Robustifying principal component analysis with spatial sign vectors

Sara Taskinen[a,b,*], Inge Koch[c], Hannu Oja[d]

[a] *Department of Mathematics and Statistics, University of Jyväskylä*
[b] *School of Mathematics and Statistics, University of New South Wales*
[c] *School of Mathematical Sciences, University of Adelaide*
[d] *Tampere School of Public Health, University of Tampere*

Abstract

In this paper, we apply orthogonally equivariant spatial sign covariance matrices as well as their affine equivariant counterparts in principal component analysis. The influence functions and asymptotic covariance matrices of eigenvectors based on robust covariance estimators are derived to compare the robustness and efficiency properties. We show that especially the estimators that use pairwise differences of the observed data have very good efficiency properties providing practical robust alternatives to classical sample covariance matrix based methods.

Keywords: Affine equivariance, efficiency, influence function, robustness, spatial sign vector

1. Introduction

Principal component analysis (PCA) is widely used for finding lower dimensional structure in data, and is commonly applied to high-dimensional data. PCA represents data by a smaller number of components that account for the variability in the data. This dimension reduction step can be followed by other multivariate methods, such as regression, discriminant analysis, cluster analysis, *etc.*

In classical PCA the sample mean and the sample covariance matrix are used to derive the principal components. These two estimators are highly sen-

*Corresponding author: Sara Taskinen, Department of Mathematics and Statistics, FI-40014 University of Jyväskylä, Tel. +358 14 260 2991, Email: sara.l.taskinen@jyu.fi

sitive to outlying observations, and classical PCA becomes unreliable, when outliers are encountered. Several robust competitors to classical PCA estimators have been proposed in the literature. A natural way to robustify PCA is to use robust location and scatter estimators instead of the sample mean and sample covariance matrix when estimating the eigenvalues and eigenvectors of the population covariance matrix. See *e*.g. Devlin et al. (1981).

Locantore et al. (1999) and Marden (1999) proposed a simple robust alternative to PCA. In spherical PCA, the data are first transformed to multivariate spatial sign vectors. These are simply defined to be the vectors of unit length that point into the direction of original data points. The sample covariance matrix of spatial sign vectors, that is, the spatial sign covariance matrix (Visuri et al., 2000), is then used to obtain the eigenvectors of interest. As spherical PCA uses only information on directions of the data points, its efficiency properties are poor especially under the multinormal model. To obtain a more efficient estimator, the symmetrised spatial sign covariance matrix (Visuri et al., 2000) can be used. This estimator is defined as the spatial sign covariance matrix based on pairwise differences of the observed data. The use of symmetrised spatial sign covariance matrix for PCA was discussed in Visuri et al. (2000), but its theoretical properties have not yet been studied.

In this paper, we compare different spatial sign covariance matrices as well as their affine equivariant counterparts, namely, the Tyler's shape matrix (Tyler, 1987) and the Dümbgen's shape matrix (Dümbgen, 1998) in a PCA framework. We assume that the number of observations in data is greater than the number of variables, so $n > p$. As shown by Tyler (2010), the estimation of robust affine equivariant scatter matrices in case $p > n$ is meaningless as resulting scatter matrices will be proportional to the sample covariance matrix. Notice that the spatial sign covariance matrices are not affine equivariant, and can be easily computed for data with more variables than observations. Applications are given for example in Locantore et al. (1999) and Gervini (2008). The paper is organised as follows. In Section 2, PCA based on different spatial sign covariance matrices is considered. In Section 3, influence functions and asymptotic efficiencies are studied. We show that symmetrised estimators, in particular, have very good efficiency properties thereby providing practical and robust alternatives to classical sample covariance matrix based methods. The paper concludes with some final remarks in Section 4.

