The Affine Equivariant Sign Covariance Matrix: Asymptotic Behavior and Efficiencies

Esa Ollila[1] Hannu Oja[1] Christophe Croux[2]

July 2, 2001

[1] *Department of Mathematics and Statistics, University of Jyväskylä, P.O. Box 35, FIN-40351 Jyväskylä, Finland*
(esaolli@cc.jyu.fi, ojahannu@cc.jyu.fi)

[2] *ECARES and Institut de Statistique, Université Libre de Bruxelles, CP-114, Av F.D. Roosevelt 50, B-1050 Bruxelles, Belgium*
(ccroux@ulb.ac.be)

Abstract

In the paper we consider the affine equivariant Sign Covariance Matrix (SCM) introduced by Visuri *et al.* (2000). 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 the sample covariance matrix estimate for heavy tailed distributions. Simulations confirmed these findings for finite-sample efficiencies.

Key words: Affine equivariance; covariance and correlation matrices; efficiency; eigenvectors and eigenvalues; influence function; multivariate median; multivariate sign; robustness;

1 Introduction

Let $\boldsymbol{x}$ be a k-dimensional random vector with finite second moments. Denote $\Sigma = \mathrm{Cov}(\boldsymbol{x})$ its covariance matrix, which we suppose to be non-singular. The eigenvalue 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 Σ in its columns and Λ is a diagonal matrix with the corresponding eigenvalues $\lambda_1, \ldots, \lambda_k$ as diagonal elements. We may also state the eigenvalue decomposition in the form

$$\Sigma = \lambda\ P\ \Lambda^*\ P^T = \lambda\ \Sigma^*$$

where $\lambda = (\lambda_1 \cdots \lambda_k)^{1/k}$ is the geometrical mean of the eigenvalues and $\Lambda = \lambda\Lambda^*$. The matrix Λ^* 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 (1996) use the terms *scale*, *shape* and *orientation* for the items λ, Λ^* and P.

In this paper we consider the affine equivariant *Sign Covariance Matrix* (SCM) which can be used to estimate the shape Λ^* and orientation P of the covariance matrix. Under a specified elliptical model distribution a consistent estimate of Σ can be obtained. The SCM estimator has been proposed by Visuri, Koivunen and Oja (2000), 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 (Hettmansperger, Nyblom and Oja, 1994) and for MANOVA (Hettmansperger and Oja, 1994). For a review of multivariate signs and ranks, see Oja (1999).

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.* (1981) by means of M-estimators. More recently, Croux and Haesbroeck (2000) computed influence functions and efficiencies for eigenvectors and eigenvalues of high breakdown estimators of covariance. A PCA based on the sign covariance matrix will not have a high breakdown point, but will be 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, while 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.

2 Affine Equivariant Sign Covariance Matrix

In the univariate case the sign of x with respect to θ is the derivative of

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

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

Next we extend this definition to the multivariate setting. Assume that $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ are k-variate ($k > 1$) observations. The multivariate Oja (1983) median $\widehat{\boldsymbol{\theta}}$ minimizes the criterion function

$$\sum_{i_1<\ldots<i_k} V(\boldsymbol{x}_i,\ldots,\boldsymbol{x}_{i_k},\boldsymbol{\theta})$$

where

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

is the volume of the k-variate simplex determined by the vertices $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k$ along with $\boldsymbol{x}$. To shorten the notations, write $I = (i_1,\ldots,i_{k-1})$ with $1 \le i_1 < \ldots < i_{k-1} \le n$, 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 $\boldsymbol{x}$ with respect to $\boldsymbol{\theta}$ is the gradient of

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

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

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

The empirical *multivariate signs* with respect to $\widehat{\boldsymbol{\theta}}$ are then defined, as in the univariate case, by

$$\widehat{\boldsymbol{S}}_i = \boldsymbol{S}_n(\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}}) \qquad i = 1,\ldots,n\,,$$

where $\widehat{\boldsymbol{\theta}}$ is the multivariate Oja median. These multivariate signs are thus centered, so $\sum_i \widehat{\boldsymbol{S}}_i = \mathbf{0}$. The **sign covariance matrix** (SCM) is now simply defined as the usual covariance matrix computed from the empirical multivariate signs:

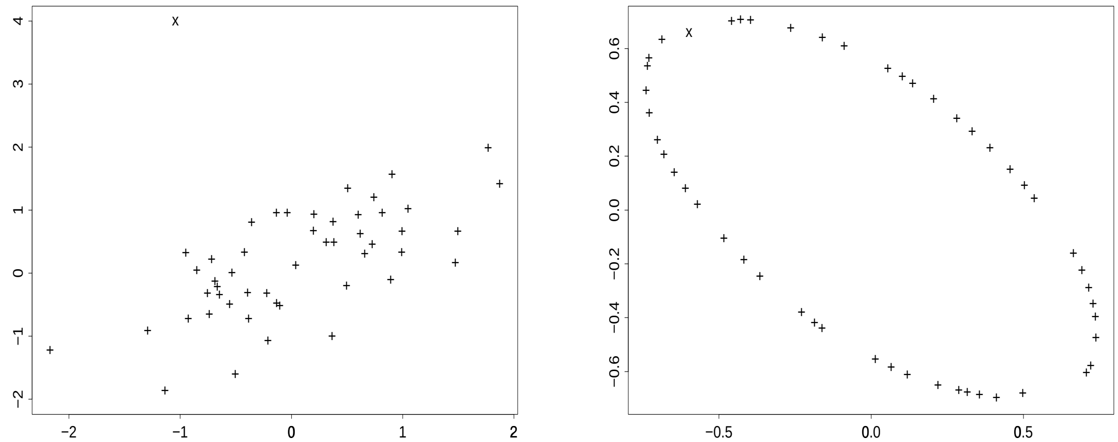
$$\widehat{D} = \text{ave}_i\left\{\widehat{\boldsymbol{S}}_i\widehat{\boldsymbol{S}}_i^T\right\} .$$

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

Lemma 1 *Let the sign vectors* $\widehat{\boldsymbol{S}_i^*}$ *and the SCM* $\widehat{D}^*$ *be calculated from the transformed observations* $\boldsymbol{x}_i^* = A\boldsymbol{x}_i + \boldsymbol{b}$ *for a nonsingular matrix* A *and* $\boldsymbol{b}$ *a* k*-dimensional vector. Then*

$$\widehat{\boldsymbol{S}_i^*} = \text{abs}\{\det(A)\}(A^{-1})^T\widehat{\boldsymbol{S}}_i \quad \textit{and} \quad \widehat{D}^* = \det(A)^2(A^{-1})^T\widehat{D}A^{-1}.$$

The proof of Lemma 1 is, as all the other proofs, in the Appendix. In Figure 1, a bivariate data set is pictured (left panel) together with the corresponding sign vectors $\widehat{\boldsymbol{S}}_i$ (right panel). We see that the signs move the data points towards the periphery of an ellipse. The sign vector points in the direction of the observations, while its magnitude depends on the dispersion of the data in the space orthogonal to the sign vector. The form of this ellipse is therefore merely determined by the *inverse* of the covariance structure of the data. As can be seen from Figure 1, this structure has not been influenced by the outlier (marked by $\times$) present in the data cloud.



a) b)

Figure 1: Representation of a bivariate data cloud (left panel) together with the corresponding sign vectors (right panel).

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 $E_F\big[V(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k,\boldsymbol{\theta})\big]$ or solves $\nabla E_F\big[V(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k,T(F))\big]=\mathbf{0}$. The population multivariate sign of $\boldsymbol{x}$ with respect to $\boldsymbol{\theta}$ is given by

$$\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\theta}) = E_F\big[\mathrm{sign}\big\{\boldsymbol{e}^T(I;\boldsymbol{\theta})(\boldsymbol{x}-\boldsymbol{\theta})\big\}\boldsymbol{e}(I;\boldsymbol{\theta})\big]\,,$$

where the expectation is taken over $k-1$ independent observations from F with indices listed in I. Note that the population sign of $\boldsymbol{x}$ with respect to Oja median, $\boldsymbol{S}_F\big(\boldsymbol{x};T(F)\big)$, has expected value zero. Finally, the population sign covariance matrix is

$$D = D(F) = E_F\big[\boldsymbol{S}_F\big(\boldsymbol{x};T(F)\big)\boldsymbol{S}_F^T\big(\boldsymbol{x};T(F)\big)\big]\,.$$

Note that $D(F)$ exists if the first moments of F are finite. For existence of the classical sample covariance matrix the stronger condition of finite second moments is required.

