_0) = \mathbb{E}_{F_0}[\boldsymbol{S}_{F_0}(\boldsymbol{X}; \mathbf{0})\boldsymbol{S}_{F_0}^T(\boldsymbol{X}; \mathbf{0})] = (c_{F_0}^2/k)I_k.$$

We used here that $E(U_i^2) = 1/k$ and $E(U_iU_j) = 0$, where $U_i$ and $U_j$ are distinct components of $\boldsymbol{U}$. (See Lemma A.1 in the appendix.)

Let $\boldsymbol{Z}$ be a random vector from a standardized spherical distribution $F_0$ and write $\boldsymbol{X} = \Sigma^{1/2}\boldsymbol{Z} + \boldsymbol{\mu}$, where $\Sigma$ is a positive definite symmetric $k \times k$ matrix and $\boldsymbol{\mu}$ a $k$-vector. Then the distribution $F$ of $\boldsymbol{X}$ is said to be elliptically symmetric with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$. Because of the affine equivariance of the Oja median, we have $T(F) = \boldsymbol{\mu}$. Moreover, due to affine equivariance of the population sign of $\boldsymbol{x} = \Sigma^{1/2}\boldsymbol{z} + \boldsymbol{\mu}$ with respect to $\boldsymbol{\mu}$ (as in Lemma 1),

$$\boldsymbol{S}_F(\boldsymbol{x}; \boldsymbol{\mu}) = \det(\Sigma^{1/2})\Sigma^{-1/2}\boldsymbol{S}_{F_0}(\boldsymbol{z}; \mathbf{0}),$$

implying that population signs at elliptical distributions are lying on the ellipsoid with center at the origin. The population SCM equals then

$$D(F) = (c_{F_0}^2/k)\det(\Sigma)\Sigma^{-1}.$$

### 3.2. Location-scale model

A similar intimate relation between the covariance matrix and the SCM still holds in a wider family of distributions which we call a multivariate location-scale family.

We start with a standardized (so having mean vector $\mathbf{0}$ and unit covariance matrix) random vector $\boldsymbol{Z}$ whose distribution is *reflection and permutation invariant* in the sense that $G\boldsymbol{Z} \sim \boldsymbol{Z}$ ($G\boldsymbol{Z}$ and $\boldsymbol{Z}$ have the same distribution) for every permutation or reflection $k \times k$ matrix $G$. A permutation matrix is obtained by permuting the rows or columns of the identity matrix and a reflection matrix is a diagonal matrix with diagonal elements $\pm 1$. For reflection and permutation invariant distributions, the marginal variables are identically distributed, symmetric about zero and uncorrelated. A *location-scale model* is then given by the family of distributions of the random vector $\boldsymbol{X} = \Sigma^{1/2}\boldsymbol{Z} + \boldsymbol{\mu}$ for every non-singular symmetric $k \times k$ matrix $\Sigma$ and $k$-vector $\boldsymbol{\mu}$. The mean vector and the covariance matrix of $\boldsymbol{X}$ are again $\boldsymbol{\mu}$ and $\Sigma$, respectively. Elliptical distributions belong to a location-scale family. A non-elliptic example is given by the choice where the margins of $\boldsymbol{Z}$ are iid random variables from a Laplace distribution with mean 0.

**Theorem 1.** *Let the distribution $F_0$ of $\boldsymbol{Z}$ be reflection and permutation invariant with mean vector $\mathbf{0}$ and covariance matrix $I_k$. Denote by $F$ the distribution of $\boldsymbol{X} = \Sigma^{1/2}\boldsymbol{Z} + \boldsymbol{\mu}$. Then $\mathbb{E}_F[\boldsymbol{S}_F\{\boldsymbol{X}; T(F)\}] = \mathbf{0}$ and $D(F) = w_{F_0}\det(\Sigma)\Sigma^{-1}$, where $w_{F_0}$ is a constant depending on $F_0$ only.*

Theorem 1 states that the population SCM is proportional to the inverse of the covariance matrix in a location-scale model. This implies that the eigenvectors of the population SCM equal the eigenvectors of the population covariance matrix. Moreover, the corresponding standardized eigenvalues of the population SCM are the inverses of the standardized eigenvalues of $\Sigma$.

The functional $T(F)$ in Theorem 1 was taken to be the Oja median, but it can be replaced by any affine equivariant location functional satisfying $T(F_0) = \mathbf{0}$. Also, in the next section, when considering the asymptotic behavior of the SCM, the Oja median may be replaced by any $\sqrt{n}$-convergent estimate of $\boldsymbol{\mu}$.

## 4. Influence function and the asymptotic behavior of the SCM

### 4.1. Influence function and limiting distribution

Consider a sample of iid random vectors $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ from a *symmetric* $k$-variate distribution $F$ with finite second-order moments. We say that $\boldsymbol{X}$ follows a symmetric distribution if there exists a vector $\boldsymbol{\mu}$ such that the distributions of $\boldsymbol{X} - \boldsymbol{\mu}$ and $-(\boldsymbol{X} - \boldsymbol{\mu})$ are the same. Note that the location-scale model and the elliptical models are subclasses of symmetric distributions. For symmetric $F$, the population Oja median $T(F)$ is the symmetry center $\boldsymbol{\mu}$ and we suppose without loss of generality that $\boldsymbol{\mu} = \mathbf{0}$.

We will use the shorthand notations, $\boldsymbol{E}(I) = \boldsymbol{E}(I; \mathbf{0})$ and $\boldsymbol{S}_n(\boldsymbol{x}) = \boldsymbol{S}_n(\boldsymbol{x}; \mathbf{0})$. The following lemma shows that $\text{ave}_i\{\boldsymbol{S}_n(\boldsymbol{X}_i)\boldsymbol{S}_n^T(\boldsymbol{X}_i)\}$ is asymptotically equivalent with a $U$-statistic.

**Lemma 2.** *Write* $K=(i_1,\dots,i_{2k-1})\subset\{1,\dots,n\}$ *and consider the U-statistic* $U_n=\binom{n}{2k-1}^{-1}\sum_K g(K)$ *with kernel*

$$\begin{aligned} g(K) &= g(\boldsymbol{X}_{i_1},\dots,\boldsymbol{X}_{i_{2k-1}})\\ &= \frac{(k-1)!^2}{(2k-1)!}\sum_{I\cup J\cup\{i\}=K}\operatorname{sign}\{\boldsymbol{E}^T(I)\boldsymbol{X}_i\}\operatorname{sign}\{\boldsymbol{E}^T(J)\boldsymbol{X}_i\}\boldsymbol{E}(I)\boldsymbol{E}^T(J). \end{aligned}$$

*Then* $\sqrt{n}[U_n-\mathrm{ave}_i\{\boldsymbol{S}_n(\boldsymbol{X}_i)\boldsymbol{S}_n^T(\boldsymbol{X}_i)\}]\xrightarrow{\mathrm{P}}0$.

The next lemma permits the replacement of the sample estimate $\hat{\boldsymbol{\theta}}$ by the population value $\boldsymbol{\mu}$ of the location estimator in asymptotical considerations.

**Lemma 3.** *With the notations stated above*,

$$\sqrt{n}[\mathrm{ave}_i\{\boldsymbol{S}_n(\boldsymbol{X}_i;\hat{\boldsymbol{\theta}})\boldsymbol{S}_n^T(\boldsymbol{X}_i;\hat{\boldsymbol{\theta}})\}-\mathrm{ave}_i\{\boldsymbol{S}_n(\boldsymbol{X}_i)\boldsymbol{S}_n^T(\boldsymbol{X}_i)\}]\xrightarrow{\mathrm{P}}0$$

*for any* $\sqrt{n}$*-convergent location estimate* $\hat{\boldsymbol{\theta}}$.

Before we continue with the derivation of the limiting distribution of the SCM, we will compute its influence function. The influence function (IF) of a functional $T$ at $F$ measures the effect of an infinitesimal contamination located at a single point $\boldsymbol{x}_0$. We thus consider the contaminated distribution $F_\epsilon=(1-\epsilon)F+\epsilon\Delta_{\boldsymbol{x}_0}$ where $\Delta_{\boldsymbol{x}_0}$ is a distribution putting all its mass at $\boldsymbol{x}_0$. The influence function is now defined as

$$\mathrm{IF}(\boldsymbol{x}_0;T,F)=\lim_{\epsilon\downarrow 0}\frac{T(F_\epsilon)-T(F)}{\epsilon}=\frac{\partial}{\partial\epsilon}T(F_\epsilon)\bigg|_{\epsilon=0}.$$

The IF is a tool to describe robustness properties of an estimator, but it can also be used to compute asymptotic variance. See [11].

The IF of the multivariate sign of $\boldsymbol{x}$ with respect to any $\boldsymbol{\theta}$ is simply given by

$$\begin{aligned} &\mathrm{IF}(\boldsymbol{x}_0;\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\theta}),F)\\ &\quad=(k-1)\{\mathbb{E}_F[\operatorname{sign}\{\boldsymbol{E}^T(I;\boldsymbol{\theta})(\boldsymbol{X}-\boldsymbol{\theta})\}\boldsymbol{E}(I;\boldsymbol{\theta})|\boldsymbol{X}_{i_1}=\boldsymbol{x}_0,\boldsymbol{X}=\boldsymbol{x}]\\ &\qquad-\ \boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\theta})\}, \end{aligned} \tag{4}$$

since $\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\theta})$ is a $U$-statistic with kernel of order $k-1$. The IF of the population SCM is given next.

**Theorem 2.** *For a symmetric distribution F with center* $\boldsymbol{\mu}$, *the influence function of the SCM functional D at F is given by*

$$\begin{aligned} \mathrm{IF}(\boldsymbol{x}_0;D,F)=&\ \boldsymbol{S}_F(\boldsymbol{x}_0;\boldsymbol{\mu})\boldsymbol{S}_F^T(\boldsymbol{x}_0;\boldsymbol{\mu})+\mathbb{E}_F[\mathrm{IF}(\boldsymbol{x}_0;\boldsymbol{S}_F(\boldsymbol{X};\boldsymbol{\mu}),F)\boldsymbol{S}_F^T(\boldsymbol{X};\boldsymbol{\mu})]\\ &+\ \mathbb{E}_F[\boldsymbol{S}_F(\boldsymbol{X};\boldsymbol{\mu})\mathrm{IF}(\boldsymbol{x}_0;\boldsymbol{S}_F(\boldsymbol{X};\boldsymbol{\mu}),F)^T]-D(F). \end{aligned} \tag{5}$$

The main result of this section is stated now. We use "vec" as operator working on matrices: $\text{vec}(A)$ vectorize matrix $A$ by stacking the columns of the matrix on top of each other.

**Theorem 3.** *Assume that $F$ is a $k$-variate symmetric distribution $F$ with finite second-order moments. Then $\hat{D} \xrightarrow{\text{P}} D$ and $\sqrt{n}\,\text{vec}(\hat{D} - D)$ has a limiting multinormal distribution with zero mean and asymptotic covariance matrix*

$$\text{ASV}(\hat{D}; F) = \mathbb{E}_F[\text{vec}\{\text{IF}(\boldsymbol{X}; D, F)\}\text{vec}\{\text{IF}(\boldsymbol{X}; D, F)\}^T].$$

The SCM is therefore asymptotically normal under the restriction of finite second-order moments. Note that asymptotic normality of the sample covariance matrix requires existence of the fourth-order moments.

### 4.2. Special case: the elliptical model

In case of elliptically symmetric model distribution, it is possible to render Eqs. (4) and (5) much more explicit. Special attention will be given to the important class of multivariate normal and $t$-distributions.

First consider a spherical $F_0$ with symmetry center $\mathbf{0}$ and covariance matrix $I_k$. The influence function of the sign of $\boldsymbol{x}$ with respect to $\mathbf{0}$ at $F_0$ is (cf. Lemma A.3 in the appendix)

$$\text{IF}(\boldsymbol{x}_0; \boldsymbol{S}_{F_0}(\boldsymbol{x}; \mathbf{0}), F_0) = (k-1)\left\{\delta c'_{F_0} \frac{(I_k - \boldsymbol{u}_0\boldsymbol{u}_0^T)\boldsymbol{x}}{||(I_k - \boldsymbol{u}_0\boldsymbol{u}_0^T)\boldsymbol{x}||} - \boldsymbol{S}_{F_0}(\boldsymbol{x}; \mathbf{0})\right\}, \tag{6}$$

where $\delta = ||\boldsymbol{x}_0||$ and $\boldsymbol{u}_0 = \boldsymbol{x}_0/\delta$ is the unit vector in the direction of $\boldsymbol{x}_0$. Note that the influence function of the sign vector is unbounded as it is a linear function of the length of the perturbing vector $\boldsymbol{x}_0$. The constant $c'_{F_0}$ above is defined as

$$c'_{F_0} = c_{F_0} \frac{(\frac{k-1}{2})\Gamma^2(\frac{k-1}{2})}{\mathbb{E}_{F_0}(R)\Gamma^2(\frac{k}{2})}, \tag{7}$$

where $R$ is the length of a random vector from $F_0$ and $c_{F_0}$ has been defined in (3).