2. PCA based on sign covariance matrices

2.1. Notation

Let $\boldsymbol{x}$ be a random vector from a p-variate distribution with symmetry center $\boldsymbol{\mu}$ and positive definite and symmetric $p \times p$ ($PDS(p)$) covariance matrix Σ. The spectral decomposition of the population covariance matrix will be denoted by $\Sigma = \Gamma\Lambda\Gamma^T$, where Γ is an orthogonal matrix containing the eigenvectors $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p$ of Σ as columns and Λ is a diagonal matrix with corresponding eigenvalues $\lambda_1 > \cdots > \lambda_p > 0$ as diagonal values. We may also write

$$\Sigma = [Det(\Sigma)]^{1/p}\, \Gamma\Lambda_\Sigma\Gamma^T,$$

where $Det(\Sigma)$ is the Wilks' generalized variance, and Λ_Σ is a diagonal matrix of standardized eigenvalues

$$\lambda_{\Sigma,i} = \frac{\lambda_i}{(\lambda_1 \cdots \lambda_p)^{1/p}}. \tag{1}$$

The geometric mean of the standardized eigenvalues thus equals one. The parameters $[Det(\Sigma)]^{1/p}$, Λ_Σ and Γ define the scale, shape and orientation of the data cloud. See for example Paindaveine (2008) and Frahm (2009).

Let F_x be the cumulative distribution function of $\boldsymbol{x}$. Then $C(F_x)$ is a scatter matrix functional if it is $PDS(p)$ and affine equivariant in the sense that

$$C(F_{Ax+b}) = A\, C(F_x)\, A^T, \tag{2}$$

where A is any nonsingular $p \times p$ matrix and $\boldsymbol{b}$ is any p-vector. We then write

$$C(F_x) = P_C \Lambda_C P_C^T \tag{3}$$

for the spectral decomposition of $C(F_x)$, where $P_C = P_C(F_x)$ is an orthogonal matrix that contains the eigenvectors $\boldsymbol{p}_{C,1}, \ldots, \boldsymbol{p}_{C,p}$ of $C(F_x)$ as columns, and $\Lambda_C = \Lambda_C(F_x)$ is a diagonal matrix of the corresponding eigenvalues arranged in decreasing order $\lambda_{C,1} > \cdots > \lambda_{C,p}$.

We call $V(F_x)$ a shape matrix functional if it is $PDS(p)$ with $Det(V) = 1$ and affine equivariant in the sense that

$$V(F_{Ax+b}) = \{Det[A(V(F_x))A^T]\}^{-1/p}\, A\, V(F_x)\, A^T. \tag{4}$$

The spectral decomposition of $V(F_x)$ can be written in a similar fashion to that of $C(F_x)$. For a detailed discussion of scatter matrix and shape matrix functionals, see Oja (2010) and references therein.

Finally, notice that in robustness and efficiency studies we assume the elliptical model F, that is, the density function of a random vector $\boldsymbol{x} \sim F$ is of the form

$$f(\boldsymbol{x}) = |\Sigma|^{-1/2} f_0(\Sigma^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})), \tag{5}$$

where $f_0(\boldsymbol{z}') = \exp\{-\rho(||\boldsymbol{z}'||)\}$ and function ρ defines the distribution of the standardised variable $\boldsymbol{z}' = \Sigma^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})$ (in this paper, the square root matrices $\Sigma^{-1/2}$ and $\Sigma^{1/2}$ are taken to be symmetric). To fix Σ, we assume that ρ is such that $E[||\boldsymbol{z}'||^2] = p$. Notice also that under the elliptical model, different shape matrices estimate the same population quantity $V = [\det(\Sigma)]^{-1/p}\,\Sigma$. The orthogonal transformation $\boldsymbol{z} = \Gamma^T(\boldsymbol{x} - \boldsymbol{\mu})$ defines variables from a centered distribution with covariance matrix Λ. Throughout the paper this distribution is denoted by F_Λ.