3 The Relation Between Σ and the SCM

3.1 Elliptic 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, 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. Moreover $\boldsymbol{u}$ is uniformly distributed on the periphery of 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 (cfr. Lemma 6 in the Appendix)

$$\boldsymbol{S}_{F_0}(\boldsymbol{x};\boldsymbol{0}) = \frac{\Gamma^k(\frac{k}{2})E_{F_0}^{k-1}[r]}{\sqrt{\pi}\Gamma^{k-1}(\frac{k+1}{2})}\,\boldsymbol{u} = c_{F_0}\,\boldsymbol{u}\,, \tag{3}$$

which is a constant times the direction vector or spatial sign. Thus population signs at F_0 are on the periphery of a sphere with radius c_{F_0}. For example, in case of the k-variate standard normal distribution, which we will denote by Φ, we use that r^2 follow a chi-square distribution with k degrees of freedom to get

$$c_\Phi = \frac{2^{\frac{k-1}{2}}\Gamma(\frac{k}{2})}{\sqrt{\pi}}. \tag{4}$$

From (3) it readily follow that the theoretical SCM of F_0 equals

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

We used here that $E[u_i^2] = 1/k$ and $E[u_iu_j] = 0$, $i \neq j$, where u_i and u_j are distinct components of $\boldsymbol{u}$. (At several places throughout the paper we need to compute moments involving the components of $\boldsymbol{u}$; see Lemma 5 in the Appendix.)

Next we construct a random variable with an elliptic distribution: let the distribution F_0 of $\boldsymbol{z}$ be spherical with mean vector $\boldsymbol{0}$ and covariance matrix I_k and write

$$\boldsymbol{x} = \Sigma^{1/2}\boldsymbol{z} + \boldsymbol{\mu},$$

where Σ is a positive definite symmetric $k \times k$ matrix and $\boldsymbol{\mu}$ a vector of length k. Then the distribution F of $\boldsymbol{x}$ is said to be elliptically symmetric with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Due to the affine equivariance of the Oja median, we have $T(F) = \boldsymbol{\mu}$. Moreover, due to affine equivariance of the population sign of $\boldsymbol{x}$ with respect to $\boldsymbol{\mu}$ (as in Lemma 1)

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

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 variable $\boldsymbol{z}$ 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 ± 1. For reflection and permutation invariant distributions, the marginal variables are identically distributed, symmetric about zero and uncorrelated. Distributions with independent margins symmetric about zero as well as distributions spherically symmetric around the origin are naturally reflection and permutation invariant.

If $\boldsymbol{z}$ has a reflection and permutation invariant k-variate standardized distribution, then it generates a **location-scale model** given by the family of distributions of the random variables

$$\boldsymbol{x} = \Sigma^{1/2}\boldsymbol{z} + \boldsymbol{\mu}$$

for every nonsingular symmetric $k \times k$ matrix Σ and $k \times 1$ vector $\boldsymbol{\mu}$. The mean vector and the covariance matrix of $\boldsymbol{x}$ are again $\boldsymbol{\mu}$ and Σ, respectively.

An elliptical model is a special case of a location-scale family. A nonelliptic example is given by the choice where the margins of $\boldsymbol{z}$ are i.i.d. random variables from a Laplace distribution with expected value zero.

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 F the distribution of $\boldsymbol{x} = \Sigma^{1/2}\boldsymbol{z} + \boldsymbol{\mu}$. Then*

$$E_F\left[\boldsymbol{S}_F\big(\boldsymbol{x}; T(F)\big)\right] = \mathbf{0} \quad \textit{and} \quad D(F) = w_{F_0} \det(\Sigma)\Sigma^{-1}$$

where w_{F_0} is a constant depending on F_0 only.

Theorem 1 shows that the Sign Covariance Matrix is proportional to the inverse of the covariance matrix (and therefore also to the matrix of pairwise partial correlations) 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 Σ.

The functional $T(F)$ in the above Theorem was taken to be the Oja median, but it can be replaced by any affine equivariant location function satisfying $T(F_0) = \mathbf{0}$. Also when considering the asymptotic behavior of the SCM in the next section, 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 random sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ from a *symmetric* k-variate distribution F with finite second moments. We say that $\boldsymbol{x}$ follows a symmetric distributions if there exists a vector $\boldsymbol{\mu}$ such that the distributions of $\boldsymbol{x}-\boldsymbol{\mu}$ and $-(\boldsymbol{x}-\boldsymbol{\mu})$ are the same. Notice 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 $\boldsymbol{\mu} = \mathbf{0}$.

We will use as shorthand notations

$$\boldsymbol{e}(I) = \boldsymbol{e}(I; \mathbf{0}) \quad \text{and} \quad \boldsymbol{S}_n(\boldsymbol{x}) = \boldsymbol{S}_n(\boldsymbol{x}; \mathbf{0}).$$

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 *For any $K = \{i_1, \ldots, i_{2k-1}\} \subset \{1, \ldots, n\}$, write*

$$g(K) = g(\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_{2k-1}}) = \frac{(k-1)!^2}{(2k-1)!} \sum_{I \cup J \cup \{i\} = K} \text{sign}\{\boldsymbol{e}^T(I)\boldsymbol{x}_i\}\text{sign}\{\boldsymbol{e}^T(J)\boldsymbol{x}_i\}\boldsymbol{e}(I)\boldsymbol{e}^T(J)\,.$$

Consider then the U-statistic with kernel g:

$$U_n = \frac{1}{\binom{n}{2k-1}} \sum_K g(K).$$

We have that $\sqrt{n}\left(U_n - \text{ave}_i\{\boldsymbol{S}_n(\boldsymbol{x}_i)\boldsymbol{S}_n(\boldsymbol{x}_i)^T\}\right) \xrightarrow{P} 0.$

The next Lemma allows to replace the sample estimate $\widehat{\boldsymbol{\theta}}$ by the population value $\boldsymbol{\mu}$ of the location estimator for asymptotical considerations.

Lemma 3 *With the notations stated above,*

$$\sqrt{n}\left(\text{ave}_i\{\boldsymbol{S}_n(\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}})\boldsymbol{S}_n^T(\boldsymbol{x}_i;\widehat{\boldsymbol{\theta}})\} - \text{ave}_i\{\boldsymbol{S}_n(\boldsymbol{x}_i)\boldsymbol{S}_n^T(\boldsymbol{x}_i)\}\right) \xrightarrow{P} 0$$

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

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

$$F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_{\boldsymbol{x}_0}$$

where $\Delta_{\boldsymbol{x}_0}$ is the cumulative distribution function of a distribution with probability mass one at $\boldsymbol{x}_0$. The influence function is now defined as

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

The IF is a tool to describe robustness properties of an estimator, but it can also be used to compute asymptotic variance (cfr. Hampel *et al.* (1986) for more information on influence functions).

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

$$\text{IF}(\boldsymbol{x}_0; \boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\theta}), F) = (k-1)\big\{E_F\big[\text{sign}\big\{\boldsymbol{e}^T(I;\boldsymbol{\theta})(\boldsymbol{x}-\boldsymbol{\theta})\big\}\boldsymbol{e}(I;\boldsymbol{\theta})|\boldsymbol{x}_{i_1}=\boldsymbol{x}_0\big] - \boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\theta})\big\}, \tag{5}$$

since $\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\theta})$ is a U-statistic with kernel of order $k-1$. For the Sign Covariance Matrix we obtain

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

$$\begin{aligned}\text{IF}(\boldsymbol{x}_0; D, F) = \boldsymbol{S}_F\big(\boldsymbol{x}_0;\boldsymbol{\mu}\big)\boldsymbol{S}_F^T(\boldsymbol{x}_0;\boldsymbol{\mu}) + E_F\big[\text{IF}(\boldsymbol{x}_0;\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\mu}),F)\,\boldsymbol{S}_F^T(\boldsymbol{x};\boldsymbol{\mu})\big] \\ + E_F\big[\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\mu})\,\text{IF}(\boldsymbol{x}_0;\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\mu}),F)^T\big] - D(F).\end{aligned} \tag{6}$$

The main result of this section is stated now. We use “vec” as operator working on matrices: 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 $\widehat{D} \xrightarrow{P} D$ and $\sqrt{n}\,\mathrm{vec}(\widehat{D} - D)$ has a limiting multinormal distribution with zero mean and asymptotic variance-covariance matrix*

$$\mathrm{ASV}(\widehat{D}; F) = E_F[\mathrm{vec}\{\mathrm{IF}(\boldsymbol{x}; D, F)\}\mathrm{vec}\{\mathrm{IF}(\boldsymbol{x}; D, F)\}^T].$$