Starting from (6) and Theorem 2, the next result has been proven.

**Theorem 4.** *For a spherical distribution $F_0$ with mean vector $\mathbf{0}$ and covariance matrix $I_k$, the influence function of the SCM functional $D$ at $F_0$ is given by*

$$\text{IF}(\boldsymbol{x}_0; D, F_0) = \alpha(||\boldsymbol{x}_0||)\boldsymbol{x}_0\boldsymbol{x}_0^T - \beta(||\boldsymbol{x}_0||)D(F_0),$$

*where $D(F_0) = (c_{F_0}^2/k)I_k$, and $\alpha$ and $\beta$ are two real-valued functions, depending only on $F_0$, and defined as*

$$\alpha(\delta) = c_{F_0}^2\delta^{-2}\{1 - 2\delta\mathbb{E}_{F_0}^{-1}(R)\},$$

$$\beta(\delta) = 2k - 1 - 2\delta k\mathbb{E}_{F_0}^{-1}(R),$$

*where $R$ is the length of a random vector from $F_0$.*

Note that $\delta^2\alpha(\delta)$ and $\beta(\delta)$ are linear in $\delta$ and that they depend on the dimension $k$. For $\boldsymbol{x}_0 = \delta\boldsymbol{u}_0$, the influence function may be rewritten as $c^2_{F_0}\{1 - 2\delta\mathbb{E}^{-1}_{F_0}(R)\}\boldsymbol{u}_0\boldsymbol{u}_0^T - \beta(\delta)D(F_0)$, which reveals more clearly the linearity of the influence function. The SCM procedure has therefore a "least absolute deviations" character, and also an unbounded influence function. Furthermore, the breakdown point of the SCM functional is zero. Because of affine equivariance of SCM, we may now easily derive the influence function at an elliptically symmetric distribution $F$.

**Corollary 1.** *Let $F$ be an elliptical distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ and let $F_0$ be the corresponding standardized distribution having mean vector $\mathbf{0}$ and covariance matrix $I_k$. Then*

$$\mathrm{IF}(\boldsymbol{x}_0; D, F) = \alpha(d(\boldsymbol{x}_0))\det(\Sigma)\Sigma^{-1}(\boldsymbol{x}_0 - \boldsymbol{\mu})(\boldsymbol{x}_0 - \boldsymbol{\mu})^T\Sigma^{-1} - \beta(d(\boldsymbol{x}_0))D(F)$$

*with $d^2(\boldsymbol{x}_0) = (\boldsymbol{x}_0 - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}_0 - \boldsymbol{\mu})$ the squared mahalanobis distance of $\boldsymbol{x}_0$, $D(F) = (c^2_{F_0}/k)\det(\Sigma)\Sigma^{-1}$, and where the two functions $\alpha$ and $\beta$ depend on $F_0$ and are as in Theorem* 4.

**Remark.** A referee noticed the similarity with the influence function of $D(F)$ and that of the functional $H(F) = \det\{\mathrm{Cov}(F)\}[\mathrm{Cov}(F)]^{-1}$, where

$$\mathrm{Cov}(F) = \mathbb{E}_F[(\boldsymbol{X} - \mathbb{E}_F[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}_F[\boldsymbol{X}])^T]$$

is the functional representation of the classical covariance matrix estimator. The influence function of $H(F)$ is then

$$\mathrm{IF}(\boldsymbol{x}_0; H, F) = \alpha^*(d(\boldsymbol{x}_0))\det(\Sigma)\Sigma^{-1}(\boldsymbol{x}_0 - \boldsymbol{\mu})(\boldsymbol{x}_0 - \boldsymbol{\mu})^T\Sigma^{-1} - \beta^*(d(\boldsymbol{x}_0))H(F),$$

where $\alpha^*(\delta) = -1$ and $\beta^*(\delta) = k - 1 - \delta^2$.

Using Theorems 3 and 4, it is now possible to find out expressions for the asymptotic covariance matrix of the SCM. Before that, we need to introduce some notations. A commutation matrix $I_{k,k}$, is a $k^2 \times k^2$ block matrix with $(i,j)$-block being equal to a $k \times k$ matrix that is 1 at entry $(j,i)$ and zero elsewhere. Recall that the Kronecker product of $k \times k$ matrices $A$ and $B$, denoted by $A \otimes B$, is a $k^2 \times k^2$-block matrix with $k \times k$-blocks, the $(i,j)$-block equal to $a_{ij}B$. For relations of Kronecker products, commutation matrices and vec-operator, the reader is referred to [15]. Now write $\hat{D}_{ii}$ for an on-diagonal element of the matrix $\hat{D}$ and $\hat{D}_{ij}$, with $i \neq j$, for an off-diagonal element.

**Corollary 2.** *At a spherical distribution $F_0$, the covariance matrix of the limiting distribution of $\sqrt{n}\,\mathrm{vec}(\hat{D} - D)$ is given by*

$$\mathrm{ASV}(\hat{D}_{12}; F_0)(I_{k^2} + I_{k,k}) + \mathrm{ASC}(\hat{D}_{11}, \hat{D}_{22}; F_0)\,\mathrm{vec}(I_k)\,\mathrm{vec}(I_k)^T,$$

*and at an elliptical distribution with parameters* $\boldsymbol{\mu}$ *and* $\Sigma$ *it is given by*

$$\frac{k^2}{c_{F_0}^4}[\mathrm{ASV}(\hat{D}_{12};F_0)(I_{k^2}+I_{k,k})(D\otimes D)+\mathrm{ASC}(\hat{D}_{11},\hat{D}_{22};F_0)\mathrm{vec}(D)\mathrm{vec}(D)^T],$$

*where* $\mathrm{ASC}(\hat{D}_{11},\hat{D}_{22};F_0)$ *is the asymptotic covariance between two distinct on-diagonal elements and* $\mathrm{ASV}(\hat{D}_{12};F_0)$ *is the asymptotic variance of an off-diagonal element of the SCM.*

Notice also that

$$\mathrm{ASV}(\hat{D}_{11};F_0)=2\mathrm{ASV}(\hat{D}_{12};F_0)+\mathrm{ASC}(\hat{D}_{11},\hat{D}_{22};F_0).$$

The limiting distribution of the SCM is therefore characterized by 2 numbers: the asymptotic variance of an off-diagonal element $\mathrm{ASV}(\hat{D}_{12};F_0)$ and the asymptotic variance of an on-diagonal element $\mathrm{ASV}(\hat{D}_{11};F_0)$. After some lengthy but straightforward calculations (Lemma A.1 is useful here), we obtained

$$\mathrm{ASV}(\hat{D}_{ij};F_0)=\frac{c_{F_0}^4\{4k\mathbb{E}_{F_0}^{-2}(R)-3\}}{k(k+2)},\quad i\neq j,$$

$$\mathrm{ASV}(\hat{D}_{ii};F_0)=\frac{c_{F_0}^4\{4k^2(k^2-1)\mathbb{E}_{F_0}^{-2}(R)-2+6k-4k^3\}}{k^2(k+2)}.$$

The above variances (and also the functions $\alpha$ and $\beta$ of Theorem 4) can be made explicit by calculating $\mathbb{E}_{F_0}(R)$ at the specified model distribution $F_0$. For example, for a multivariate standard normal $F_0=\Phi$ and a $k$-variate standardized (so having unit covariance matrix) spherical $t$-distribution with $v$ degrees of freedom $F_0=t_v$ we have

$$\mathbb{E}_{\Phi}(R)=\frac{\sqrt{2}\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})}\quad\text{and}\quad\mathbb{E}_{t_v}(R)=\frac{\sqrt{v-2}\Gamma(\frac{v-1}{2})\Gamma(\frac{k+1}{2})}{\Gamma(\frac{v}{2})\Gamma(\frac{k}{2})}.$$

Note that the related multivariate one sample and several samples sign test statistics [20] are conditionally and asymptotically distribution-free ('non-parametric' in that sense), but naturally (and this happens also in the univariate case), the companion location and scatter estimates are not even asymptotically distribution free; however, they are Fisher consistent in a wide class of distributions.

## 5. Principal components analysis based on the SCM

Assume that the $k$-variate cdf $F$ has a covariance matrix $\Sigma$ with distinct eigenvalues $\lambda_1>\cdots>\lambda_k>0$ and respective eigenvectors $\boldsymbol{v}_1,\dots,\boldsymbol{v}_k$, and write $\Sigma=P\Lambda P^T$ for its spectral decomposition (as defined in Section 1). Further, let $\Lambda^*$ be the diagonal matrix of standardized eigenvalues as defined in (1). Consequently, $D(F)$ has distinct eigenvalues $0<\lambda_{D,1}(F)<\cdots<\lambda_{D,k}(F)$ and we write $\boldsymbol{v}_{D,1}(F),\dots,\boldsymbol{v}_{D,k}(F)$ for the corresponding eigenvectors, and $P_D(F)\Lambda_D(F)P_D(F)^T$ for the spectral decomposition of $D(F)$. Further, let $\hat{P}_D\hat{\Lambda}_D\hat{P}_D^T$ be the spectral decomposition of $\hat{D}$, thus having the eigenvalues $\hat{\lambda}_{D,1} < \cdots < \hat{\lambda}_{D,k}$ of $\hat{D}$ as diagonal elements of $\hat{\Lambda}_D$ and the corresponding eigenvectors $\hat{\boldsymbol{v}}_{D,1}, \ldots, \hat{\boldsymbol{v}}_{D,k}$ of $\hat{D}$ as column vectors of $\hat{P}_D$. Let $\Lambda_D^*(F)$ be a diagonal matrix having as diagonal elements $\lambda_{D,1}^*(F), \ldots, \lambda_{D,k}^*(F)$, the inverses of the standardized eigenvalues of $D(F)$. We use the obvious notations $\hat{\lambda}_{D,j}^*$, $j = 1, \ldots, k$ and $\hat{\Lambda}_D^*$ for corresponding elements obtained from $\hat{D}$. Theorem 1 yields

$$P_D(F) = P, \quad \Lambda_D(F) = w_F \Lambda^{-1} \quad \text{and} \quad \Lambda_D^*(F) = \Lambda^*$$

for $F$ belonging to a location-scale model. This means that the orientation of the SCM matrix is the same as for the covariance matrix, whereas the inverses of the eigenvalues of $\hat{D}$ allow to measure the shape of $\Sigma$.

Next, we derive the influence functions for eigenvector and eigenvalue functionals at an elliptical model.