2.2. PCA based on spatial sign covariance matrices

Locantore et al. (1999) and Marden (1999) introduced a simple robust alternative to classical principal component analysis based on spatial sign vectors. Write the spatial sign function as

$$\boldsymbol{S}(\boldsymbol{x}) = \begin{cases} \|\boldsymbol{x}\|^{-1}\boldsymbol{x}, & \boldsymbol{x} \neq \boldsymbol{0} \\ \boldsymbol{0}, & \boldsymbol{x} = \boldsymbol{0}, \end{cases}$$

where $||\boldsymbol{x}||$ denotes the Euclidean norm of $\boldsymbol{x}$. The spatial sign function thus projects data points onto the unit sphere making resulting estimators highly robust. The population spatial sign covariance matrix is defined as

$$SCov(F_x) = E\left[\boldsymbol{S}(\boldsymbol{x} - \boldsymbol{\mu}(F_x))\boldsymbol{S}^T(\boldsymbol{x} - \boldsymbol{\mu}(F_x))\right], \tag{6}$$

where $\boldsymbol{\mu}(F_x)$ is the population spatial median that solves $E[\boldsymbol{S}(\boldsymbol{x}-\boldsymbol{\mu}(F_x))] = \boldsymbol{0}$. Recall that the spatial median is a higly robust rotation equivariant estimator of symmetry center $\boldsymbol{\mu}$. It has a 50% breakdown point and a bounded influence function. A fast algorithm for computing spatial medians is given in Vardi and Zhang (2001).

In spherical principal component analysis, the matrix of eigenvectors, $P_S = P_S(F_x)$, and the diagonal matrix $\Lambda_S = \Lambda_S(F_x)$ of the eigenvalues of $SCov(F_x)$ are computed as in (3). The robust principal component scores are then given by $\boldsymbol{z} = {P_S}^T(\boldsymbol{x} - \boldsymbol{\mu}(F_x))$.

Croux et al. (2002) examined efficiency properties of the eigenvectors based on the spatial sign covariance matrix, and showed that especially for multivariate normal models, the efficiencies are very low. This is due to the fact that the spatial sign covariance matrix only uses information on the directions of the data points. Motivated by these results, in the following we consider principal component analysis based on more efficient covariance estimators.

Define the population symmetrised spatial sign covariance matrix as

$$SSCov(F_x) = E\left[\boldsymbol{S}(\boldsymbol{x}_1 - \boldsymbol{x}_2)\boldsymbol{S}(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T\right] \tag{7}$$

(Visuri et al., 2000). Since $SSCov(F_x)$ is based on pairwise differences, it is computed without the need for the location functional. Via pairwise differences one also uses information on distances of the data points from the center of data. This feature makes the symmetrised spatial sign covariance matrix more efficient than the spatial sign covariance matrix. The eigenvectors and eigenvalues are now obtained from the spectral decomposition of $SSCov(F_x)$ and denoted by $P_{SS} = P_{SS}(F_x)$ and $\Lambda_{SS} = \Lambda_{SS}(F_x)$. The principal component scores can be computed as before using *e*.g. spatial median.

The next theorem summarizes the relationship between the eigenvectors and eigenvalues of the population covariance matrices (6) and (7) with those of Σ. A proof is given in the Appendix.

Theorem 1. *Let F be an elliptical distribution with location center $\boldsymbol{\mu}$ and covariance matrix $\Sigma = \Gamma\Lambda\Gamma^T$. Then*

$$SCov(F_x) = \Gamma\Lambda_S\Gamma^T \quad \textit{and} \quad SSCov(F_x) = \Gamma\Lambda_S\Gamma^T,$$

where

$$\Lambda_S = E\left[\frac{\Lambda^{1/2}\boldsymbol{u}\boldsymbol{u}^T\Lambda^{1/2}}{\boldsymbol{u}^T\Lambda\boldsymbol{u}}\right] = diag(\lambda_{S,1}, \ldots, \lambda_{S,p}), \tag{8}$$

$$\lambda_{S,i} = E\left[\frac{\lambda_{\Sigma,i}u_i^2}{\lambda_{\Sigma,1}u_1^2 + \cdots + \lambda_{\Sigma,p}u_p^2}\right], \tag{9}$$

$\boldsymbol{u} = (u_1, \ldots, u_p)^T$ is uniformly distributed on the unit sphere, and $\lambda_{\Sigma,i}$s are the standardised eigenvalues of Σ as defined in (1).