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

4.2 Special case: the elliptical model

In case of an elliptically symmetric model distribution, it is possible to render the equations (5) and (6) 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 (cfr. Lemma 7 in the Appendix)

$$\mathrm{IF}(\boldsymbol{x}_0; \boldsymbol{S}_{F_0}(\boldsymbol{x}; \mathbf{0}), F_0) = (k-1)\left\{\delta\, c'_{F_0} \frac{(I_k - \boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{x}}{\|(I_k - \boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{x}\|} - \boldsymbol{S}_{F_0}(\boldsymbol{x}; \mathbf{0})\right\} \tag{7}$$

where $\delta = \|\boldsymbol{x}_0\|$ and $\boldsymbol{u} = \boldsymbol{x}_0/\delta$ is the unit vector in the direction of $\boldsymbol{x}_0$. The constant c'_{F_0} is defined as

$$c'_{F_0} = c_{F_0} \frac{\left(\frac{k-1}{2}\right)\Gamma^2\left(\frac{k-1}{2}\right)}{E_{F_0}[r]\Gamma^2\left(\frac{k}{2}\right)} \tag{8}$$

where r is the norm of k-variate vector $\boldsymbol{y} \sim F_0$ and c_{F_0} has been defined by (3)

Starting from (7) 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*

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

where $D(F_0) = (c^2_{F_0}/k)I_k$, and α and β are two real valued functions, only depending on F_0 and defined as

$$\alpha(\delta) = c^2_{F_0}\delta^{-2}\{\,1 - 2\delta E^{-1}_{F_0}[r]\,\} \qquad \textit{and} \qquad \beta(\delta) = (2k-1) - 2\delta k E^{-1}_{F_0}[r].$$

Note that $\delta^2\alpha$ and β are linear in δ. For $\boldsymbol{x} = \delta \boldsymbol{u}$, the influence function may be rewritten as $c_{F_0}^2\{\,1 - 2\delta E_{F_0}^{-1}[r]\,\}\boldsymbol{u}\boldsymbol{u}^T - \beta(\delta)D(F_0)$, 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. By the affine equivariance property of SCM (cfr. Lemma 2), 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 Σ and let F_0 be the corresponding standardized distribution having mean vector $\mathbf{0}$ and covariance matrix I_k. Then*

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

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

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 Magnus and Neudecker (1988). Now write $\widehat{D}_{ii}$ for a diagonal element of the matrix $\widehat{D}$ and $\widehat{D}_{ij}$, 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 $\sqrt{n}\,\mathrm{vec}(\widehat{D} - D)$ is given by*

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

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

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

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

Notice also that

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

The limiting distribution of the SCM is therefore characterized by 2 numbers: the asymptotic variance of off-diagonal element $\mathrm{ASV}(\widehat{D}_{12}; F_0)$ and the asymptotic variance of an on-diagonal element $\mathrm{ASV}(\widehat{D}_{11}; F_0)$. After some lengthy but straightforward calculations (Lemma 5 is useful here), we obtained

$$\mathrm{ASV}(\widehat{D}_{ij}; F_0) = \frac{c_{F_0}^4 \left\{4kE_{F_0}^{-2}[r] - 3\right\}}{k(k+2)}, \quad i \neq j$$
$$\mathrm{ASV}(\widehat{D}_{ii}; F_0) = \frac{c_{F_0}^4 \left\{4k^2(k^2-1)E_{F_0}^{-2}[r] - 2 + 6k - 4k^3\right\}}{k^2(k+2)}.$$

The above variances (and also the functions α and β of Theorem 4) can be made explicit by calculating $E_{F_0}[r]$ 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 ν degrees of freedom $F_0 = t_\nu$ we have

$$E_\Phi[r] = \frac{\sqrt{2}\,\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \quad \text{and} \quad E_{t_\nu}[r] = \frac{\sqrt{\nu-2}\,\Gamma(\frac{\nu-1}{2})\Gamma(\frac{k+1}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{k}{2})}.$$

In the next section we will study the limit behavior of the eigenvector and standardized eigenvalue estimates based on the Sign Covariance Matrix.

5 Principal Components Analysis based on the SCM

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

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

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, while the inverses of the eigenvalues of $\widehat{D}$ allow to measure the shape of Σ.

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

Theorem 5 *Let F be elliptical distribution with parameters $\boldsymbol{\mu}$ and Σ and let F_0 be the corresponding standardized distribution. The influence functions of the eigenvectors and eigenvalues of D at F are then given by*

$$\mathrm{IF}(\boldsymbol{x}; \boldsymbol{v}_{D,j}, F) = \tilde{\alpha}(d(\boldsymbol{x})) \sum_{\substack{i=1 \\ i \neq j}}^{k} \frac{z_i z_j}{\lambda_j - \lambda_i} \boldsymbol{v}_i \,,$$
$$\mathrm{IF}(\boldsymbol{x}; \lambda_{D,j}, F) = \alpha(d(\boldsymbol{x})) \det(\Sigma)(z_j/\lambda_j)^2 - \beta(d(\boldsymbol{x})) \det(\Sigma)(c^2_{F_0}/k)\lambda_j^{-1},$$

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

As in Croux and Haesbroeck (2000), we can rewrite the influence function for the eigenvectors of the SCM in the form

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

where $\mathrm{IF}(\boldsymbol{x}; \boldsymbol{v}_{\mathrm{Cov},j}, F)$ is the influence function of the eigenvector obtained from the classical covariance matrix estimator Cov, having functional representation

$$\mathrm{Cov}(F) = E_F\left[(\boldsymbol{x} - E_F[\boldsymbol{x}])(\boldsymbol{x} - E_F[\boldsymbol{x}])^T\right] .$$

The influence function for the eigenvectors of the classical eigenvector estimator has already been obtained by Critchley (1985). 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 Figure 2. We also compared with the

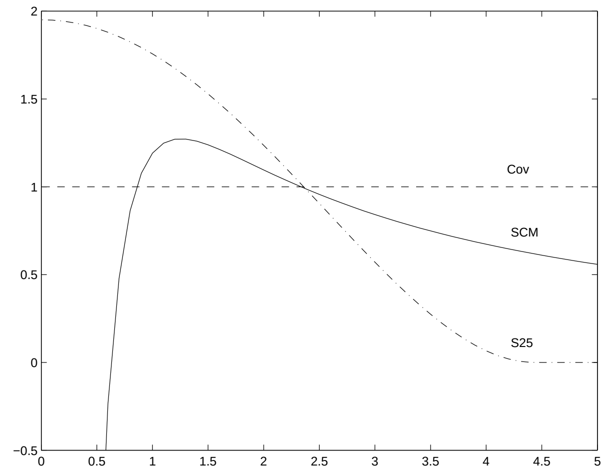


Figure 2: The function $\tilde{\alpha}(\delta)$ for the SCM estimator, the classical covariance matrix estimator and the 25 percent breakdown S-estimator at the bivariate normal model ($F = \Phi$ and $k = 2$).

$\tilde{\alpha}(\delta)$ function of a high breakdown estimator: the multivariate Biweight S-estimator (Rousseeuw and Leroy, 1987, and Davies, 1987), which have already been considered by Croux and Haesbroeck (2000). Note that the $\tilde{\alpha}(\delta)$ for the classical estimator is constant and equal to one. From Figure 2 we see that observations far away from the origin, so for δ large, receive much less weight using SCM instead of the classical estimator. For the high breakdown estimators the downweighting of outliers is much stronger, which render 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.* (1997). They observed that the inlier effect becomes smaller and smaller with increasing k. Note that the influence function for the SCM remains bounded in the neighborhood of the origin.

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

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

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

$$\begin{aligned}\text{ASV}(\widehat{P}_D; F) &= E_F\left[\text{vec}\{\text{IF}(\boldsymbol{x}; P_D, F)\}\text{vec}\{\text{IF}(\boldsymbol{x}; P_D, F)\}^T\right], \\ \text{ASV}(\widehat{\Lambda}_D; F) &= E_F\left[\text{vec}\{\text{IF}(\boldsymbol{x}; \Lambda_D, F)\}\text{vec}\{\text{IF}(\boldsymbol{x}; \Lambda_D, F)\}^T\right]\end{aligned}$$

to calculate the asymptotic covariance matrices.