**Theorem 5.** *Let $F$ be elliptical distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ having distinct eigenvalues, and write $F_0$ for the corresponding standardized distribution. The influence functions of the eigenvectors and eigenvalues of $D$ at $F$ are then given by*

$$\mathrm{IF}(\boldsymbol{x}_0; \boldsymbol{v}_{D,j}, F) = \tilde{\alpha}(d(\boldsymbol{x}_0)) \sum_{\substack{i=1 \\ i \neq j}}^{k} \frac{z_i z_j}{\lambda_j - \lambda_i} \boldsymbol{v}_i,$$

$$\mathrm{IF}(\boldsymbol{x}_0; \lambda_{D,j}, F) = \alpha(d(\boldsymbol{x}_0)) \det(\Sigma)(z_j/\lambda_j)^2 - \beta(d(\boldsymbol{x}_0)) \det(\Sigma)(c_{F_0}^2/k)\lambda_j^{-1},$$

*where $z_j = \boldsymbol{v}_j^T(\boldsymbol{x}_0 - \boldsymbol{\mu})$ for $j = 1, \ldots, k$, $d^2(\boldsymbol{x}_0) = (\boldsymbol{x}_0 - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}_0 - \boldsymbol{\mu})$ and $\tilde{\alpha}(\delta) = -(k/c_{F_0}^2)\alpha(\delta)$.*

As in [7], we can rewrite the influence function for the eigenvectors of the SCM in the form

$$\mathrm{IF}(\boldsymbol{x}_0; \boldsymbol{v}_{D,j}, F) = \tilde{\alpha}(d(\boldsymbol{x}_0)) \mathrm{IF}(\boldsymbol{x}_0; \boldsymbol{v}_{\mathrm{Cov},j}, F),$$

where $\mathrm{IF}(\boldsymbol{x}_0; \boldsymbol{v}_{\mathrm{Cov},j}, F)$ is the influence function of the eigenvector functional of $\mathrm{Cov}(F)$ obtained by Critchley [4]. The function $\delta \to \tilde{\alpha}(\delta)$ is telling us how much more or less weight an observation receives when computing eigenvectors from the SCM instead of from the sample covariance matrix. It is instructive to have a look at the form of this function, pictured in Fig. 2. We also compared with the $\tilde{\alpha}(\delta)$ function of a high breakdown estimator: the multivariate Biweight $S$-estimator [9,24], which has already been considered by Croux and Haesbroeck [7]. Note that the $\tilde{\alpha}(\delta)$ for the classical estimator is constant and equal to one. From Fig. 2 we see that observations far away from the origin, so for $\delta$ large, receive much less weight using the SCM instead of the classical estimator. For the high breakdown estimators the downweighting of outliers is much stronger, which renders these estimators more robust, but they will also be less efficient. Note that observations very close to the center have a relatively large effect on the SCM. This *inlier-effect* is also observed for the spatial median, and has been discussed by Brown et al. [3]. They observed that

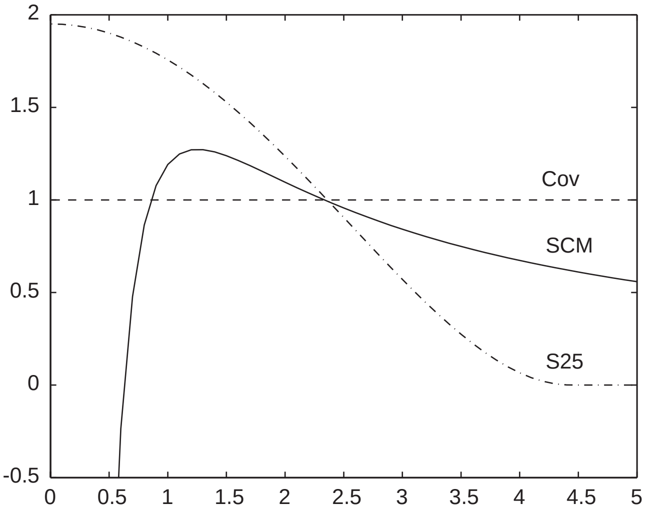


Fig. 2. The function $\tilde{\alpha}(\delta)$ for the SCM estimator, the classical covariance matrix estimator and the 25 percent breakdown $S$-estimator at the bivariate normal model ($F = \Phi$ and $k = 2$).

the inlier effect becomes smaller and smaller with increasing $k$. The influence function for the SCM remains bounded in the neighborhood of the origin, that is, while $|\tilde{\alpha}(d(\boldsymbol{x}_0))| \to \infty$ as $d(\boldsymbol{x}_0) \to 0$, $||\mathrm{IF}(\boldsymbol{x}_0; \boldsymbol{v}_{D,j}, F)||$ remains bounded.

In the paper, we set the sign of the eigenvectors in such a way that the first element of each vector is positive. This is needed to obtain uniquely defined eigenvectors. The following theorem shows that the estimators $\hat{\lambda}_{D,j}$ and $\hat{\boldsymbol{v}}_{D,j}$ have regular asymptotic behavior.

**Theorem 6.** *Let the distribution $F$ belong to a location-scale model with covariance matrix $\Sigma$ having distinct eigenvalues. Then $\hat{P}_D \xrightarrow{\mathrm{P}} P$ and $\sqrt{n}\,\mathrm{vec}(\hat{P}_D - P)$ has a limiting normal distribution with zero mean. Furthermore, $\hat{\Lambda}_D \xrightarrow{\mathrm{P}} \Lambda_D$ and $\sqrt{n}\,\mathrm{vec}(\hat{\Lambda}_D - \Lambda_D)$ has a limiting normal distribution with zero mean.*

For elliptical distributions, we can be more rigorous than in Theorem 6, and use

$$\mathrm{ASV}(\hat{P}_D; F) = \mathbb{E}_F[\mathrm{vec}\{\mathrm{IF}(\boldsymbol{X}; P_D, F)\}\mathrm{vec}\{\mathrm{IF}(\boldsymbol{X}; P_D, F)\}^T],$$

$$\mathrm{ASV}(\hat{\Lambda}_D; F) = \mathbb{E}_F[\mathrm{vec}\{\mathrm{IF}(\boldsymbol{X}; \Lambda_D, F)\}\mathrm{vec}\{\mathrm{IF}(\boldsymbol{X}; \Lambda_D, F)\}^T]$$

to calculate the asymptotic covariance matrices. See also [7,8] for recent discussions on asymptotic distributions of the estimates of eigenvectors and eigenvalues.

**Corollary 3.** *Let $F$ be an elliptical distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\Sigma$ having distinct eigenvalues, and write $F_0$ for the corresponding standardized* *distribution. Then,* $\sqrt{n}\,\mathrm{vec}(\hat{P}_D - P)$ *and* $\sqrt{n}\,\mathrm{vec}(\hat{\Lambda}_D - \Lambda_D)$ *has a limiting normal distribution with zero mean, and* $\hat{P}_D$ *and* $\hat{\Lambda}_D$ *are asymptotically independent. The covariance matrix of* $\hat{\boldsymbol{v}}_{D,j}$ *and the covariance matrix of* $\hat{\boldsymbol{v}}_{D,i}$ *and* $\hat{\boldsymbol{v}}_{D,j}$, $i \neq j$, *in the limiting distribution are given by*

$$\mathrm{ASV}(\hat{\boldsymbol{v}}_{D,j};F) = \frac{k^2}{c_{F_0}^4}\mathrm{ASV}(\hat{D}_{12};F_0)\sum_{\substack{i=1\\ i\neq j}}^{k}\frac{\lambda_i\lambda_j}{(\lambda_j-\lambda_i)^2}\boldsymbol{v}_i\boldsymbol{v}_i^T,$$

$$\mathrm{ASC}(\hat{\boldsymbol{v}}_{D,i},\hat{\boldsymbol{v}}_{D,j};F) = \frac{k^2}{c_{F_0}^4}\mathrm{ASV}(\hat{D}_{12};F_0)\frac{-\lambda_i\lambda_j}{(\lambda_j-\lambda_i)^2}\boldsymbol{v}_j\boldsymbol{v}_i^T,$$

*correspondingly. The variance of* $\hat{\lambda}_{D,j}$ *and the covariance of* $\hat{\lambda}_{D,i}$ *and* $\hat{\lambda}_{D,j}$, $i \neq j$, *in the limiting distribution are given by*

$$\mathrm{ASV}(\hat{\lambda}_{D,j};F) = \frac{\det(\Sigma)^2}{\lambda_j^2}\mathrm{ASV}(\hat{D}_{11};F_0),$$

$$\mathrm{ASC}(\hat{\lambda}_{D,i},\hat{\lambda}_{D,j};F) = \frac{\det(\Sigma)^2}{\lambda_i\lambda_j}\mathrm{ASC}(\hat{D}_{11},\hat{D}_{22};F_0)$$

*correspondingly.*

The asymptotic covariance matrix for the eigenvector estimates based on the sample covariance matrix $\widehat{\mathrm{Cov}}$ is given by

$$\mathrm{ASV}(\hat{\boldsymbol{v}}_{\mathrm{Cov},j};F) = \mathrm{ASV}(\widehat{\mathrm{Cov}}_{12};F_0)\sum_{\substack{i=1\\ i\neq j}}^{k}\frac{\lambda_i\lambda_j}{(\lambda_j-\lambda_i)^2}\boldsymbol{v}_i\boldsymbol{v}_i^T,$$

where $\widehat{\mathrm{Cov}}_{12}$ (by symmetry) can be regarded as any off-diagonal element of the sample covariance matrix $\widehat{\mathrm{Cov}}$ (e.g. [4]). This means that the asymptotic relative efficiency (ARE) of the estimates $\hat{\boldsymbol{v}}_{D,j}$ based on the SCM with respect to the estimates $\hat{\boldsymbol{v}}_{\mathrm{Cov},j}$ based on the sample covariance matrix at an elliptical distribution $F$ is given by

$$\mathrm{ARE}(\hat{\boldsymbol{v}}_{\mathrm{Cov},j},\hat{\boldsymbol{v}}_{D,j};F) = \frac{\mathrm{ASV}(\widehat{\mathrm{Cov}}_{12};F_0)}{(k^2/c_{F_0}^4)\mathrm{ASV}(\hat{D}_{12};F_0)}.$$

For example, at the standardized $t$-distribution $(F_0 = t_\nu)$, $\mathrm{ASV}(\widehat{\mathrm{Cov}}_{12};t_\nu) = (\nu-2)/(\nu-4)$ for $\nu > 4$ and hence the AREs are readily calculable using the formulas of Section 4.2. Table 1A lists the AREs calculated for multivariate $t$-distributions for several dimensions and degrees of freedom. Efficiencies for multinormal distributions, which correspond to the limiting case of the degrees of freedom ($\nu = \infty$), are also given. The efficiencies are very high in the normal case, and they get larger with increasing dimension. At the multivariate $t$-distributions, the estimates based on SCM outperform the classical estimators, especially at the heavier tailed distributions. Table 1B lists the same asymptotic efficiencies, but now relative

Table 1
AREs of the SCM eigenvector estimates with respect to those based on the sample covariance matrix at $t$-distribution for several values of the dimension $k$ and degrees of freedom $v$. (B) lists the AREs with respect to the MLE

| $k$ | $v=5$ | $v=6$ | $v=8$ | $v=15$ | $v=\infty$ |
|---|---|---|---|---|---|
| (A) | | | | | |
| 2 | 2.000 | 1.447 | 1.184 | 1.031 | 0.956 |
| 3 | 1.960 | 1.429 | 1.179 | 1.038 | 0.973 |
| 5 | 1.905 | 1.400 | 1.167 | 1.040 | 0.987 |
| 10 | 1.843 | 1.365 | 1.148 | 1.036 | 0.996 |
| 15 | 1.816 | 1.349 | 1.139 | 1.032 | 0.998 |
| $\infty$ | 1.752 | 1.310 | 1.114 | 1.022 | 1.000 |
| (B) | | | | | |
| 2 | 0.857 | 0.904 | 0.947 | 0.975 | 0.956 |
| 3 | 0.816 | 0.873 | 0.929 | 0.976 | 0.973 |
| 5 | 0.762 | 0.827 | 0.897 | 0.968 | 0.987 |
| 10 | 0.696 | 0.768 | 0.850 | 0.946 | 0.996 |
| 15 | 0.666 | 0.739 | 0.825 | 0.932 | 0.998 |
| $\infty$ | 0.584 | 0.655 | 0.743 | 0.865 | 1.000 |

to maximum likelihood estimates (MLE) for the respective multivariate $t$-distributions, the latter being the most efficient estimates at the model distribution. Recall that the sample covariance matrix is the MLE at the normal model. We see that also these efficiencies remain fairly high. Only when the number of degrees of freedom becomes too low, there is a serious loss in efficiency w.r.t. the MLE.

The asymptotic behavior of the standardized eigenvalues will be studied in the next section, where we will show that their relative asymptotic efficiencies are exactly the same as those of the eigenvector estimates.

## 6. Estimating covariance and correlation

The SCM allows to estimate shape and orientation of the underlying covariance matrix, but it is also possible to construct an affine equivariant estimator for $\Sigma$ based on the SCM. (Maronna and Yohai [16] give an overview of existing estimators of multivariate scatter.) Suppose that $F$ belongs to a location-scale family generated by $F_0$, where $F_0$ has been specified and is therefore supposed to be known. Define now

$$C(F) = \left[\frac{\det\{D(F)\}}{w_{F_0}}\right]^{1/(k-1)} D(F)^{-1}$$

and write $C(\boldsymbol{X}) = C(G)$ whenever $\boldsymbol{X} \sim G$. Using the equivariance properties of the SCM (see Lemma 1), it follows that $C$ is affine equivariant in the regular sense: $C(A\boldsymbol{X}+\boldsymbol{b}) = AC(\boldsymbol{X})A^T$ with $A$ any regular non-singular $k \times k$ matrix and for any

$k$-vector $\boldsymbol{b}$. Moreover, by Theorem 1, one then has that $C(F) = \Sigma$ meaning that $C$ is a Fisher consistent functional for $\Sigma$ at a location-scale model. Particularly at elliptical models, we set $w_{F_0} = c^2_{F_0}/k$ and obtain an affine equivariant scatter matrix estimator for $\Sigma$. For example, at the normal model we get as sample estimate

$$\hat{C} = \left[\frac{\det(\hat{D})}{c^2_{\Phi}/k}\right]^{1/(k-1)} \hat{D}^{-1}.$$

Note that also other affine equivariant scatter matrix estimators, including the *minimum covariance determinant* (MCD)-estimator of Rousseeuw [23] and multivariate $S$-estimators, need a scaling factor to attain consistency for $\Sigma$ at the model distribution. Without such a scaling factor they only estimate orientation and shape, but not the size of the scatter matrix.