Under the elliptical model, the eigenvectors of the population spatial sign covariance matrices therefore coincide with those of Σ and the eigenvalues are related via (9). The relationship between the eigenvalues implies that it is possible to obtain an algorithm for computing estimates for Λ_Σ starting from Λ_S: Let $\Lambda_\Sigma^{(0)}$ be some initial value for Λ_Σ, *e.g.* $\Lambda_\Sigma^{(0)} = I_p$. Then the iteration step suggested by (8) is

$$\Lambda_\Sigma^{(k+1)} \leftarrow \left[E\left(\frac{\boldsymbol{u}\boldsymbol{u}^T}{\boldsymbol{u}^T \Lambda_\Sigma^{(k)} \boldsymbol{u}} \right) \right]^{-1} \Lambda_S.$$

The result is scaled so that $\det(\Lambda_\Sigma^{(k+1)}) = 1$. In practice, an estimate for Λ_Σ is obtained by replacing the above expectation by the sample average.

In Visuri et al. (2000) statistical properties of spatial sign covariance matrices were studied. Corresponding matrices are not genuine scatter matrices as they satisfy (2) only when A is an orthogonal $p \times p$ matrix. However, in the context of PCA, this orthogonal equivariance property suffices, since it yields principal component scores that are invariant under orthogonal and location transformations. The main advantage of spatial sign covariance matrices is that they are very easy and fast to compute in practice.

2.3. PCA based on affine equivariant spatial sign covariance matrices

Hettmansperger and Randles (2002) proposed affine equivariant counterparts of spatial median and spatial sign covariance matrix. In their approach, the original data points are first transformed to $\boldsymbol{y} = V(F_x)^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}(F_x))$, where $\boldsymbol{\mu}(F_x)$ is a p-vector and $V(F_x)$ is a shape matrix. Then the affine equivariant counterparts of the spatial median and spatial sign covariance matrix are those $\boldsymbol{\mu}(F_x)$ and $V(F_x)$ that are found as the solutions of the following M-estimation equations

$$E[\boldsymbol{S}(\boldsymbol{y})] = \boldsymbol{0} \quad \text{and} \quad p\, E[\boldsymbol{S}(\boldsymbol{y})\boldsymbol{S}(\boldsymbol{y})^T] = I_p. \tag{10}$$

The resulting functional $V(F_x)$ is known as Tyler's (1987) shape matrix and the location functional $\boldsymbol{\mu}(F_x)$ is the transformation retransformation spatial median using Tyler's shape matrix (Hettmansperger and Randles, 2002). In the following we will denote $V(F_x) = SCov_2(F_x)$.

The affine equivariant counterpart of the symmetrised spatial sign covariance matrix is obtained similarly to the above: this time we use pairwise differences of transformed observations, that is, $V(F_x)$ is chosen to satisfy

$$p\, E[\boldsymbol{S}(\boldsymbol{y}_1 - \boldsymbol{y}_2)\boldsymbol{S}(\boldsymbol{y}_1 - \boldsymbol{y}_2)^T] = I_p. \tag{11}$$

The resulting functional $V(F_x)$ is known as Dümbgen's (1998) shape matrix, which we denote by $SSCov_2(F_x)$. Note that, since Dümbgen's estimator uses differences of transformed observations, it avoids estimation of the location parameter.

As in (3), the spectral decomposition of $V(F_x)$ can be used to compute the eigenvectors and eigenvalues of different shape matrices, and the robust principal component scores. It then follows from (4) that under the elliptical model,

$$SCov_2(F_x) = \Gamma \Lambda_\Sigma \Gamma^T$$

and similarly for $SSCov_2(F_x)$. The eigenvectors and eigenvalues of $SCov_2(F_x)$ and $SSCov_2(F_x)$ therefore coincide with the eigenvectors and standardised eigenvalues of Σ as defined in (1).