Corollary 3 *Let F be elliptical distribution with parameters $\boldsymbol{\mu}$ and Σ and let F_0 be the corresponding standardized distribution. Then, $\sqrt{n}\,\text{vec}(\widehat{P}_D - P)$ and $\sqrt{n}\,\text{vec}(\widehat{\Lambda}_D - \Lambda_D)$ has a limiting normal distribution with zero mean and $\widehat{P}_D$ and $\widehat{\Lambda}_D$ are asymptotically independent. The covariance matrix of $\widehat{\boldsymbol{v}}_{D,j}$ and the covariance matrix of $\widehat{\boldsymbol{v}}_{D,i}$ and $\widehat{\boldsymbol{v}}_{D,j}$, $i \neq j$, in the limiting distribution are given by*

$$\begin{aligned}\text{ASV}(\widehat{\boldsymbol{v}}_{D,j}; F) &= \frac{k^2}{c_{F_0}^4}\text{ASV}(\widehat{D}_{12}; F_0)\sum_{\substack{i=1\\ i\neq j}}^{k}\frac{\lambda_i\lambda_j}{(\lambda_j-\lambda_i)^2}\boldsymbol{v}_i\boldsymbol{v}_i^T \\ \text{ASC}(\widehat{\boldsymbol{v}}_{D,i}, \widehat{\boldsymbol{v}}_{D,j}; F) &= \frac{k^2}{c_{F_0}^4}\text{ASV}(\widehat{D}_{12}; F_0)\frac{-\lambda_i\lambda_j}{(\lambda_j-\lambda_i)^2}\,\boldsymbol{v}_j\boldsymbol{v}_i^T,\end{aligned}$$

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

$$\begin{aligned}\text{ASV}(\widehat{\lambda}_{D,j}; F) &= \frac{\det(\Sigma)^2}{\lambda_j^2}\text{ASV}(\widehat{D}_{11}; F_0) \\ \text{ASC}(\widehat{\lambda}_{D,i}, \widehat{\lambda}_{D,j}; F) &= \frac{\det(\Sigma)^2}{\lambda_i\lambda_j}\text{ASC}(\widehat{D}_{11}, \widehat{D}_{22}; F_0),\end{aligned}$$

correspondingly.

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

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

where $\widehat{\text{Cov}}_{12}$ (by symmetry) can be taken as any off-diagonal element of the sample covariance matrix $\widehat{\text{Cov}}$ (e.g. Critchley 1985). This means that the asymptotic efficiency of the estimates $\widehat{\boldsymbol{v}}_{D,j}$ based on the sample SCM relative to the estimates $\widehat{\boldsymbol{v}}_{\text{Cov},j}$ based on the sample covariance matrix at an elliptical distribution F is given by

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

dimension	A. 5	6	8	15	∞	B. 5	6	8	15	∞
	Degrees of freedom					Degrees of freedom				
2	2.000	1.447	1.184	1.031	0.956	0.857	0.904	0.947	0.975	0.956
3	1.960	1.429	1.179	1.038	0.973	0.816	0.873	0.929	0.976	0.973
5	1.905	1.400	1.167	1.040	0.987	0.762	0.827	0.897	0.968	0.987
10	1.843	1.365	1.148	1.036	0.996	0.696	0.768	0.850	0.946	0.996
15	1.816	1.349	1.139	1.032	0.998	0.666	0.739	0.825	0.932	0.998
∞	1.752	1.310	1.114	1.022	1.000	0.584	0.655	0.743	0.865	1.000

Table 1: Asymptotic efficiencies of the SCM eigenvector estimates relative to those based on the sample covariance matrix at t-distribution for several values of the dimension and degrees of freedom. Table B lists the asymptotic efficiencies relative to the MLE.

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 asymptotic relative efficiencies are readily calculable using the formulas of Section 4.2. Table 1A lists the asymptotic relative efficiencies 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. As we can see from Table 1A, 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 list the same asymptotic efficiencies, but now relative 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 eigenvectors will be studied in the next Section, where it will be shown that their relative asymptotic efficiencies are exactly the same as those of the eigenvector estimates.

6 Estimating Covariance and Correlation

The Sign Covariance Matrix allows to estimate shape and orientation of the underlying covariance matrix, but it is also possible to construct an affine equivariant estimator for Σ based on the SCM. (Maronna and Yohai 1998 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

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

and write $\mathrm{C}(\boldsymbol{x}) = \mathrm{C}(G)$ whenever $\boldsymbol{x} \sim G$. Using the equivariance properties of the SCM (see Lemma 1), it follows that C is affine equivariant:

$$\mathrm{C}(A\boldsymbol{x} + \boldsymbol{b}) = A\mathrm{C}(\boldsymbol{x})A^T$$

with A any regular $k \times k$ matrix and for any k-vector $\boldsymbol{b}$. Moreover, by Theorem 1, one then has that

$$\mathrm{C}(F) = \Sigma\,,$$

meaning that C is a Fisher consistent functional for Σ 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 Σ. For example, at the normal model we get as sample estimate

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

with c_Φ defined in (4). Note that also other affine equivariant scatter matrix estimators, including the MCD and multivariate S-estimators, need a scaling factor to attain consistency for Σ 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 follows, after applying some matrix differentiation rules, from Corollary 1:

$$\mathrm{IF}(\boldsymbol{x}_0; \mathrm{C}, F) = \tilde{\alpha}(d(\boldsymbol{x}))(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T - \tilde{\beta}(d(\boldsymbol{x}))\Sigma \tag{9}$$

where $\tilde{\alpha}(d(\boldsymbol{x})) = -(k/c^2_{F_0})\alpha(d(\boldsymbol{x}))$ was already defined in Section 5 and $\tilde{\beta}(d(\boldsymbol{x})) \equiv 1$. In Figure 3 we picture $\mathrm{IF}(\boldsymbol{x}_0; \mathrm{C}_{12}, F_0)$ for a typical off-diagonal element of C and in Figure 4 for a typical on-diagonal element of C, with $F_0 = \Phi$. We compared with the influence functions for the classical estimator, which are also represented. From the Figures, we see that all these 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.

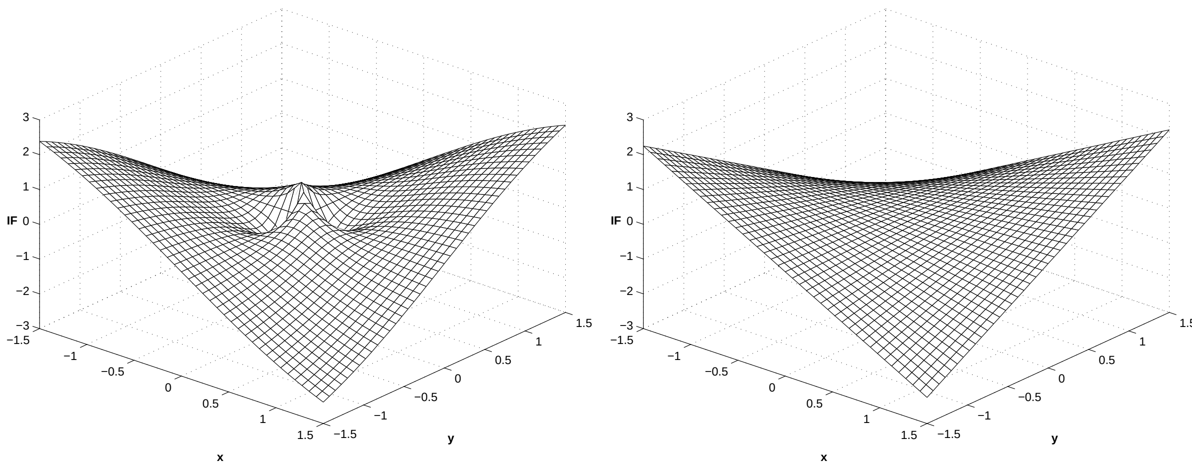


Figure 3: Influence function for an off-diagonal element of the SCM-estimator C (left panel) and for the classical covariance estimator Cov (right panel) at the normal model with $k = 2$.

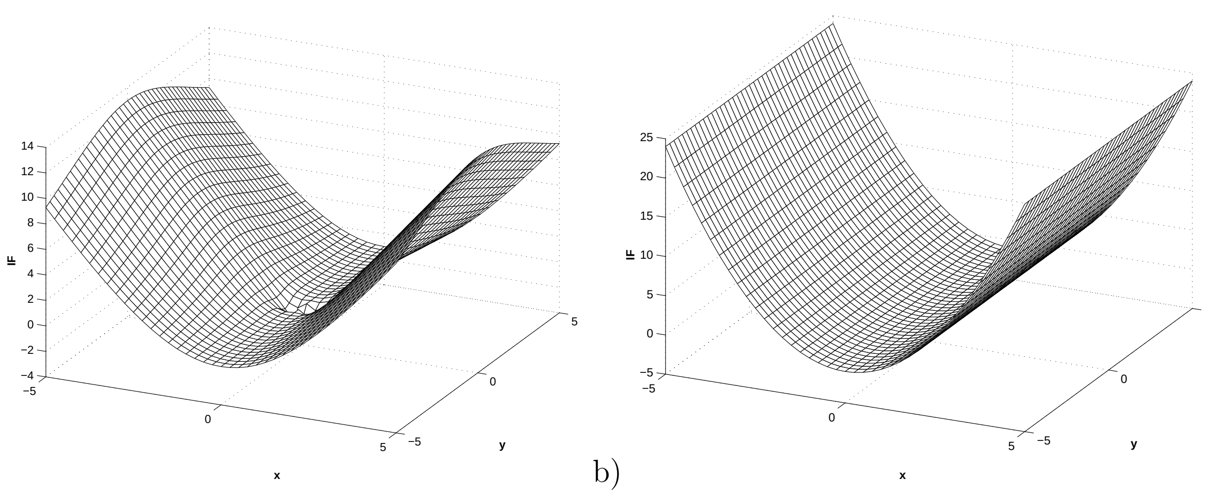


Figure 4: Influence function for the on-diagonal element of the SCM-estimator C (left panel) and for the classical covariance estimator Cov (right panel) at the normal model with $k = 2$.