An expression for the IF of $C$ at elliptical distributions easily follows from Corollary 1, after using some matrix differentiation rules for the determinant and the inverse of a non-singular matrix (see e.g. [15, Chapter 8, Theorem 1 and 3]):

$$\mathrm{IF}(\boldsymbol{x}_0; C, F) = \tilde{\alpha}(d(\boldsymbol{x}_0))(\boldsymbol{x}_0 - \boldsymbol{\mu})(\boldsymbol{x}_0 - \boldsymbol{\mu})^T - \tilde{\beta}(d(\boldsymbol{x}_0))\Sigma, \tag{8}$$

where $\tilde{\alpha}(\delta) = -(k/c^2_{F_0})\alpha(\delta)$ was already defined in Section 5 and $\tilde{\beta}(\delta) \equiv 1$. Note that for the classical estimator $\mathrm{Cov}(F)$, $\mathrm{IF}(\boldsymbol{x}_0; \mathrm{Cov}, F) = (\boldsymbol{x}_0 - \boldsymbol{\mu})(\boldsymbol{x}_0 - \boldsymbol{\mu})^T - \Sigma$, as was shown in [4]. In Fig. 3 we picture $\mathrm{IF}(\boldsymbol{x}_0; C_{12}, F_0)$ for a typical off-diagonal and on-diagonal element of $C$ with $F_0 = \Phi$. We compared with the influence functions for the classical estimator. From the figures we see that the influence functions are smooth, but unbounded. But the increase in influence when an observation tends away from the center of the distribution is much slower for the SCM-based covariance matrix estimator than for the classical procedure. Notice that the inlier-effect is visible in the figures for SCM.

Similar pictures have been depicted by Croux and Haesbroeck [6], who also computed asymptotic efficiencies for several estimators of the off- and on-diagonal elements of $\Sigma$. For the off-diagonal elements, there is no work to do, since one readily can check that

$$\mathrm{ARE}(\hat{\mathrm{Cov}}_{12}, \hat{C}_{12}; F_0) = \frac{\mathrm{ASV}(\hat{\mathrm{Cov}}_{12}; F_0)}{(k^2/c^4_{F_0})\mathrm{ASV}(\hat{D}_{12}; F_0)} = \mathrm{ARE}(\hat{\boldsymbol{v}}_{\mathrm{Cov},j}, \hat{\boldsymbol{v}}_{D,j}; F)$$

corresponding to the numbers in Table 1. For the on-diagonal elements there are some extra computations to be done. AREs for multivariate $t$ and normal distributions are given in Table 2. Again we see that at the normal model, the efficiencies are very high. At $t$-distributions the SCM-based estimators outperform the classical estimators. We observe that the relative efficiencies for the on-diagonal elements are in general higher than for the estimates of the off-diagonal elements when comparing to Cov, but the reverse is true when we compare to the MLE.

The influence function of any affine equivariant scatter matrix estimator can be written in the form (8), but of course with different $\tilde{\alpha}$ and $\tilde{\beta}$ (cf. Lemma 1 of Croux and Haesbroeck [7]). Obtaining the $\tilde{\alpha}$ and $\tilde{\beta}$ functions for the affine equivariant scatter matrix estimator $C$ is also useful for further applications. For example, Croux

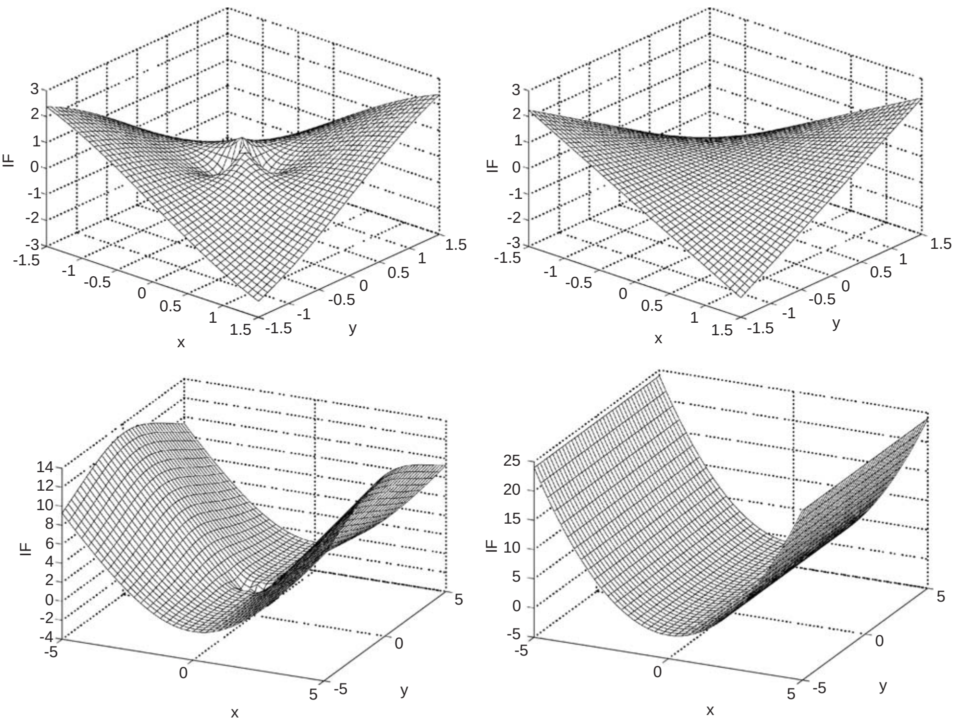


Fig. 3. Influence functions for an off-diagonal element (first row) and on-diagonal element (second row) of the SCM-estimator $C$ (first column) and the classical covariance estimator Cov (second column) at the normal model.

Table 2
(A) Lists the ARE of the on-diagonal element of SCM-estimate $\hat{C}$ with respect to on-diagonal element of the sample covariance matrix $\hat{\mathrm{Cov}}$ at $t$-distribution with selected values of dimension $k$ and degrees of freedom $v$. (B) lists the corresponding efficiencies relative to MLE

| $k$ | $v=5$ | $v=6$ | $v=8$ | $v=15$ | $v=\infty$ |
|---|---|---|---|---|---|
| (A) | | | | | |
| 2 | 2.286 | 1.589 | 1.250 | 1.044 | 0.935 |
| 3 | 2.227 | 1.562 | 1.243 | 1.054 | 0.960 |
| 5 | 2.148 | 1.522 | 1.225 | 1.057 | 0.981 |
| 10 | 2.060 | 1.472 | 1.198 | 1.050 | 0.994 |
| 15 | 2.023 | 1.450 | 1.185 | 1.046 | 0.997 |
| $\infty$ | 1.934 | 1.396 | 1.152 | 1.031 | 1.000 |
| (B) | | | | | |
| 2 | 0.857 | 0.908 | 0.952 | 0.974 | 0.935 |
| 3 | 0.795 | 0.859 | 0.923 | 0.974 | 0.960 |
| 5 | 0.716 | 0.791 | 0.875 | 0.961 | 0.981 |
| 10 | 0.625 | 0.707 | 0.805 | 0.928 | 0.994 |
| 15 | 0.585 | 0.667 | 0.770 | 0.907 | 0.997 |
| $\infty$ | 0.483 | 0.558 | 0.658 | 0.810 | 1.000 |

and Dehon [5] obtained results for robust discriminant analysis based on any affine equivariant scatter matrix estimators. Knowledge of $\tilde{\alpha}$ and $\tilde{\beta}$ allows for immediate application of their results.

From $C(F)$ we can in the usual way obtain an estimator $R(F)$ of the population correlation matrix. We write $\hat{R}$ for the corresponding estimate obtained from $\hat{C}$. Note that $\hat{R}$ can be computed directly from the SCM, since

$$\hat{R}_{ij} = \frac{[\hat{C}]_{ij}}{\sqrt{[\hat{C}]_{ii}[\hat{C}]_{jj}}} = \frac{[\hat{D}^{-1}]_{ij}}{\sqrt{[\hat{D}^{-1}]_{ii}[\hat{D}^{-1}]_{jj}}}$$

for $1 \leqslant i,j \leqslant k$. Since $C$ is an affine equivariant scatter matrix estimator, the influence function of $R$ follows immediately from Lemma 2 of Croux and Haesbroeck [7]:

$$\mathrm{IF}(\boldsymbol{x}_0; R, F) = \tilde{\alpha}(d(\boldsymbol{x}_0))\mathrm{IF}(\boldsymbol{x}_0; \mathrm{Corr}, F),$$

where $\mathrm{IF}(\boldsymbol{x}_0; \mathrm{Corr}, F)$ is the influence function of the classical correlation matrix. Furthermore, the influence functions for the eigenvectors and eigenvalues based on the correlation matrix $R$ follow directly from Croux and Haesbroeck [7, Theorem 2]. ARE of the estimates of correlation matrix at an elliptical distributions $F$ are therefore, as in Section 5 for the eigenvector estimates, given by $\mathrm{ARE}(\hat{\mathrm{Corr}}_{12}, \hat{R}_{12}; F) = \mathrm{ARE}(\hat{\mathrm{Cov}}_{12}, \hat{C}_{12}; F_0)$. The correlation depends both on the orientation and on the shape of the matrix $\Sigma$, but their ARE only depend on one number. Note also that when a PCA is based on the correlation matrix $\hat{R}$, both the eigenvalues and eigenvectors of $\hat{R}$ are estimating the same quantities (at location-scale model) as when using the classical correlation matrix estimate.

Let us now study the asymptotic behavior of the standardized eigenvalues of $\hat{C}$, which are the same as $\hat{\lambda}^*_{D,j}$ $(j = 1, \ldots, k)$, the inverses of the standardized eigenvalues of the SCM $\hat{D}$. Herefore we will use the following lemma, valid for any regular affine equivariant estimator of scatter.

**Theorem 7.** *Let* $\hat{\boldsymbol{\lambda}}_C = (\hat{\lambda}_{C,1}, \ldots, \hat{\lambda}_{C,k})^T$ *be the eigenvalue estimates of* any *affine equivariant scatter matrix estimate* $\hat{C}$ *possessing an influence function. Assume that* $\hat{\boldsymbol{\lambda}}_C$ *is consistent with a limiting multivariate normal distribution and asymptotic covariance matrix* $\mathbb{E}_F[\mathrm{IF}(\boldsymbol{X}; \boldsymbol{\lambda}_C, F)\mathrm{IF}(\boldsymbol{X}; \boldsymbol{\lambda}_C, F)^T]$. *Let F be elliptical distribution with parameters* $\boldsymbol{\mu}$ *and* $\Sigma$ *and let* $F_0$ *be the corresponding standardized distribution. Then*

$$\sqrt{n}(\ln \hat{\lambda}^*_{C,j} - \ln \lambda^*_j) \xrightarrow{\mathrm{d}} N(0, \mathrm{ASV}(\ln \hat{\lambda}^*_{C,j}; F))$$

*with*

$$\mathrm{ASV}(\ln \hat{\lambda}^*_{C,j}; F) = \{2(k-1)/k\}\mathrm{ASV}(\hat{C}_{12}; F_0)$$

*for* $j = 1, \ldots, k$.