Robustness and asymptotic properties of affine equivariant spatial sign covariance matrices were studied in Tyler (1987), Dümbgen (1998), Dümbgen and Tyler (2005) and Sirkiä et al. (2007) among others. Corresponding estimates can be computed using simple iterative algorithms based on estimating equations (Hettmansperger and Randles, 2002; Oja and Randles, 2004). In the next section we compare different spatial covariance matrices in a PCA framework.

3. Robustness and efficiency properties of eigevectors

3.1. Influence functions

The influence function measures the robustness of a functional T against a single outlier $\boldsymbol{x}$ (Hampel et al., 1986). Let $F_\epsilon = (1-\epsilon)F + \epsilon\Delta_{\boldsymbol{x}}$ denote a contaminated distribution, where $\Delta_{\boldsymbol{x}}$ is the cdf of a distribution with probability one at point $\boldsymbol{x}$. The influence function of T is defined by

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

A continuous and bounded influence function indicates good local robustness properties of an estimator.

Write $C(F_x)$ for a scatter matrix functional decomposed as in (3). Croux and Haesbroeck (2000) showed that under an elliptical distribution F with distinct eigenvalues of the covariance matrix Σ, the influence function of the

eigenvector functional $\boldsymbol{p}_{C,j}$ is

$$IF(\boldsymbol{x}; \boldsymbol{p}_{C,j}, F) = \sum_{\substack{k=1 \\ k \neq j}}^{p} \frac{1}{\lambda_j - \lambda_k} \left[\boldsymbol{\gamma}_k^T IF(\boldsymbol{x}; C, F) \boldsymbol{\gamma}_j \right] \boldsymbol{\gamma}_k, \tag{12}$$

where $\boldsymbol{\gamma}_k$ is the k-th eigenvector of Σ as defined in Section 2.1. Influence functions in case of multiple multiple roots are studied in Tanaka (1988).

To derive influence functions for eigenvectors based on orthogonally equivariant $SCov(F_x)$ and $SSCov(F_x)$, we proceed as in the proof of Theorem 1 and use the orthogonal transformation $\boldsymbol{z} = \Gamma^T(\boldsymbol{x} - \boldsymbol{\mu})$.

Theorem 2. *Let $\boldsymbol{x}$ be a random vector from an elliptical distribution F with parameters $\boldsymbol{\mu}$ and Σ. Then the influence function of j-th eigenvector functional $\boldsymbol{p}_{S,j}$ based on $SCov(F_x)$ is given by*

$$IF(\boldsymbol{x}; \boldsymbol{p}_{S,j}, F) = \sum_{\substack{k=1 \\ k \neq j}}^{p} \frac{1}{\lambda_{S,j} - \lambda_{S,k}} \frac{z_k z_j}{||\boldsymbol{z}||^2} \boldsymbol{\gamma}_k,$$

and the influence function of $\boldsymbol{p}_{SS,j}$ based on $SSCov(F_x)$ is

$$IF(\boldsymbol{x}; \boldsymbol{p}_{SS,j}, F) = \sum_{\substack{k=1 \\ k \neq j}}^{p} \frac{2}{\lambda_{S,j} - \lambda_{S,k}} E_{\boldsymbol{z}_2} \left[\frac{(z_{1k} - z_{2k})(z_{1j} - z_{2j})}{||\boldsymbol{z}_1 - \boldsymbol{z}_2||^2} \right] \boldsymbol{\gamma}_k,$$

where $\boldsymbol{z}$, $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are independent observations from F_Λ and $\lambda_{S,j}$s are defined in (9).