	A.	Degrees of freedom				B.	Degrees of freedom			
dimension	5	6	8	15	∞	5	6	8	15	∞
2	2.286	1.589	1.250	1.044	0.935	0.857	0.908	0.952	0.974	0.935
3	2.227	1.562	1.243	1.054	0.960	0.795	0.859	0.923	0.974	0.960
5	2.148	1.522	1.225	1.057	0.981	0.716	0.791	0.875	0.961	0.981
10	2.060	1.472	1.198	1.050	0.994	0.625	0.707	0.805	0.928	0.994
15	2.023	1.450	1.185	1.046	0.997	0.585	0.667	0.770	0.907	0.997
∞	1.934	1.396	1.152	1.031	1.000	0.483	0.558	0.658	0.810	1.000

Table 2: Table A lists the asymptotic efficiencies of the on-diagonal element of SCM-estimate $\widehat{C}$ relative to on-diagonal element of the sample covariance matrix $\widehat{\text{Cov}}$ at t-distribution with selected values of dimension and degrees of freedom. Table B lists the corresponding efficiencies relative to MLE.

Similar pictures have been depicted by Croux and Haesbroeck (1999), who also computed asymptotic efficiencies for several estimators of the off- and on-diagonal elements of Σ. For the on-diagonal elements, there is no work to do, since one readily can check that

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

corresponding to the numbers in Table 1. For the off-diagonal elements there are some extra computations to be done. Relative efficiencies for multivariate $t-$ and normal distributions are given in Table 2. One 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 in true when we compare to the MLE.

The influence function of any affine equivariant scatter matrix estimator can be written in the form (9), but of course with different $\tilde{\alpha}$ and $\tilde{\beta}$ (cfr lemma 1 of Croux and Haesbroeck 2000). 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 and Dehon (2001) 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 $\text{C}(F)$ we can in the usual way obtain an estimator $R(F)$ of the population correlation matrix. We write $\widehat{R}$ for the corresponding estimate obtained from $\widehat{\text{C}}$. Note that $\widehat{R}$ can be computed directly from the SCM, since

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

for $1 \leq i, j \leq k$. Since C is an affine equivariant scatter matrix estimator, the influence function of R follows immediately from Lemma 2 of Croux and Haesbroeck (2000):

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

where Corr stands for the classical correlation matrix. Relative asymptotic efficiencies of the estimates of correlation matrix at an elliptical distributions F are therefore, as in Section 5 for the eigenvector estimates, given by

$$\mathrm{ARE}(\widehat{\mathrm{Corr}}_{12}, \widehat{R}_{12}; F) = \mathrm{ARE}(\widehat{\mathrm{Cov}}_{12}, \widehat{\mathrm{C}}_{12}; F_0).$$

The correlation depend both on the orientation and on the shape of the matrix Σ, but their asymptotic relative efficiencies only depend on one number.

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

Lemma 4 *Let* $\widehat{\boldsymbol{\lambda}}_C = (\widehat{\lambda}_{C,1}, \ldots, \widehat{\lambda}_{C,k})^T$ *be the eigenvalue estimates of* any *affine equivariant scatter matrix estimate* $\widehat{C}$ *possessing an influence function and assume that* $\widehat{\boldsymbol{\lambda}}_C$ *is consistent with a limiting normal distribution and asymptotic covariance matrix* $E_F[\mathrm{IF}(\boldsymbol{x}; \boldsymbol{\lambda}_C, F)\mathrm{IF}(\boldsymbol{x}; \boldsymbol{\lambda}_C, F)^T]$. *Let* F *be elliptical distribution with parameters* $\boldsymbol{\mu}$ *and* Σ *and let* F_0 *be the corresponding standardized distribution. Then*

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

with

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

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

Particularly, we get for the SCM

$$\mathrm{ASV}(\ln \widehat{\lambda}^*_{D,j}; F) = \frac{2(k-1)}{k}\mathrm{ASV}(\widehat{\mathrm{C}}_{12}; F_0) = \frac{2k(k-1)}{c^4_{F_0}}\mathrm{ASV}(\widehat{D}_{12}; F_0)\,. \qquad (10)$$

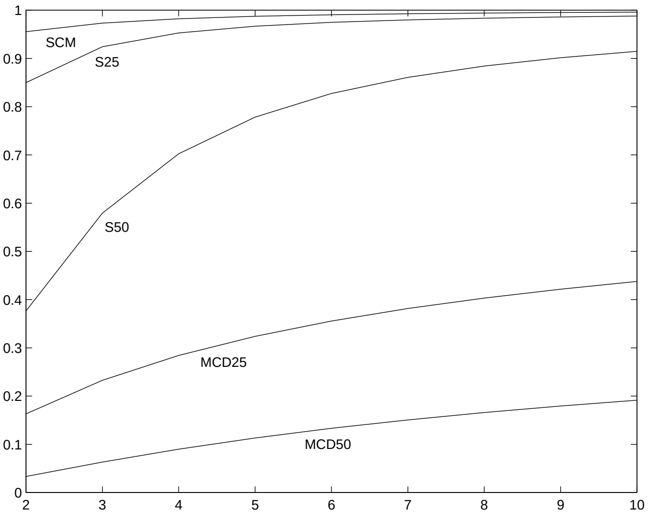


Figure 5: Efficiencies of the standardized eigenvalues in function of the dimension at the normal model for the SCM estimator and 25/50 percent breakdown MCD estimator and biweight S-estimator.

Next, write $\widehat{\lambda}^*_{\text{Cov},j}$, $j = 1, \ldots, k$ for the standardized eigenvalue estimates based on the sample covariance matrix $\widehat{\text{Cov}}$. The asymptotic efficiency of the standardized eigenvalue estimates $\widehat{\lambda}^*_{D,j}$ relative to $\widehat{\lambda}^*_{\text{Cov},j}$ for elliptical F is again given by

$$\text{ARE}(\widehat{\lambda}^*_{\text{Cov},j}, \widehat{\lambda}^*_{D,j}; F) = \frac{\text{ASV}(\ln \widehat{\lambda}^*_{\text{Cov},j}; F)}{\text{ASV}(\ln \widehat{\lambda}^*_{D,j}; F)} = \frac{\text{ASV}(\widehat{\text{Cov}}_{12}; F_0)}{\text{ASV}(\widehat{\text{C}}_{12}; F_0)}, \tag{11}$$

which also equals the asymptotic relative efficiencies $\text{ARE}(\widehat{\boldsymbol{v}}_{\text{Cov},j}, \widehat{\boldsymbol{v}}_{D,j}; F)$ of the eigenvector estimates, and which have already been reported in Table 1. These efficiencies also equal the efficiencies of the SCM regression slope coefficient estimates relative to corresponding estimates based on the LS regression (see Ollila *et al.* 2001).

We compared the efficiencies (11) with those obtained for the *Minimum Covariance Determinant* (MCD) estimate (Rousseeuw, 1985) and and those for the Biweight S-estimate (both with 25% and 50% breakdown point) at the normal model. We refer to Croux and Haesbroeck (1999) and Lopuhäa (1989) for asymptotic properties of the scatter MCD and S-estimators. In Figure 5 we pictured the efficiency of the estimates of the standardized eigenvalues of Σ as a function of the dimension k. We see that the SCM is clearly the most efficient. The S-estimator with 25% is a competitor, but the other estimators seem to result in a too high loss of efficiency.

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 ν degrees of freedom and covariance matrix $\Sigma = \text{diag}(1, \ldots, k)$. Our choices are $k = 2, 3$ and $\nu = 5, 6, 8, 15, \infty$, where $\nu = \infty$ 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 $\arccos\{|\boldsymbol{v}_1^T \widehat{\boldsymbol{v}}_1|\}$ where $\widehat{\boldsymbol{v}}_1$ 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 first eigenvector is then

$$\text{MSE}(\widehat{\boldsymbol{v}}_1) = \frac{1}{m} \sum_{j=1}^{m} (\arccos\{|\boldsymbol{v}_1^T \widehat{\boldsymbol{v}}_1^{(j)}|\})^2$$

where $\widehat{\boldsymbol{v}}_1^{(j)}$ is the estimate for the first eigenvector computed from the jth 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 $\lambda_1^* = \log\{k/(k!^{1/k})\}$, yielding as MSE

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

where $(\widehat{\lambda}_1^*)^{(j)}$ is the estimate for the first standardized eigenvalue computed from the jth 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.