Particularly, we get for the SCM,

$$\mathrm{ASV}(\ln \hat{\lambda}^*_{D,j}; F) = \frac{2(k-1)}{k} \mathrm{ASV}(\hat{C}_{12}; F_0) = \frac{2k(k-1)}{c^4_{F_0}} \mathrm{ASV}(\hat{D}_{12}; F_0). \tag{9}$$

Next, write $\hat{\lambda}^*_{\mathrm{Cov},j}, j = 1, \ldots, k$ for the standardized eigenvalue estimates based on the sample covariance matrix $\hat{\mathrm{Cov}}$. The ARE of the standardized eigenvalue estimates $\hat{\lambda}^*_{D,j}$ w.r.t $\hat{\lambda}^*_{\mathrm{Cov},j}$ for elliptical $F$ is again given by

$$\mathrm{ARE}(\hat{\lambda}^*_{\mathrm{Cov},j}, \hat{\lambda}^*_{D,j}; F) = \frac{\mathrm{ASV}(\ln \hat{\lambda}^*_{\mathrm{Cov},j}; F)}{\mathrm{ASV}(\ln \hat{\lambda}^*_{D,j}; F)} = \frac{\mathrm{ASV}(\hat{\mathrm{Cov}}_{12}; F_0)}{\mathrm{ASV}(\hat{C}_{12}; F_0)}. \tag{10}$$

(See Table 1 for the efficiency calculations.)

We compared efficiencies (10) with those obtained for the MCD estimate and those for the Biweight $S$-estimate (both with 25% and 50% breakdown point) at the normal model. We refer to Croux and Haesbroeck [6] and Lopuhaä [14] for asymptotic properties of the scatter MCD and $S$-estimators. In Fig. 4 we pictured the efficiency of the estimates of the standardized eigenvalues of $\Sigma$ as a function of the dimension $k$. We see that the SCM is clearly the most efficient. The $S$-estimator with 25% breakdown point is a competitor, but the other estimators seem to result in a too high loss of efficiency. The main advantage of SCM is in its high efficiency at the normal and $t$-distributions. If these model distributions are appropriate, then also using the SCM is sensible. Nevertheless, as the SCM is not robust, the use of high breakdown estimators, like the MCD or the $S$-estimator is more useful, if the presence of large amount of outliers are suspected.

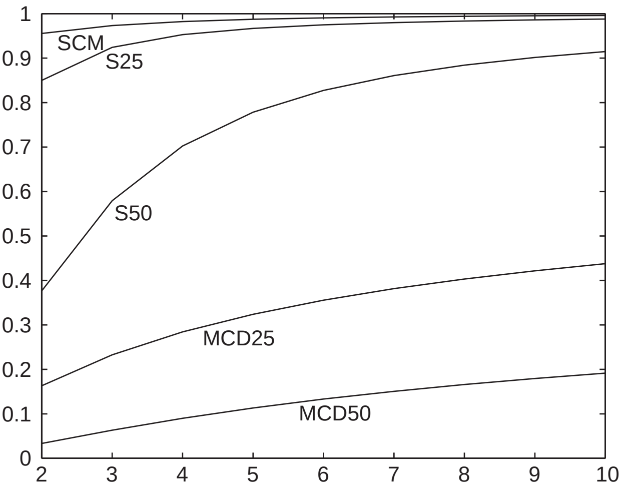


Fig. 4. Efficiencies of the standardized eigenvalues as a function of the dimension at the normal model for the SCM estimator and 25/50 percent breakdown MCD estimator and biweight $S$-estimator.

## 7. Finite sample efficiency

In the preceding sections asymptotic efficiencies were obtained for the SCM eigenvector and standardized eigenvalue estimates relative to corresponding estimates based on the sample covariance matrix. In this section, finite-sample efficiencies are obtained by means of a modest simulation study.

For $m = 1000$ samples of sizes $n = 20, 50, 100, 300$ observations were generated from a $k$-variate elliptical $t$-distribution with $v$ degrees of freedom and covariance matrix $\Sigma = \mathrm{diag}(1, \ldots, k)$. Our choices are $k = 2, 3$ and $v = 5, 6, 8, 15, \infty$, where $v = \infty$ corresponds to multinormal samples. The estimated quantities were the direction of the first eigenvector and the logarithm of the first standardized eigenvalue. The error in direction is here $\arccos\{|\boldsymbol{v}_1^T \hat{\boldsymbol{v}}_1|\}$ where $\hat{\boldsymbol{v}}_1$ is the estimated first eigenvector and $\boldsymbol{v}_1 = (0, \ldots, 0, 1)^T$ is the value to be estimated. The mean squared error (MSE) for the estimator of the first eigenvector is then

$$\mathrm{MSE}(\hat{\boldsymbol{v}}_I) = \frac{1}{m} \sum_{j=1}^{m} (\arccos\{|\boldsymbol{v}_1^T \hat{\boldsymbol{v}}_1^{(j)}|\})^2,$$

where $\hat{\boldsymbol{v}}_1^{(j)}$ is the estimate for the first eigenvector computed from the $j$th generated sample. The errors in shape will be measured as the deviation of the logarithm of the estimated standardized eigenvalue from the logarithm of the 'true' first standardized eigenvalue $\lambda_1^* = k/(k!^{1/k})$, yielding as MSE

$$\mathrm{MSE}(\log \hat{\lambda}^*) = \frac{1}{m} \sum_{j=1}^{m} (\log(\hat{\lambda}_1^*)^{(j)} - \log \lambda_1^*)^2,$$

where $(\hat{\lambda}_1^*)^{(j)}$ is the estimate for the first standardized eigenvalue computed from the $j$th generated sample. The estimated efficiencies are now computed as the ratios of the simulated mean squared errors of the SCM based procedure with respect to the sample covariance matrix based procedure. They are reported in Table 3.

For dimensions $k = 2$ and 3, the simulations show that, for $n \geqslant 300$, the asymptotic efficiencies approximate well the finite sample efficiencies. We do not claim that this finding will be true for all dimensions, and extensive simulation studies should be needed to find the values of the sample size (probably depending on the dimension) such that the normal approximation becomes reasonable. Note that somewhat slower convergence is seen at $v = 5$ showing quite serious loss of efficiency for very small samples (cases $n = 20$ and 50). This may be due to the fact that for $v = 5$ the sample covariance matrix is performing better than what the large-sample efficiency indicates (notice also that $v = 5$ is the smallest value of degrees of freedom of the $t$-distribution for which the sample covariance matrix is asymptotically normal).

Finally, we discuss the computation of the SCM. In the above simulation study, for $k = 3$ with $n = 100$ and 300, the data were centered using the spatial median estimate due to high computational cost for the Oja median at large samples. As already mentioned, replacing the Oja median by another $\sqrt{n}$-consistent estimate, being easier to compute, does not affect the asymptotics. A recent algorithm by

Table 3
Finite sample efficiencies of the SCM eigenvector and standardized eigenvalue estimates (reported between parentheses) relative to eigenvector and standardized eigenvalue estimates based on the sample covariance matrix. Samples were generated from a $k$-variate $t$-distribution with $\nu$ degrees of freedom and $\Sigma = \text{diag}(1, \ldots, k)$

| | $\nu = 5$ | $\nu = 6$ | $\nu = 8$ | $\nu = 15$ | $\nu = \infty$ |
|---|---|---|---|---|---|
| $k = 2$ | | | | | |
| $n = 20$ | 1.034 | 1.015 | 1.032 | 1.002 | 0.945 |
| | (1.154) | (1.104) | (1.038) | (1.012) | (0.942) |
| $n = 50$ | 1.180 | 1.196 | 1.124 | 1.076 | 0.922 |
| | (1.274) | (1.149) | (1.127) | (1.025) | (0.974) |
| $n = 100$ | 1.479 | 1.327 | 1.167 | 1.025 | 0.948 |
| | (1.357) | (1.209) | (1.143) | (1.039) | (0.982) |
| $n = 300$ | 1.866 | 1.413 | 1.210 | 1.026 | 0.953 |
| | (1.570) | (1.293) | (1.180) | (1.037) | (0.939) |
| $n = \infty$ | 2.000 | 1.447 | 1.184 | 1.031 | 0.956 |
| $k = 3$ | | | | | |
| $n = 20$ | 1.045 | 1.028 | 1.003 | 0.999 | 0.983 |
| | (1.191) | (1.111) | (1.070) | (1.013) | (0.951) |
| $n = 50$ | 1.201 | 1.164 | 1.056 | 1.022 | 0.981 |
| | (1.355) | (1.216) | (1.111) | (0.997) | (0.967) |
| $n = 100$ | 1.307 | 1.261 | 1.154 | 1.016 | 0.964 |
| | (1.391) | (1.239) | (1.114) | (1.020) | (0.956) |
| $n = 300$ | 1.777 | 1.409 | 1.168 | 1.052 | 0.972 |
| | (1.402) | (1.350) | (1.132) | (1.026) | (0.979) |
| $n = \infty$ | 1.960 | 1.429 | 1.179 | 1.038 | 0.973 |

Ronkainen et al. [22] that calculates an approximate solution (estimate of an estimate) for the Oja median could have been used for centering the data as well. However, for small values of $n$ and $k$ one can easily compute the Oja median using linear programming algorithm [18]. Note that, given the Oja median, the computation of a multivariate sign is explicit but requires enumeration of $O(n^{k-1})$ hyperplanes, and we need to compute $n$ multivariate signs. Therefore, it may often be too computing intensive to consider all these hyperplanes for higher values of $k$. Again, a faster approximate version of the SCM using a sample of hyperplanes is easily constructed [21, Appendix A.1].

## Appendix A. Proofs and additional lemmas

**Lemma A.1.** *For a random vector $\boldsymbol{U} = (U_1, \ldots, U_k)^T$ uniformly distributed on the unit sphere, one has that*

$$\text{(a)}\ \mathbb{E}[\boldsymbol{U}\boldsymbol{U}^T] = (1/k)\boldsymbol{I}_k,\ \ \text{(b)}\ \mathbb{E}[U_i^2 U_j^2] = \{k(k+2)\}^{-1},$$

$$\text{(c)}\ \mathbb{E}[U_i^4] = 3\{k(k+2)\}^{-1},\ \ \text{(d)}\ \mathbb{E}[(1-U_i^2)^{1/2}] = \frac{\Gamma^2(\frac{k}{2})}{\Gamma(\frac{k-1}{2})\Gamma(\frac{k+1}{2})},$$

(e) $\mathbb{E}\left[\frac{U_i^2}{(1-U_j^2)^{1/2}}\right] = \frac{\Gamma^2(\frac{k}{2})}{2\Gamma^2(\frac{k+1}{2})}$,

*where $U_i$ and $U_j$ are distinct elements of $\boldsymbol{U}$. Moreover,*

(f) $\mathbb{E}[|\det(\boldsymbol{U}_1 \cdots \boldsymbol{U}_k)|] = \mathbb{E}[|\det(\boldsymbol{U}_1 \cdots \boldsymbol{U}_k)||\boldsymbol{U}_1] = \frac{\Gamma^k(\frac{k}{2})}{\sqrt{\pi}\Gamma^{k-1}(\frac{k+1}{2})}$,

*where $\boldsymbol{U}_1, \dots, \boldsymbol{U}_k$ are random vectors uniformly distributed on the unit sphere.*

**Proof.** Here we only prove item (f), items (a)–(e) are fairly straightforward and left as exercise for the reader. Now let $R_i^2$, $i = 1, \dots, k$ be independent random variables from a $\chi_k^2$ distribution. Consequently, $\boldsymbol{X}_i = R_i \boldsymbol{U}_i$, $i = 1, \dots, k$ are independent observations from the $k$-variate standard normal distribution. Then

$$|\det(\boldsymbol{X}_1 \cdots \boldsymbol{X}_k)| = |\det(\boldsymbol{U}_1 \cdots \boldsymbol{U}_k)| \prod_{i=1}^{k} R_i \sim \prod_{i=1}^{k} \chi_i$$

with independent chi-square variables $\chi_1^2, \dots, \chi_k^2$ (cf. Lemma 1 in [17]). Thus,

$$\prod_{i=1}^{k} \mathbb{E}[\chi_i] = \mathbb{E}^k[\chi_k]\mathbb{E}[|\det(\boldsymbol{U}_1 \cdots \boldsymbol{U}_k)|]$$

which, by using $\mathbb{E}[\chi_j] = \Gamma(\frac{j+1}{2})\Gamma^{-1}(\frac{j}{2})\sqrt{2}$, gives the result. □

**Lemma A.2.** *At a spherical distribution $F_0$,*

$$\boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0}) = c_{F_0}\boldsymbol{u} \quad \textit{with} \quad c_{F_0} = \frac{\Gamma^k(\frac{k}{2})\mathbb{E}_{F_0}^{k-1}(R)}{\sqrt{\pi}\Gamma^{k-1}(\frac{k+1}{2})}.$$

**Proof.** Let $\boldsymbol{X}_i = R_i \boldsymbol{U}_i$, $i = 1, \dots, k-1$ be independent observations from $F_0$ and write $\boldsymbol{x} = r\boldsymbol{u}$, where $r = ||\boldsymbol{x}||$. Then with the aid of Lemma A.1 (f),

$$\mathbb{E}_{F_0}[|\det(\boldsymbol{X}_1 \cdots \boldsymbol{X}_{k-1}\boldsymbol{x})|] = r\mathbb{E}_{F_0}^{k-1}(R)\mathbb{E}[|\det(\boldsymbol{U}_1 \cdots \boldsymbol{U}_{k-1}\boldsymbol{u})|] = r\frac{\Gamma^k(\frac{k}{2})\mathbb{E}_{F_0}^{k-1}(R)}{\sqrt{\pi}\Gamma^{k-1}(\frac{k+1}{2})} = rc_{F_0},$$

so that $\boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0}) = \nabla_{\boldsymbol{x}}\mathbb{E}_{F_0}[|\det(\boldsymbol{X}_1 \cdots \boldsymbol{X}_{k-1}\boldsymbol{x})|] = c_{F_0}\boldsymbol{u}$.