To compare the influence functions of the theorem, we plot the norm of the influence function of the first eigenvector for different spatial covariance matrices and for the usual sample covariance matrix using a simple bivariate normal model with covariance matrix $\Sigma = diag(2, 1)$. The influence function of the first eigenvector based on the sample covariance matrix is computed using (14) below. Figure 1 shows that the eigenvector based on $SSCov(F_x)$ (right plot) has some very desirable properties: At the centre of the distribution, the shape of the influence function is similar to that of the sample covariance matrix (left plot), but points far away from the centre have no influence on the estimator. In comparison, $SCov(F_x)$ (middle plot) downweights the outliers, but at the centre of the distribution the shape of

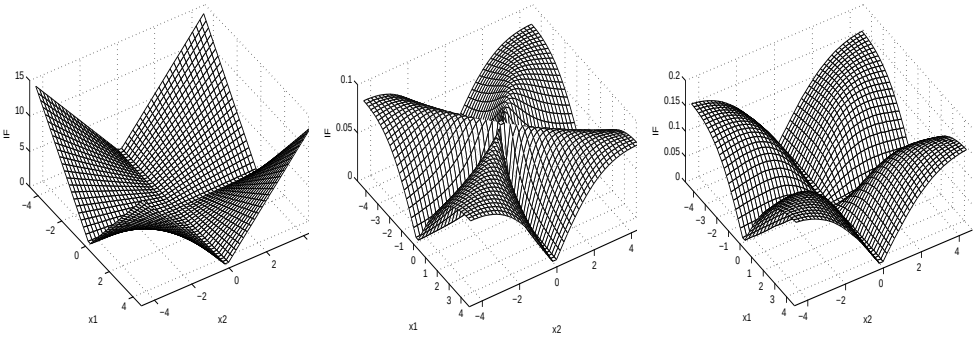

Figure 1: Norm of the influence function of the first eigenvector for the sample covariance matrix (left), the spatial sign covariance matrix (middle) and the symmetrised spatial sign covariance matrix (right) with $F = N(\mathbf{0}, diag(2, 1))$.

the influence function is different from that of the sample covariance matrix indicating poor efficiency properties.

We next derive the influence functions of eigenvector estimators based on $SCov_2(F_x)$ and $SSCov_2(F_x)$. Due to affine equivariance, the corresponding influence functions are easily obtained using the transformation $\boldsymbol{z}' = \Lambda^{-1/2}\Gamma^T(\boldsymbol{x} - \boldsymbol{\mu}) = r\boldsymbol{u}$, where $r = ||\boldsymbol{z}'||$ and $\boldsymbol{u} = r^{-1}\boldsymbol{z}'$, and r and $\boldsymbol{u}$ are independent. At elliptical F, the influence functions of all shape matrices $V(F_x)$ are of the form

$$IF(\boldsymbol{x}; V, F) = \alpha_V(r)\,[Det(\Sigma)]^{-1/p}\,\Sigma^{1/2}\left[\boldsymbol{u}\boldsymbol{u}^T - \frac{1}{p}I_p\right]\Sigma^{1/2}, \tag{13}$$

where $\alpha_V(r)$ is a real-valued weight function that depends on $V(F_x)$ and the underlying spherical distribution (Taskinen et al., 2006). Combining (12) and (13), the influence function of the eigenvector functional $\boldsymbol{p}_{V,j}$ based on any affine equivariant estimator $V(F_x)$ is thus

$$IF(\boldsymbol{x}; \boldsymbol{p}_{V,j}, F) = \alpha_V(r)\sum_{\substack{k=1\\k\neq j}}^{p}\frac{\sqrt{\lambda_j\lambda_k}}{\lambda_j - \lambda_k}\,u_k u_j\,\boldsymbol{\gamma}_k, \tag{14}$$

and comparisons between different estimators may be based on $\alpha_V(r)$ functions only. For $SCov_2(F_x)$ and $SSCov_2(F_x)$ these functions are derived in Tyler (1987) and Sirkiä et al. (2007). We denote them as