First of all, note that the finite sample efficiencies converge well to the asymptotic ones listed under $n = \infty$. Somewhat slower convergence is seen at $\nu = 5$ showing quite serious loss of efficiency for very small samples (cases $n = 20$ and $n = 50$). This may be due to the fact that for $\nu = 5$ the sample covariance matrix is performing better than what the large-sample efficiency indicates (notice also that $\nu = 5$ is the smallest value of degrees of freedom of the t-distribution for which the sample covariance matrix is asymptotically normal).

		Degrees of freedom				
		5	6	8	15	∞
$k=2$	$n=20$	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

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 ν degrees of freedom and $\Sigma = \text{diag}(1, \ldots, k)$.

In the case of $k = 3$ with $n = 100$ and $n = 300$ 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 change the asymptotics. Computation of a multivariate sign requires enumeration of $O(n^{k-1})$ hyperplanes, and we need to compute n multivariate signs. Therefore it might be too computing intensive to consider all these hyperplanes for higher values of k. In the latter case one could generate a subset of all possible hyperplanes passing the origin, $k-1$ observations and the observation in which we want to compute the sign. Hereby we obtain a stochastic version of the SCM, being tractable also in high dimensions.

Appendix: Proofs and Additional Lemmas

Lemma 5 *For a random vector $\boldsymbol{u} = (u_1, \ldots, u_k)^T$ uniformly distributed on the periphery of the unit sphere, one has that*

$$\textbf{a)}\; E[\boldsymbol{u}\boldsymbol{u}^T] = \frac{1}{k} I_k \quad \textbf{b)}\; E[u_i^2 u_j^2] = \frac{1}{k(k+2)} \quad \textbf{c)}\; E\left[\sqrt{1-u_i^2}\right] = \frac{\Gamma^2(\frac{k}{2})}{\Gamma(\frac{k-1}{2})\Gamma(\frac{k+1}{2})}$$

$$\textbf{d)}\; E[u_i^4] = \frac{3}{k(k+2)} \qquad \textbf{e)}\; E\left[\frac{u_i^2}{\sqrt{1-u_j^2}}\right] = \frac{1}{2}\left[\frac{\Gamma(\frac{k}{2})}{\Gamma(\frac{k+1}{2})}\right]^2 ,$$

where u_i and u_j are distinct elements of $\boldsymbol{u}$. Moreover,

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

where $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k$ are random vectors uniformly distributed on the periphery of the unit sphere.

Proof. Here we only proof item f), items a)-e) are fairly straightforward and left as exercise for the reader. Now let r_i^2, $i = 1, \ldots, k$ be independent random variables from a χ_k^2 distribution. Consequently, $\boldsymbol{x}_i = r_i \boldsymbol{u}_i$, $i = 1, \ldots, k$ 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 $\chi_1^2, \ldots, \chi_k^2$. (cfr. Lemma 1 in Möttönen *et al*, 1998). Thus

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

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

Lemma 6 *At a spherical distribution F_0,*

$$\boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0}) = c_{F_0}\, \boldsymbol{u} \quad \textit{with} \quad c_{F_0} = \frac{\Gamma^k(\frac{k}{2}) E_{F_0}^{k-1}[r]}{\sqrt{\pi}\Gamma^{k-1}(\frac{k+1}{2})} .$$

Proof Let $\boldsymbol{x}_i = r_i \boldsymbol{u}_i,\ i = 1, \ldots, k-1$ be independent observations from F_0 and let $\boldsymbol{x} = \delta \boldsymbol{u}$ be fixed,where $\delta = \|\boldsymbol{x}\|$. Then with the aid of Lemma 5 f),

$$\begin{aligned} E_{F_0}[|\det(\boldsymbol{x}_1 \cdots \boldsymbol{x}_{k-1}\, \boldsymbol{x})|\,\big|\,\boldsymbol{x}] &= \delta E_{F_0}^{k-1}[r] E\big[\,|\det(\boldsymbol{u}_1 \cdots \boldsymbol{u}_{k-1}\, \boldsymbol{u})|\,\big|\,\boldsymbol{u}\,\big] \\ &= \delta \frac{\Gamma^k(\frac{k}{2}) E_{F_0}^{k-1}[r]}{\sqrt{\pi}\Gamma^{k-1}(\frac{k+1}{2})} = \delta\, c_{F_0}, \end{aligned}$$

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

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

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

the transformed vector of cofactors equals $\boldsymbol{e}^*(I; \widehat{\boldsymbol{\theta}}^*) = \det(A)(A^{-1})^T \boldsymbol{e}(I; \widehat{\boldsymbol{\theta}})$. Consequently

$$\mathrm{sign}\{\boldsymbol{e}^*(I; \widehat{\boldsymbol{\theta}}^*)^T (\boldsymbol{x}^* - \widehat{\boldsymbol{\theta}}^*)\} = \mathrm{sign}\{\det(A)\}\mathrm{sign}\{\boldsymbol{e}(I; \widehat{\boldsymbol{\theta}})^T (\boldsymbol{x} - \widehat{\boldsymbol{\theta}})\}.$$

By definition of $\widehat{\boldsymbol{S}_i^*} = \mathrm{ave}_I \big\{\mathrm{sign}\big[\boldsymbol{e}^*(I; \widehat{\boldsymbol{\theta}}^*)^T (\boldsymbol{x}_i^* - \widehat{\boldsymbol{\theta}}^*)\big] \boldsymbol{e}^*(I; \widehat{\boldsymbol{\theta}}^*)\big\}$ and $\widehat{D}^* = \mathrm{ave}_i \left\{\widehat{\boldsymbol{S}_i^*} \big(\widehat{\boldsymbol{S}_i^*}\big)^T\right\}$ the stated expressions follow.

Proof of Theorem 1 It is straightforward to see, using invariance in distribution properties, that

$$E_{F_0}[\boldsymbol{S}_{F_0}(\boldsymbol{z}; \boldsymbol{0})] = \boldsymbol{0} \quad \text{and} \quad D(F_0) = w_{F_0} I_k$$

where w_{F_0} is a positive constant depending on F_0. The affine equivariance property of $\widehat{\boldsymbol{S}}_i$ and $\widehat{D}$ stated in Lemma 1 also hold for the theoretical counterparts and consequently $E_F[\boldsymbol{S}_F(\boldsymbol{x}; T(F))] = \boldsymbol{0}$ 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

$$E_F\big[\, E_F[\mathrm{sign}\{\boldsymbol{e}^T(I)\boldsymbol{x}_i\}\mathrm{sign}\{\boldsymbol{e}^T(J)\boldsymbol{x}_i\}\boldsymbol{e}(I)\boldsymbol{e}^T(J) \,|\, \boldsymbol{x}_i]\,\big] = E_F[\boldsymbol{S}_F(\boldsymbol{x}_i; \boldsymbol{0})\boldsymbol{S}_F(\boldsymbol{x}_i; \boldsymbol{0})^T]\,.$$

So that $E_F[g(K)] = D(F)$.

Next write

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

The statistic V_n can be further decomposed to a weighted sum of $k-1$ U-statistics having kernels with finite expectations and where each weighted term in the sum is $o_p(1/n)$. It follows that $\sqrt{n}\,V_n \xrightarrow{P} 0$ 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

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

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

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

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

$$\begin{aligned}
&\sqrt{n}\,\text{ave}_i\{\boldsymbol{S}_n(\boldsymbol{x}_i;n^{-1/2}\boldsymbol{\theta})\boldsymbol{S}_n^T(\boldsymbol{x}_i;n^{-1/2}\boldsymbol{\theta}) - \boldsymbol{S}_n(\boldsymbol{x}_i)\boldsymbol{S}_n^T(\boldsymbol{x}_i)\} \\
&\qquad \xrightarrow{P} E_F\big\{[\nabla\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{0})]^T\boldsymbol{\theta}\boldsymbol{S}_F^T(\boldsymbol{x};\boldsymbol{0})\big\} + E_F\big\{\boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{0})\boldsymbol{\theta}^T\nabla\boldsymbol{S}_F^T(\boldsymbol{x};\boldsymbol{0})\big\} = 0
\end{aligned}$$

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

Proof of Theorem 2 By writing

$$D(F_\varepsilon) = (1-\varepsilon)E_F\big[\boldsymbol{S}_{F_\varepsilon}\big(\boldsymbol{x};T(F_\varepsilon)\big)\boldsymbol{S}_{F_\varepsilon}^T\big(\boldsymbol{x};T(F_\varepsilon)\big)\big] + \varepsilon\,\boldsymbol{S}_{F_\varepsilon}\big(\boldsymbol{x}_0;T(F_\varepsilon)\big)\boldsymbol{S}_{F_\varepsilon}^T\big(\boldsymbol{x}_0;T(F_\varepsilon)\big)$$