**Proof of Lemma 1.** First, note that the Oja median is affine equivariant: $\hat{\boldsymbol{\theta}}^* = A\hat{\boldsymbol{\theta}} + \boldsymbol{b}$. Since

$$\begin{aligned}\boldsymbol{e}^*(I; \hat{\boldsymbol{\theta}}^*)^T\boldsymbol{x} &= \det(\boldsymbol{x}_{i_1}^* - \hat{\boldsymbol{\theta}}^* \quad \cdots \quad \boldsymbol{x}_{i_{k-1}}^* - \hat{\boldsymbol{\theta}}^* \quad \boldsymbol{x}) \\ &= \det(A(\boldsymbol{x}_{i_1} - \hat{\boldsymbol{\theta}} \quad \cdots \quad \boldsymbol{x}_{i_{k-1}} - \hat{\boldsymbol{\theta}} \quad A^{-1}\boldsymbol{x})) = \det(A)\boldsymbol{e}(I; \hat{\boldsymbol{\theta}})^T A^{-1}\boldsymbol{x},\end{aligned}$$

the transformed vector of cofactors equals $\boldsymbol{e}^*(I; \hat{\boldsymbol{\theta}}^*) = \det(A)(A^{-1})^T\boldsymbol{e}(I; \hat{\boldsymbol{\theta}})$. Consequently, $\mathrm{sign}\{\boldsymbol{e}^*(I; \hat{\boldsymbol{\theta}}^*)^T(\boldsymbol{x}^* - \hat{\boldsymbol{\theta}}^*)\} = \mathrm{sign}\{\det(A)\}\,\mathrm{sign}\{\boldsymbol{e}(I; \hat{\boldsymbol{\theta}})^T(\boldsymbol{x} - \hat{\boldsymbol{\theta}})\}$. By definition of $\hat{\boldsymbol{S}}_i^* = \text{ave}_I\{\text{sign}[\boldsymbol{e}^*(I;\hat{\boldsymbol{\theta}}^*)^T(\boldsymbol{x}_i^* - \hat{\boldsymbol{\theta}}^*)]\boldsymbol{e}^*(I;\hat{\boldsymbol{\theta}}^*)\}$ and $\hat{D}^* = \text{ave}_i\{\hat{\boldsymbol{S}}_i^*(\hat{\boldsymbol{S}}_i^*)^T\}$ the stated expressions follow. □

**Proof of Theorem 1.** It is straightforward to see, using invariance in distribution properties, that $\mathbb{E}_{F_0}[\boldsymbol{S}_{F_0}(\boldsymbol{Z};\boldsymbol{0})] = \boldsymbol{0}$ and $D(F_0) = w_{F_0}I_k$, where $w_{F_0}$ is a positive constant depending on $F_0$. The affine equivariance property of $\hat{\boldsymbol{S}}_i$ and $\hat{D}$ stated in Lemma 1 also hold for the theoretical counterparts and consequently $\mathbb{E}_F[\boldsymbol{S}_F(\boldsymbol{X};T(F))] = \boldsymbol{0}$ and $D(F) = \det(\Sigma)\Sigma^{-1/2}D(F_0)\Sigma^{-1/2} = w_{F_0}\det(\Sigma)\Sigma^{-1}$. □

**Proof of Lemma 2.** First note that the expectation of the kernel is

$$\begin{aligned}&\mathbb{E}_F[\mathbb{E}_F[\text{sign}\{\boldsymbol{E}^T(I)\boldsymbol{X}_i\}\,\text{sign}\{\boldsymbol{E}^T(J)\boldsymbol{X}_i\}\boldsymbol{E}(I)\boldsymbol{E}^T(J)|\boldsymbol{X}_i]]\\ &\quad = \mathbb{E}_F[\boldsymbol{S}_F(\boldsymbol{X}_i;\boldsymbol{0})\boldsymbol{S}_F(\boldsymbol{X}_i;\boldsymbol{0})^T],\end{aligned}$$

so $\mathbb{E}_F[g(K)] = D(F)$.

Then notice that

$$\begin{aligned}&\text{ave}_i\{\boldsymbol{S}_n(\boldsymbol{X}_i)\boldsymbol{S}_n(\boldsymbol{X}_i)^T\}\\
&\quad = \frac{1}{n\binom{n}{k-1}^2}\sum_i\sum_I\sum_J \text{sign}\{\boldsymbol{E}^T(I)\boldsymbol{X}_i\}\,\text{sign}\{\boldsymbol{E}^T(J)\boldsymbol{X}_i\}\boldsymbol{E}(I)\boldsymbol{E}^T(J)\\
&\quad = \frac{(n-k+1)(n-k)\cdots(n-2k+2)}{n^2(n-1)(n-2)\cdots(n-k+2)}U_n\\
&\qquad + \frac{1}{n\binom{n}{k-1}^2}\sum_{\substack{I\cap\{i\}=J\cap\{i\}=\emptyset\\ I\cap J\neq\emptyset}} \text{sign}\{\boldsymbol{E}^T(I)\boldsymbol{X}_i\}\,\text{sign}\{\boldsymbol{E}^T(J)\boldsymbol{X}_i\}\boldsymbol{E}(I)\boldsymbol{E}^T(J)\\
&\quad = \{1 + O(1/n)\}U_n + V_n.\end{aligned}$$

The statistic $V_n$ can be further decomposed to a sum $V_n = V_{n,1} + \cdots + V_{n,k-1}$, where $V_{n,j}$, $j = 1,\ldots,k-1$, is the sum of terms where $I$ and $J$ have $j$ joint indices. Then $V_{n,j} = c_{n,j}U_{n,j}$, where $U_{n,j}$ is a $U$-statistic converging in probability to its finite expectation and $c_{n,j} = O(1/n^j)$. It follows that $\sqrt{n}\,V_n \xrightarrow{\text{P}} 0$ and the lemma is proven. □

**Proof of Lemma 3.** We only sketch the proof here. The first step (straightforward but quite tedious) is to note that

$$\sqrt{n}[\boldsymbol{S}_n(\boldsymbol{x};n^{-1/2}\boldsymbol{\theta}) - \boldsymbol{S}_n(\boldsymbol{x})] \xrightarrow{\text{P}} [\nabla\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{0})]^T\boldsymbol{\theta}$$

uniformly in $||\boldsymbol{\theta}||<\eta$ and $\boldsymbol{x}$ for a certain $\eta>0$. Further note that

$$\begin{aligned}&\sqrt{n}[\boldsymbol{S}_n(\boldsymbol{x};\boldsymbol{\theta})\boldsymbol{S}_n^T(\boldsymbol{x};\boldsymbol{\theta}) - \boldsymbol{S}_n(\boldsymbol{x})\boldsymbol{S}_n^T(\boldsymbol{x})]\\
&\quad = \sqrt{n}\{\boldsymbol{S}_n(\boldsymbol{x};\boldsymbol{\theta}) - \boldsymbol{S}_n(\boldsymbol{x})\}\boldsymbol{S}_n^T(\boldsymbol{x}) + \boldsymbol{S}_n(\boldsymbol{x})[\sqrt{n}\{\boldsymbol{S}_n(\boldsymbol{x};\boldsymbol{\theta}) - \boldsymbol{S}_n(\boldsymbol{x})\}]^T\\
&\qquad + n^{-1/2}[\sqrt{n}\{\boldsymbol{S}_n(\boldsymbol{x};\boldsymbol{\theta}) - \boldsymbol{S}_n(\boldsymbol{x})\}][\sqrt{n}\{\boldsymbol{S}_n(\boldsymbol{x};\boldsymbol{\theta}) - \boldsymbol{S}_n(\boldsymbol{x})\}]^T.\end{aligned}$$

Then since $\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{0})$ is an odd function, $\nabla \boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{0})$ an even function and since $F$ is symmetric,

$$\begin{aligned}&\sqrt{n}\,\mathrm{ave}_i\{\boldsymbol{S}_n(\boldsymbol{X}_i;n^{-1/2}\boldsymbol{\theta})\boldsymbol{S}_n^T(\boldsymbol{X}_i;n^{-1/2}\boldsymbol{\theta})-\boldsymbol{S}_n(\boldsymbol{X}_i)\boldsymbol{S}_n^T(\boldsymbol{X}_i)\}\\&\quad\xrightarrow{\mathrm{P}}\mathbb{E}_F\{[\nabla\boldsymbol{S}_F(\boldsymbol{X};\boldsymbol{0})]^T\boldsymbol{\theta}\boldsymbol{S}_F^T(\boldsymbol{X};\boldsymbol{0})\}+\mathbb{E}_F\{\boldsymbol{S}_F(\boldsymbol{X};\boldsymbol{0})\boldsymbol{\theta}^T\nabla\boldsymbol{S}_F(\boldsymbol{X};\boldsymbol{0})\}=0\end{aligned}$$

uniformly in $||\boldsymbol{\theta}||<\eta$. The result then follows as $\sqrt{n}\,\hat{\boldsymbol{\theta}}$ is bounded in probability. □

**Proof of Theorem 2.** By writing

$$\begin{aligned}D(F_\varepsilon)&=(1-\varepsilon)\mathbb{E}_F[\boldsymbol{S}_{F_\varepsilon}(\boldsymbol{X};T(F_\varepsilon))\boldsymbol{S}_{F_\varepsilon}^T(\boldsymbol{X};T(F_\varepsilon))]\\&\quad+\varepsilon\,\boldsymbol{S}_{F_\varepsilon}(\boldsymbol{x}_0;T(F_\varepsilon))\boldsymbol{S}_{F_\varepsilon}^T(\boldsymbol{x}_0;T(F_\varepsilon))\end{aligned}$$

and taking the derivative of $D(F_\varepsilon)$ with respect to $\varepsilon$ and evaluating at 0 and using $T(F)=\boldsymbol{\mu}$, we get (assuming the order of the expectation and the differentiation can be reversed)

$$\begin{aligned}\mathrm{IF}(\boldsymbol{x}_0;D,F)=&-D(F)+\boldsymbol{S}_F(\boldsymbol{x}_0;\boldsymbol{\mu})\boldsymbol{S}_F^T(\boldsymbol{x}_0;\boldsymbol{\mu})\\&+\mathbb{E}_F\left[\frac{\partial}{\partial\varepsilon}\boldsymbol{S}_{F_\varepsilon}(\boldsymbol{X};T(F_\varepsilon))\boldsymbol{S}_{F_\varepsilon}^T(\boldsymbol{X};T(F_\varepsilon))\bigg|_{\varepsilon=0}\right].\end{aligned}\tag{A.1}$$

Next step is to note that

$$\begin{aligned}\frac{\partial}{\partial\varepsilon}\boldsymbol{S}_{F_\varepsilon}(\boldsymbol{x};T(F_\varepsilon))\bigg|_{\varepsilon=0}=&-\nabla_{\boldsymbol{x}}\mathbb{E}_F[\mathrm{sign}\{\boldsymbol{E}^T(I;\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})\}\boldsymbol{E}^T(I;\boldsymbol{0})]\mathrm{IF}(\boldsymbol{x}_0;T,F)\\&+\mathrm{IF}(\boldsymbol{x}_0;\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\mu}),F).\end{aligned}$$

Since $\boldsymbol{S}_F^T(\boldsymbol{x};\boldsymbol{\mu})$ is an odd function and $\nabla_{\boldsymbol{x}}\mathbb{E}_F[\mathrm{sign}\{\boldsymbol{E}^T(I;\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})\}\boldsymbol{E}^T(I;\boldsymbol{0})]$ is an even function of $(\boldsymbol{x}-\boldsymbol{\mu})$, and since $F$ is symmetric,

$$\mathbb{E}_{\boldsymbol{X}}[\nabla_{\boldsymbol{X}}\mathbb{E}_F[\mathrm{sign}\{\boldsymbol{E}^T(I;\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})\}\boldsymbol{E}^T(I;\boldsymbol{0})\mid\boldsymbol{X}]\mathrm{IF}(\boldsymbol{x}_0;T,F)\boldsymbol{S}_F^T(\boldsymbol{X};\boldsymbol{\mu})]=0.$$

Therefore

$$\begin{aligned}&\mathbb{E}_F\left[\frac{\partial}{\partial\varepsilon}\boldsymbol{S}_{F_\varepsilon}(\boldsymbol{X};T(F_\varepsilon))\boldsymbol{S}_{F_\varepsilon}^T(\boldsymbol{X};T(F_\varepsilon))\bigg|_{\varepsilon=0}\right]\\&\quad=\mathbb{E}_F\left[\frac{\partial}{\partial\varepsilon}\boldsymbol{S}_{F_\varepsilon}(\boldsymbol{X};T(F_\varepsilon))\bigg|_{\varepsilon=0}\boldsymbol{S}_F^T(\boldsymbol{X};\boldsymbol{\mu})\right]+\mathbb{E}_F\left[\boldsymbol{S}_F(\boldsymbol{X};\boldsymbol{\mu})\frac{\partial}{\partial\varepsilon}\boldsymbol{S}_{F_\varepsilon}^T(\boldsymbol{X};T(F_\varepsilon))\bigg|_{\varepsilon=0}\right]\\&\quad=\mathbb{E}_F[\mathrm{IF}(\boldsymbol{x}_0;\boldsymbol{S}_F(\boldsymbol{X};\boldsymbol{\mu}),F)\,\boldsymbol{S}_F^T(\boldsymbol{X};\boldsymbol{\mu})]+\mathbb{E}_F[\boldsymbol{S}_F(\boldsymbol{X};\boldsymbol{\mu})\mathrm{IF}(\boldsymbol{x}_0;\boldsymbol{S}_F(\boldsymbol{X};\boldsymbol{\mu}),F)^T].\end{aligned}$$

Substituting the above equation in (A.1) gives the stated result. □

**Proof of Theorem 3.** Lemmas 2 and 3 imply that $\sqrt{n}\,(\hat{D}-U_n)\xrightarrow{\mathrm{p}}0$. This together with general properties of $U$-statistics gives the stated result. Note that for the limiting normality of the $U$-statistic $U_n$ it is enough to assume that the second-order moments exists. □

**Lemma A.3.** *The influence function of the population sign of* $\boldsymbol{x}$ *with respect to* $\mathbf{0}$ *at a standardized spherical distribution* $F_0$ *is given by Eq.* (6).

**Proof.** Write $\boldsymbol{E}(I;\mathbf{0}) = \boldsymbol{E}(I)$. First note that if $P$ is a rotation matrix, then

$$\mathbb{E}_{F_0}[\text{sign}\{\boldsymbol{E}^T(I)\boldsymbol{x}\}\boldsymbol{E}(I)|\boldsymbol{X}_{i_1} = P\boldsymbol{x}_0] = P\mathbb{E}_{F_0}[\text{sign}\{\boldsymbol{E}^T(I)P^T\boldsymbol{x}\}\boldsymbol{E}(I)|\boldsymbol{X}_{i_1} = \boldsymbol{x}_0]. \quad \text{(A.2)}$$

First consider the special case $\boldsymbol{x}_0 = \delta\boldsymbol{v} = \delta(1,0,\dots,0)^T$. Similarly, as in Lemma A.2, we can show that

$$\mathbb{E}[\text{sign}\{\boldsymbol{E}^T(I)\boldsymbol{x}\}\boldsymbol{E}(I)|\boldsymbol{X}_{i_1} = \boldsymbol{x}_0] = \delta c'_{F_0}\frac{(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{x}}{||(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{x}||} \tag{A.3}$$

with

$$c'_{F_0} = \frac{\Gamma^{k-1}(\frac{k-1}{2})\mathbb{E}^{k-2}_{F_0}(R')}{\sqrt{\pi}\Gamma^{k-2}(\frac{k}{2})}, \tag{A.4}$$

where $R'$ is a length of $(k-1)$-variate subvector of $\boldsymbol{X} = R\boldsymbol{U} \sim F_0$. So we may set $R' = (R^2 - X_1^2)^{1/2} = R(1 - U_1^2)^{1/2}$. Expression (7) for $c'_{F_0}$ follows by using the relation $\mathbb{E}_{F_0}(R') = \mathbb{E}_{F_0}(R)\mathbb{E}[(1 - U_1^2)^{1/2}]$ together with Lemma A.1(d), and Eq. (3) for $c_{F_0}$.

Next, consider the general case $\boldsymbol{x}_0 = \delta\boldsymbol{u}_0$ and let $P$ be a rotation matrix ($PP^T = I_k$) such that $P\boldsymbol{v} = \boldsymbol{u}_0$. Then Eq. (A.2) together with Eq. (A.3) imply that

$$\begin{aligned}
&\mathbb{E}_{F_0}[\text{sign}\{\boldsymbol{E}^T(I)\boldsymbol{x}\}\boldsymbol{E}(I) \mid \boldsymbol{X}_{i_1} = \boldsymbol{x}_0] \\
&\quad = \delta c'_{F_0}\frac{P(I_k - \boldsymbol{v}\boldsymbol{v}^T)P^T\boldsymbol{x}}{||(I_k - \boldsymbol{v}\boldsymbol{v}^T)P^T\boldsymbol{x}||} = \delta c'_{F_0}\frac{(I_k - \boldsymbol{u}_0\boldsymbol{u}_0^T)\boldsymbol{x}}{||(I_k - \boldsymbol{u}_0\boldsymbol{u}_0^T)\boldsymbol{x}||}
\end{aligned}$$

which, by using Eq. (4), gives the desired expression. □

**Proof of Theorem 4.** First, we derive the influence function for a point in the direction of the first axis, $\boldsymbol{x}_\delta = \delta\boldsymbol{v}$, with $\boldsymbol{v} = (1,0,\dots,0)^T$. By Theorem 2,

$$\begin{aligned}
\text{IF}(\boldsymbol{x}_\delta; D, F_0) = {} & \boldsymbol{S}_{F_0}(\boldsymbol{x}_\delta;\mathbf{0})\boldsymbol{S}_F^T(\boldsymbol{x}_\delta;\mathbf{0}) + \mathbb{E}_{F_0}[\text{IF}(\boldsymbol{x}_\delta;\boldsymbol{S}_{F_0}(\boldsymbol{X};\mathbf{0}),F_0)\boldsymbol{S}_{F_0}^T(\boldsymbol{X};\mathbf{0})] \\
& + \mathbb{E}_{F_0}[\boldsymbol{S}_{F_0}(\boldsymbol{X};\mathbf{0})\,\text{IF}(\boldsymbol{x}_\delta;\boldsymbol{S}_{F_0}(\boldsymbol{X};\mathbf{0}),F_0)^T] - D(F_0),
\end{aligned} \tag{A.5}$$

since $\boldsymbol{\mu} = T(F_0) = \mathbf{0}$. Then use $\boldsymbol{S}_{F_0}(\boldsymbol{x};\mathbf{0}) = c_{F_0}\boldsymbol{u}$ with $\boldsymbol{u} = \boldsymbol{x}||\boldsymbol{x}||^{-1}$, $D(F_0) = \mathbb{E}_{F_0}[\boldsymbol{S}_{F_0}(\boldsymbol{X};\mathbf{0})\boldsymbol{S}_{F_0}^T(\boldsymbol{X};\mathbf{0})]$ together with Eq. (6) to obtain

$$\begin{aligned}
&\mathbb{E}_{F_0}[\text{IF}(\boldsymbol{x}_\delta;\boldsymbol{S}_{F_0}(\boldsymbol{X};\mathbf{0}),F_0)\boldsymbol{S}_{F_0}^T(\boldsymbol{X};\mathbf{0})] \\
&\quad = (k-1)\mathbb{E}_{F_0}\left[\left\{\delta c'_{F_0}\frac{(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{X}}{||(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{X}||} - \boldsymbol{S}_{F_0}(\boldsymbol{X};\mathbf{0})\right\}\boldsymbol{S}_{F_0}^T(\boldsymbol{X};\mathbf{0})\right] \\
&\quad = (k-1)\delta c_{F_0}c'_{F_0}\mathbb{E}_{F_0}\left[\frac{(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{X}\boldsymbol{X}^T}{||(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{X}||\,||\boldsymbol{X}||}\right] - (k-1)D(F_0).
\end{aligned} \tag{A.6}$$

By noticing that

$$\mathbb{E}_{F_0}\left[\frac{(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{X}\boldsymbol{X}^T}{||(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{X}||\,||\boldsymbol{X}||}\right] = \mathbb{E}\left[\frac{U_2^2}{(1-U_1^2)^{1/2}}\right]\begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I_{k-1}\end{pmatrix}$$

and by substituting Eq. (7) for $c'_{F_0}$ and using Lemma A.1(e), Eq. (A.6) simplifies to

$$\mathbb{E}_{F_0}[\mathrm{IF}(\boldsymbol{x}_\delta; \boldsymbol{S}_{F_0}(\boldsymbol{X};\mathbf{0}), F_0)\boldsymbol{S}_{F_0}^T(\boldsymbol{X};\mathbf{0})] = \delta c_{F_0}^2 \mathbb{E}_{F_0}^{-1}(R)\begin{pmatrix} 0 & \mathbf{0}^T \\ \mathbf{0} & I_{k-1}\end{pmatrix} - (k-1)D(F_0).$$

Hence we may now write (A.5) as

$$\mathrm{IF}(\boldsymbol{x}_\delta; D, F_0) = c_{F_0}^2\begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & 2\delta\, \mathbb{E}_{F_0}^{-1}(R)\, I_{k-1}\end{pmatrix} - (2k-1)D(F_0). \tag{A.7}$$

An influence point in an arbitrary direction is obtained by setting $\boldsymbol{x}_0 = P\boldsymbol{x}_\delta = \delta\boldsymbol{p}_1$ for a well chosen rotation matrix $P = [\boldsymbol{p}_1 \cdots \boldsymbol{p}_k]$ with $P^TP = I_k$. Then $||\boldsymbol{x}_0|| = \delta$. The influence function is then given by $\mathrm{IF}(\boldsymbol{x}_0; D, F_0) = P\mathrm{IF}(\boldsymbol{x}_\delta; D, F_0)P^T$, which, after using (A.7) and some simple matrix manipulation, yields $\mathrm{IF}(\boldsymbol{x}_0; D, F_0) = \alpha(\delta)\boldsymbol{x}_0\boldsymbol{x}_0^T - \beta(\delta)D(F_0)$ with $\alpha$ and $\beta$ as stated in the theorem. □

**Proof of Corollary 1.** Affine equivariance of $D$ yields

$$\mathrm{IF}(\boldsymbol{x}_0; D, F) = \det(\Sigma)\Sigma^{-1/2}\mathrm{IF}(\Sigma^{-1/2}(\boldsymbol{x}_0 - \boldsymbol{\mu}); D, F_0)\Sigma^{-1/2}.$$

Applying Theorem 4 yields the stated expression. □

**Proof of Theorem 5.** Lemma 3 in [7] combined with $\boldsymbol{v}_{D,j}(F) = \boldsymbol{v}_j$ and $\lambda_{D,j}(F) = \det(\Sigma)(c_{F_0}^2/k)\lambda_j^{-1}$ implies that

$$\mathrm{IF}(\boldsymbol{x}_0; \lambda_{D,j}, F) = \boldsymbol{v}_j^T\mathrm{IF}(\boldsymbol{x}_0; D, F)\boldsymbol{v}_j,$$

$$\mathrm{IF}(\boldsymbol{x}_0; \boldsymbol{v}_{D,j}, F) = \frac{-k}{\det(\Sigma)c_{F_0}^2}\sum_{\substack{i=1\\ i\neq j}}^{k}\frac{\lambda_j\lambda_i}{\lambda_j - \lambda_i}\{\boldsymbol{v}_i^T\mathrm{IF}(\boldsymbol{x}_0; D, F)\boldsymbol{v}_j\}\boldsymbol{v}_i.$$

By Corollary 1 one has that

$$\begin{aligned}\boldsymbol{v}_i^T\mathrm{IF}(\boldsymbol{x}_0; D, F)\boldsymbol{v}_j^T &= \alpha(d(\boldsymbol{x}_0))\det(\Sigma)\boldsymbol{v}_i^T\Sigma^{-1}(\boldsymbol{x}_0-\boldsymbol{\mu})(\boldsymbol{x}_0-\boldsymbol{\mu})^T\Sigma^{-1}\boldsymbol{v}_j \\ &\quad - \beta(d(\boldsymbol{x}_0))\boldsymbol{v}_i^TD(F)\boldsymbol{v}_j.\end{aligned}$$

By noting that $\boldsymbol{v}_i^T\Sigma^{-1} = \lambda_i^{-1}\boldsymbol{v}_i^T$, $\boldsymbol{v}_i^TD(F)\boldsymbol{v}_j = \lambda_{D,j}\delta_{ij} = \det(\Sigma)(c_{F_0}^2/k)\lambda_j^{-1}\delta_{ij}$ ($\delta_{ij} = 1$ if $i = j$ and 0 otherwise), and then replacing $\boldsymbol{v}_i^T(\boldsymbol{x}_0 - \boldsymbol{\mu})$ by $z_i$ yields the stated expressions. □

**Proof of Theorem 6.** The proof follows from Theorem 3 the fact that, $D \to (P, \Lambda)$ is a bijection and has non-zero differentials in a neighborhood of the true value (note that we assumed distinct eigenvalues). See Theorem 3.3.A in [25] or Theorem 13.5.1 in [1].