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

$$\text{IF}(\boldsymbol{x}_0, D, F) = -D(F) + \boldsymbol{S}_F(\boldsymbol{x}_0;\boldsymbol{\mu})\boldsymbol{S}_F^T(\boldsymbol{x}_0;\boldsymbol{\mu}) + E_F\left[\frac{\partial}{\partial\varepsilon}\boldsymbol{S}_{F_\varepsilon}\big(\boldsymbol{x};T(F_\varepsilon)\big)\boldsymbol{S}_{F_\varepsilon}^T\big(\boldsymbol{x};T(F_\varepsilon)\big)\Big|_{\varepsilon=0}\right]. \tag{12}$$

Next step is to note that

$$\frac{\partial}{\partial \varepsilon} \boldsymbol{S}_{F_\varepsilon}\big(\boldsymbol{x}; T(F_\varepsilon)\big)\Big|_{\varepsilon=0} = -\nabla_{\boldsymbol{x}} E_F\big[\operatorname{sign}\big\{\boldsymbol{e}^T(I;\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})\big\}\boldsymbol{e}^T(I;\mathbf{0})\big]\operatorname{IF}(\boldsymbol{x}_0; T, F) + \operatorname{IF}(\boldsymbol{x}_0; \boldsymbol{S}_F(\boldsymbol{x};\boldsymbol{\mu}), F)\,.$$

Since $\boldsymbol{S}_F^T(\boldsymbol{x};\boldsymbol{\mu})$ is an odd, $\nabla_{\boldsymbol{x}} E_F\big[\operatorname{sign}\big\{\boldsymbol{e}^T(I;\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})\big\}\boldsymbol{e}^T(I;\mathbf{0})\big]$ is an even function and since F is symmetric,

$$E_F\left[\,\big\{\nabla_{\boldsymbol{x}} E_F\big[\operatorname{sign}\big\{\boldsymbol{e}^T(I;\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})\big\}\boldsymbol{e}^T(I;\mathbf{0})\big]\big\}\ \operatorname{IF}(\boldsymbol{x}_0;T,F)\,\boldsymbol{S}_F^T(\boldsymbol{x};\boldsymbol{\mu})\,\right] = 0\,.$$

Therefore

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

Substituting the above equation in (12) gives the stated result.

Proof of Theorem 3 Lemma 2 and Lemma 3 imply that $\sqrt{n}\,(\widehat{D}-U_n)\xrightarrow{p}0$. 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 7 *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 equation* (7)*.*

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

$$E_{F_0}[\operatorname{sign}\{\boldsymbol{e}^T(I)\boldsymbol{x}\}\boldsymbol{e}(I)|\boldsymbol{x}_{i_1}=P\boldsymbol{x}_0] = PE_{F_0}[\operatorname{sign}\{\boldsymbol{e}^T(I)P^T\boldsymbol{x}\}\boldsymbol{e}(I)|\boldsymbol{x}_{i_1}=\boldsymbol{x}_0]. \qquad (13)$$

First consider the special case $\boldsymbol{x}_0=\delta\boldsymbol{u}_0=\delta(1,0,\ldots,0)^T$. Following the proof of Lemma 6, it can be shown that

$$E[\operatorname{sign}\{\boldsymbol{e}^T(I)\boldsymbol{x}\}\boldsymbol{e}(I)|\boldsymbol{x}_{i_1}=\boldsymbol{x}_0] = \delta c'_{F_0}\frac{(I_k-\boldsymbol{u}_0\boldsymbol{u}_0^T)\boldsymbol{x}}{\|(I_k-\boldsymbol{u}_0\boldsymbol{u}_0^T)\boldsymbol{x}\|} \qquad (14)$$

with

$$c'_{F_0} = \frac{\Gamma^{k-1}(\frac{k-1}{2})E_{F_0}^{k-2}[r']}{\sqrt{\pi}\Gamma^{k-2}(\frac{k}{2})} \qquad (15)$$

where r' is a length of $(k-1)$-variate subvector of $\boldsymbol{y} = r\boldsymbol{u} \sim F_0$. So we may set $r' = \sqrt{r^2 - y_1^2} = r\sqrt{1-u_1^2}$. By using the relation $E_{F_0}[r'] = E_{F_0}[r]E[\sqrt{1-u_1^2}]$ together with Lemma 5 c), and equation (3) for c_{F_0} we get

$$c'_{F_0} = c_{F_0} \frac{\left(\frac{k-1}{2}\right)\Gamma^2(\frac{k-1}{2})}{E_{F_0}[r]\Gamma^2(\frac{k}{2})}.$$

Next consider the general case $\boldsymbol{x}_0 = \delta\boldsymbol{u}$ and let P be a rotation matrix ($PP^T = I_k$) such that $P\boldsymbol{u}_0 = \boldsymbol{u}$. Then equation (13) together with equation (14) imply that

$$E_{F_0}[\text{sign}\{\boldsymbol{e}^T(I)\boldsymbol{x}\}\boldsymbol{e}(I)|\boldsymbol{x}_{i_1} = \boldsymbol{x}_0] = \delta c'_{F_0} \frac{P(I_k - \boldsymbol{u}_0\boldsymbol{u}_0^T)P^T\boldsymbol{x}}{\|(I_k - \boldsymbol{u}_0\boldsymbol{u}_0^T)P^T\boldsymbol{x}\|} = \delta c'_{F_0} \frac{(I_k - \boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{x}}{\|(I - \boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{x}\|}$$

which, by using equation (5), 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 $\boldsymbol{v} = (1, 0, \ldots, 0)^T$. By Theorem 2,

$$\begin{aligned}\text{IF}(\boldsymbol{x}_\delta; D, F_0) = \boldsymbol{S}_{F_0}(\boldsymbol{x}_\delta; \boldsymbol{0})\boldsymbol{S}_F^T(\boldsymbol{x}_\delta; \boldsymbol{0}) + E_{F_0}\left[\text{IF}(\boldsymbol{x}_\delta; \boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0}), F_0)\, \boldsymbol{S}_{F_0}^T(\boldsymbol{x}; \boldsymbol{0})\right] \\ + E_{F_0}\left[\boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0})\, \text{IF}(\boldsymbol{x}_\delta; \boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0}), F_0)^T\right] - D(F_0)\,, \quad (16)\end{aligned}$$

since $\boldsymbol{\mu} = T(F_0) = \boldsymbol{0}$. Then use $\boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0}) = c_{F_0}\boldsymbol{u}$ with $\boldsymbol{u} = \boldsymbol{x}\,\|\boldsymbol{x}\|^{-1}$, $D(F_0) = E_{F_0}[\boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0})\boldsymbol{S}_{F_0}^T(\boldsymbol{x}; \boldsymbol{0})]$ together with equation (7) to obtain

$$\begin{aligned}&E_{F_0}\left[\text{IF}(\boldsymbol{x}_\delta; \boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0}), F_0)\, \boldsymbol{S}_{F_0}^T(\boldsymbol{x}; \boldsymbol{0})\right] \\ &= (k-1)E_{F_0}\left[\left\{\delta c'_{F_0} \frac{(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{x}}{\|(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{x}\|} - \boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0})\right\} \boldsymbol{S}_{F_0}^T(\boldsymbol{x}; \boldsymbol{0})\right] \\ &= (k-1)\delta c_{F_0} c'_{F_0} E_{F_0}\left[\frac{(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{x}\boldsymbol{x}^T}{\|(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{x}\|\,\|\boldsymbol{x}\|}\right] - (k-1)D(F_0)\,. \qquad (17)\end{aligned}$$

By noticing that

$$E_{F_0}\left[\frac{(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{x}\boldsymbol{x}^T}{\|(I_k - \boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{x}\|\,\|\boldsymbol{x}\|}\right] = E\left[\frac{u_2^2}{\sqrt{1-u_1^2}}\right]\begin{pmatrix}0 & \boldsymbol{0}^T \\ \boldsymbol{0} & I_{k-1}\end{pmatrix}$$

and by substituting equation (8) for c'_{F_0} and using Lemma 5 e), equation (17) simplifies to

$$E_{F_0}\left[\text{IF}(\boldsymbol{x}_\delta; \boldsymbol{S}_{F_0}(\boldsymbol{x}; \boldsymbol{0}), F_0)\, \boldsymbol{S}_{F_0}^T(\boldsymbol{x}; \boldsymbol{0})\right] = \delta\, c_{F_0}^2 E_{F_0}^{-1}[r]\begin{pmatrix}0 & \boldsymbol{0}^T \\ \boldsymbol{0} & I_{k-1}\end{pmatrix} - (k-1)D(F_0)\,.$$