**Proof of Corollary 3.** The asymptotic variance of $\hat{\lambda}_{D,j}$ is

$$\begin{aligned}\mathrm{ASV}(\hat{\lambda}_{D,j};F) &= \mathbb{E}_F[\mathrm{IF}(\boldsymbol{X};\lambda_{D,j},F)^2] \\ &= (\det(\Sigma)/\lambda_j)^2\mathbb{E}_F[\{\alpha(d(\boldsymbol{X}))(Z_j/\sqrt{\lambda_j})^2 - \beta(d(\boldsymbol{X}))(c_{F_0}^2/k)\}^2],\end{aligned}$$

where $Z_j = \boldsymbol{v}_j^T(\boldsymbol{X}-\boldsymbol{\mu})$.

With $Y_j = Z_j/\sqrt{\lambda_j}$, one has that $\boldsymbol{Y} = (Y_1,\ldots,Y_k)^T \sim F_0$ and $d(\boldsymbol{X}) = ||\boldsymbol{Y}||$. This yields

$$\begin{aligned}\mathrm{ASV}(\hat{\lambda}_{D,j};F) &= (\det(\Sigma)/\lambda_j)^2\mathbb{E}_{F_0}[\{\alpha(||\boldsymbol{Y}||)Y_j^2 - \beta(||\boldsymbol{Y}||)(c_{F_0}^2/k)\}^2] \\ &= (\det(\Sigma)/\lambda_j)^2\mathbb{E}_{F_0}[\mathrm{IF}(\boldsymbol{Y};D_{jj},F_0)^2] = (\det(\Sigma)/\lambda_j)^2\mathrm{ASV}(\hat{D}_{11};F_0)\end{aligned}$$

as $\mathrm{ASV}(\hat{D}_{jj};F_0) = \mathrm{ASV}(\hat{D}_{11};F_0)$ by symmetry.

For the eigenvector estimator, the asymptotic variance is given by

$$\begin{aligned}\mathrm{ASV}(\hat{\boldsymbol{v}}_{D,j};F) &= \mathbb{E}_F[\mathrm{IF}(\boldsymbol{X};\boldsymbol{v}_{D,j},F)\mathrm{IF}(\boldsymbol{X};\boldsymbol{v}_{D,j},F)^T] \\ &= \sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{\substack{l=1\\ l\neq j}}^{k}\frac{1}{\lambda_j-\lambda_i}\frac{1}{\lambda_j-\lambda_l}\frac{k^2}{c_{F_0}^4}\mathbb{E}_F[\alpha(d(\boldsymbol{X}))^2 Z_i Z_l Z_j^2]\boldsymbol{v}_i\boldsymbol{v}_l^T. \qquad \text{(A.8)}\end{aligned}$$

Using the transformation $Z_j/\sqrt{\lambda_j} = Y_j$, the expectation in (A.8) is simply

$$\begin{aligned}\mathbb{E}_F[\alpha^2(d(\boldsymbol{X}))Z_iZ_lZ_j^2] &= \sqrt{\lambda_i}\sqrt{\lambda_l}\lambda_j\mathbb{E}_{F_0}[\alpha(||\boldsymbol{Y}||)^2 Y_iY_lY_j^2] \\ &= \lambda_i\lambda_j\mathbb{E}_{F_0}[\alpha(||\boldsymbol{Y}||)^2Y_i^2Y_j^2]\delta_{il} \\ &= \lambda_i\lambda_j\mathbb{E}_{F_0}[\mathrm{IF}(\boldsymbol{Y};D_{ij},F_0)^2]\delta_{il} = \lambda_i\lambda_j\mathrm{ASV}(\hat{D}_{12};F_0)\delta_{il}\end{aligned}$$

as $\mathrm{ASV}(\hat{D}_{ij};F_0) = \mathrm{ASV}(\hat{D}_{12};F_0)$ by symmetry. Consequently, we obtain the stated expression for $\mathrm{ASV}(\hat{\boldsymbol{v}}_{D,j};F)$. The asymptotic covariances are found in a similar manner. □

**Proof of Theorem 7.** Under the stated assumption, we may write $\sqrt{n}\,(\hat{\boldsymbol{\lambda}}_C - \boldsymbol{\lambda}) \xrightarrow{\mathrm{d}} N(\boldsymbol{0},B)$, where $\boldsymbol{\lambda} = (\lambda_1,\ldots,\lambda_k)^T$ and the diagonal elements of the asymptotic covariance matrix $B = \mathbb{E}_F[\mathrm{IF}(\boldsymbol{X};\boldsymbol{\lambda}_C,F)\mathrm{IF}(\boldsymbol{X};\boldsymbol{\lambda}_C,F)^T]$ are

$$b_{jj} = \mathrm{ASV}(\hat{\lambda}_{C,j};F) = \lambda_j^2\mathrm{ASV}(\hat{C}_{11};F_0)$$

for $j=1,\ldots,k$ (Corollary 1 of Croux and Haesbroeck [7]). It is easy to derive the expression for the off-diagonal elements (limiting covariances)

$$b_{ij} = \mathrm{ASC}(\hat{\lambda}_{C,i},\hat{\lambda}_{C,j};F) = \lambda_i\lambda_j\mathrm{ASC}(\hat{C}_{11},\hat{C}_{22};F_0)$$

for $1 \leqslant i \neq j \leqslant k$.

Write

$$g(\boldsymbol{x}) = g(x_1, \dots, x_k) = \frac{k-1}{k} \ln x_j - \frac{1}{k} \sum_{i=1, i \neq j}^{k} \ln x_i.$$

By the multivariate version of the delta-method one has that

$$\sqrt{n}\,(\ln \hat{\lambda}^*_{C,j} - \ln \lambda^*_{C,j}) \xrightarrow{\mathrm{d}} N(0, [\nabla g(\boldsymbol{\lambda}_C)]^T B g(\boldsymbol{\lambda}_C)).$$

After some easy matrix algebra we may write

$$[\nabla g(\boldsymbol{\lambda}_C)]^T B\, g(\boldsymbol{\lambda}_C) = \frac{k-1}{k} \{\mathrm{ASV}(\hat{C}_{11}; F_0) - \mathrm{ASC}(\hat{C}_{11}, \hat{C}_{22}; F_0)\}.$$

It is not difficult to find out, using e.g. the general expression (8) for the influence function of any affine equivariant scatter matrix estimator, that $\mathrm{ASV}(\hat{C}_{11}; F_0) - \mathrm{ASC}(\hat{C}_{11}, \hat{C}_{22}; F_0) = 2\mathrm{ASV}(\hat{C}_{12}; F_0)$. Hence one has that

$$\mathrm{ASV}(\ln \hat{\lambda}^*_{C,j}; F) = [\nabla g(\boldsymbol{\lambda}_D)]^T B g(\boldsymbol{\lambda}_D) = \frac{2(k-1)}{k} \mathrm{ASV}(\hat{C}_{12}; F_0)$$

which completes the proof. $\square$

## References


[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, 2nd Edition, Wiley, New York, 1984.

[2] H. Bensmail, G. Celeux, Regularized Gaussian discriminant analysis through eigenvalue decomposition, J. Amer. Statist. Assoc. 91 (1996) 1743–1749.

[3] B.M. Brown, P. Hall, G.A. Young, On the effect of inliers on the spatial median, J. Multivariate Anal. 63 (1997) 88–104.

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

[5] C. Croux, C. Dehon, Robust linear discriminant analysis using *S*-estimators, Canad. J. Statist. 29 (2001) 473–492.

[6] C. Croux, G. Haesbroeck, Influence function and efficiency of the minimum covariance determinant scatter matrix estimator, J. Multivariate Anal. 71 (1999) 161–190.

[7] C. Croux, G. Haesbroeck, Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies, Biometrika 87 (2000) 603–618.

[8] H. Cui, X. He, K.W. Ng, Asymptotic distributions of principal components based on robust dispersions, Technical Report, 2002.

[9] P.L. Davies, Asymptotic behavior of *S*-estimators of multivariate location parameters and dispersion matrices, Ann. Statist. 15 (1987) 1269–1292.

[10] S.J. Devlin, R. Gnanadesikan, J.R. Kettenring, Robust estimation of dispersion matrices and principal components, J. Amer. Statist. Assoc. 76 (1981) 354–362.

[11] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel, Robust Statistics: The Approach Based on Influence Functions, Wiley, New York, 1986.

[12] T.P. Hettmansperger, J. Nyblom, H. Oja, Affine invariant multivariate one-sample sign tests, J. Roy. Statist. Soc. Ser. B 56 (1994) 221–234.

[13] T.P. Hettmansperger, H. Oja, Affine invariant multivariate multisample sign tests, J. Roy. Statist. Soc. Ser. B 56 (1994) 235–249.

[14] H.P. Lopuhaä, On the relation between *S*-estimators and *M*-estimators of multivariate location and covariance, Ann. Statist. 17 (1989) 1662–1683.

[15] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley, New York, 1988.
[16] R.A. Maronna, V. Yohai, Robust estimation of multivariate location and scatter, in: S. Kotz, C. Read, D. Banks (Eds.), Encyclopedia of Statistical Sciences Update, Vol. 2, Wiley, New York, 1998, pp. 589–596.
[17] J. Möttönen, T.P. Hettmansperger, H. Oja, J. Tienari, On the efficiency of affine invariant multivariate rank tests, J. Multivariate Statist. 66 (1998) 108–132.
[18] A. Niinimaa, H. Oja, J. Nyblom, Algorithm AS 277: the bivariate Oja median, Appl. Statist. 41 (1992) 611–633.
[19] H. Oja, Descriptive statistics for multivariate distributions, Statist. Probab. Lett. 1 (1983) 327–332.
[20] H. Oja, Affine invariant multivariate sign and rank tests and corresponding estimates: a review, Scand. J. Statist. 26 (1999) 319–343.
[21] E. Ollila, T.P. Hettmansperger, H. Oja, Estimates of regression coefficients based on the sign covariance matrix, J. Roy. Statist. Soc. Ser. B 64 (2002) 447–466.
[22] T. Ronkainen, H. Oja, P. Orponen, Computation of the multivariate Oja median, in: R. Dutter, P. Filzmoser, U. Gather, P.J. Rousseeuw (Eds.), Developments in Robust Statistics, Springer-Verlag, Heidelberg, 2003, pp. 344–359.
[23] P.J. Rousseeuw, Multivariate estimation with high breakdown point, in: W. Grossmann, G. Pflug, I. Vincze, W. Wertz (Eds.), Mathematical Statistics and Applications, Vol. B, Dordrecht, Reidel, 1985, pp. 283–297.
[24] P.J. Rousseeuw, A.M. Leroy, Robust Regression and Outlier Detection, Wiley, New York, 1987.
[25] R.J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, New York, 1980.
[26] S. Visuri, V. Koivunen, H. Oja, Sign and rank covariance matrices, J. Statist. Plann. Inference 91 (2000) 557–575.