Hence we may now write (16) as

$$\text{IF}(\boldsymbol{x}_\delta; D, F_0) = c_{F_0}^2\begin{pmatrix}1 & \boldsymbol{0}^T \\ \boldsymbol{0} & 2\delta\, E_{F_0}^{-1}[r]\, I_{k-1}\end{pmatrix} - (2k-1)D(F_0)\,. \qquad (18)$$

An influence point in an arbitrary direction is obtained by setting $\boldsymbol{x} = P\boldsymbol{x}_\delta = \delta \boldsymbol{p}_1$ for a well chosen rotation matrix $P = [\boldsymbol{p}_1 \cdots \boldsymbol{p}_k]$ with $P^T P = I_k$. The influence function is then given by

$$\mathrm{IF}(\boldsymbol{x}; D, F_0) = P\,\mathrm{IF}(\boldsymbol{x}_\delta; D, F_0)\,P^T\,,$$

which, by (18) and some simple matrix manipulation, can be written as

$$\begin{aligned} IF(\boldsymbol{x}; D, F_0) &= c_{F_0}^2 \delta^{-2} \big\{\, 1 - 2\delta\, E_{F_0}^{-1}[r] \,\big\} \boldsymbol{x}\boldsymbol{x}^T - \big\{\, (2k-1) - 2\delta k E_{F_0}^{-1}[r] \,\big\} D(F_0) \\ &= \alpha(\delta)\boldsymbol{x}\boldsymbol{x}^T - \beta(\delta) D(F_0)\,. \end{aligned}$$

Proof of Corollary 1 Affine equivariance of D yields

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

Applying Theorem 4 yields then stated expression.

Proof of Theorem 6 The proof follows from Theorem 3 the fact that, $D \to (P, \Lambda)$ is a bijection and has nonzero differentials in a neighborhood of the true value (note that we assumed distinct eigenvalues). See Theorem 3.3.A in Serfling (1980) or Theorem 13.5.1 in Anderson (1984).

Proof of Theorem 5 Lemma 3 in Croux and Haesbroeck (2000) 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

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

By Corollary 1 one has that

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

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

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

$$\begin{aligned} \mathrm{ASV}(\widehat{\lambda}_{D,j}; F) &= E_F[\mathrm{IF}(\boldsymbol{x}; \lambda_{D,j}, F)^2] \\ &= (\det(\Sigma)/\lambda_j)^2 E_F\left[\left\{\alpha(d(\boldsymbol{x}))(z_j/\sqrt{\lambda_j})^2 - \beta(d(\boldsymbol{x}))(c_{F_0}^2/k)\right\}^2\right], \end{aligned}$$

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

With $u_j = z_j/\sqrt{\lambda_j}$, one has that $\boldsymbol{u} = (u_1, \ldots, u_k)^T \sim F_0$ and $d(\boldsymbol{x}) = \|\boldsymbol{u}\|$. This yields

$$\begin{aligned}\mathrm{ASV}(\widehat{\lambda}_{D,j}; F) &= (\det(\Sigma)/\lambda_j)^2 E_{F_0}\left[\{\alpha(\|\boldsymbol{u}\|)u_j^2 - \beta(\|\boldsymbol{u}\|)(c_{F_0}^2/k)\}^2\right] \\ &= (\det(\Sigma)/\lambda_j)^2 E_{F_0}\left[\mathrm{IF}(\boldsymbol{u}; D_{jj}, F_0)^2\right] = (\det(\Sigma)/\lambda_j)^2 \mathrm{ASV}(\widehat{D}_{11}; F_0)\end{aligned}$$

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

For the eigenvector estimator, the asymptotic variance is given by

$$\begin{aligned}\mathrm{ASV}(\widehat{\boldsymbol{v}}_{D,j}; F) &= E_F[\mathrm{IF}(\boldsymbol{x}; \boldsymbol{v}_{D,j}, F)\mathrm{IF}(\boldsymbol{x}; \boldsymbol{v}_{D,j}, F)^T] \\ &= \sum_{\substack{i=1\\ i\neq j}}^{k} \sum_{\substack{l=1\\ l\neq j}}^{k} \frac{1}{\lambda_j - \lambda_i} \frac{1}{\lambda_j - \lambda_l} \frac{k^2}{c_{F_0}^4} E_F\left[\alpha(d(\boldsymbol{x}))^2 z_i z_l z_j^2\right] \boldsymbol{v}_i \boldsymbol{v}_l^T\,. \qquad (19)\end{aligned}$$

Using the transformation $z_j/\sqrt{\lambda_j} = u_j$, the expectation in (19) is simply

$$\begin{aligned}E_F\left[\alpha^2(d(\boldsymbol{x}))z_i z_l z_j^2\right] &= \sqrt{\lambda_i}\sqrt{\lambda_l}\lambda_j E_{F_0}\left[\alpha(\|\boldsymbol{u}\|)^2 u_i u_l u_j^2\right] = \lambda_i\lambda_j E_{F_0}\left[\alpha(\|\boldsymbol{u}\|)^2 u_i^2 u_j^2\right]\delta_{il} \\ &= \lambda_i\lambda_j E_{F_0}\left[\mathrm{IF}(\boldsymbol{u}; D_{ij}, F_0)^2\right]\delta_{il} = \lambda_i\lambda_j \mathrm{ASV}(\widehat{D}_{12}; F_0)\delta_{il}\end{aligned}$$

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

Proof of Corollary 4 Under the stated assumption, we may write $\sqrt{n}\,(\widehat{\boldsymbol{\lambda}}_C - \boldsymbol{\lambda}) \xrightarrow{d} N(\mathbf{0}, B)$, where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)^T$ and the diagonal elements of the asymptotic variance-covariance matrix $B = E_F[\mathrm{IF}(\boldsymbol{x}; \boldsymbol{\lambda}_C, F)\mathrm{IF}(\boldsymbol{x}; \boldsymbol{\lambda}_C, F)^T]$ are

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

for $j = 1, \ldots, k$ (Corollary 1 of Croux and Haesbroeck, 2000). It is easy to derive the expression for the off-diagonal elements (limiting covariances)

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

for $1 \leq i \neq j \leq n$.

Write

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

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

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

After some easy matrix algebra we may write

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

It is not difficult to find out, using e.g. the general expression (9) for the influence function of any affine equivariant scatter matrix estimator, that $\mathrm{ASV}(\widehat{C}_{11}; F_0) - \mathrm{ASC}(\widehat{C}_{11}, \widehat{C}_{22}; F_0) = 2\mathrm{ASV}(\widehat{C}_{12}; F_0)$. Hence one has that

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

which completes the proof.

References

Anderson, T.W. (1984) *An Introduction to Multivariate Statistical Analysis.* Second edition. Wiley, New York.

Bensmail, H. and Celeux, G. (1996). Regularized gaussian discriminant analysis through eigenvalue decomposition. *J. Am. Statist. Ass.*, **91**, 1743-1749.

Brown, B.M., Hall, P. and Young G.A. (1997). On the effect of inliers on the spatial median. *J. Multiv. Anal.*, **63**, 88-104.

Critchley, F. (1985). Influence in principal component analysis. *Biometrika*, **72**, 627-636.

Croux, C and Haesbroeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *J. Multiv. Anal.*, **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.

Croux, C and Dehon, C. (2001). Robust linear discriminant analysis using S-estimators. To appear in *Canadian J. Statist.*, vol 29.

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

Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1981). Robust estimation of dispersion matrices and principal components. *J. Am. Statist. Ass.*, **76**, 354-362.

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., Nyblom, J. and Oja, H. (1994). Affine invariant multivariate one-sample sign tests. *J. Roy. Statist. Soc. Ser. B*, **56** , 221-234.

Hettmansperger, T.P. and Oja, H. (1994). Affine invariant multivariate multisample sign tests. *J. Roy. Statist. Soc. Ser. B*, **56** , 235-249.

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

Magnus, J.R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley and Sons, New York.

Maronna, R.A. and Yohai, V. (1998). Robust estimation of multivariate location and scatter. In *Encyclopedia of Statistical Sciences Update Volume 2*, eds. S. Kotz, C. Read, and D. Banks. Wiley, New York, 589-596.

Möttönen, J., Hettmansperger, T.P., Oja, H. and Tienari, J. (1998). On the efficiency of affine invariant multivariate rank tests. *J. Multiv. Statist.*, **66**, 108-132.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Stat. and Prob. Letters*, **1**, 327-332.

Oja, H. (1999). Affine invariant multivariate sign and rank tests and corresponding estimates: a review. *Scand. J. Statist.*, **26**, 319-343.

Ollila, E., Hettmansperger, T. P. and Oja, H. (2001). Estimates of regression coefficients based on the sign covariance matrix. Conditionally accepted.

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 Leroy, A.M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Visuri, S., Koivunen, V. and Oja, H. (2000). Sign and rank covariance matrices. *J. Statist. Plann. Inference,* **91**, 557-575.