$$\alpha_{S_2}(r) = p + 2 \quad \text{and} \quad \alpha_{SS_2}(r) = 2(p+2)(1 - pg(r)), \tag{15}$$

where $g(r) = E_{\boldsymbol{z}_1'}(\boldsymbol{z}_{12}'^2/||\boldsymbol{z}_1' - r\boldsymbol{e}_1||^2)$, with $\boldsymbol{e}_1 = (1, 0, \ldots, 0)^T$. Note that $\alpha_{S_2}(r)$ is constant in r, thus making the estimates based on $SCov_2(F_x)$ highly robust, but not very efficient. As shown in Sirkiä et al. (2007), $\alpha_{SS_2}(r)$ is continuous and bounded, whereas for the sample covariance matrix, $\alpha_{Cov}(r) = r^2$, which makes the corresponding estimates highly sensitive to outliers.

3.2. Asymptotic distributions and efficiencies

Let $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ be $n \times p$ data matrix. We denote by $\hat{C}$ the estimator associated with the functional $C(F)$, that is, $\hat{C} = C(F_n)$, where F_n is the empirical distribution function based on X. The asymptotic behaviour of the eigenvector estimators based on different sign covariance matrices can be derived as in Anderson (1984) using the asymptotic normality of the corresponding covariance matrix estimators. As before, in the following theorem we assume that Σ has no multiple roots. For the limiting distribution of a Σ with multiple roots, see Anderson (1963). The *vec*-operator below vectorizes a matrix by stacking the columns of the matrix on top of each other.

Theorem 3. *Let $\hat{C}$ be any positive definite symmetric $p \times p$ matrix such that at F_Λ the limiting distribution of $\sqrt{n}\, vec(\hat{C} - \Lambda)$ is a p^2-variate (singular) normal distribution with mean zero. Write the spectral decomposition of $\hat{C}$ as $\hat{C} = \hat{P}\hat{\Lambda}\hat{P}^T$. Then the limiting distributions of $\sqrt{n}\, vec(\hat{P} - I_p)$ and $\sqrt{n}\, vec(\hat{\Lambda} - \Lambda)$ are multivariate (singular) normal and*

$$\sqrt{n}\, vec(\hat{C}-\Lambda) = ((\Lambda \otimes I_p) - (I_p \otimes \Lambda))\, \sqrt{n}\, vec(\hat{P}-I_p) + \sqrt{n}\, vec(\hat{\Lambda}-\Lambda) + o_P(1).$$

Note that the theorem does not yield the limiting distribution of $\sqrt{n}\,(\hat{p}_{ii}-1)$, where $\hat{p}_{ii}$ is the diagonal element of $\hat{P}$. In fact, $\sqrt{n}\,(\hat{p}_{ii} - 1) = o_P(1)$, as the limiting distribution of

$$n(1 - \hat{p}_{ii}^2) = n \sum_{j \neq i} \hat{p}_{ji}^2$$

is that of a weighted sum of independent chi-square random variables each with one degree of freedom.

Since $\widehat{SCov}$ and $\widehat{SSCov}$ are asymptotically normal under an elliptical F (see eg. Visuri et al. (2000)), the limiting normality of the corresponding eigenvector estimators follows from Theorem 3.

Corollary 1. *Let F be an elliptical distribution with parameters $\boldsymbol{\mu}$ and Σ. Then $\sqrt{n}(\hat{\boldsymbol{p}}_{S,j} - \boldsymbol{\gamma}_j)$ and $\sqrt{n}(\hat{\boldsymbol{p}}_{SS,j} - \boldsymbol{\gamma}_j)$ have limiting normal distributions with zero mean and asymptotic variances (ASV)*

$$ASV(\hat{\boldsymbol{p}}_{S,j}; F) = \sum_{\substack{k=1 \\ k\neq j}}^{p} \frac{1}{(\lambda_{S,j} - \lambda_{S,k})^2} E\left[\left(\frac{\sqrt{\lambda_{\Sigma,j}\lambda_{\Sigma,k}}u_j u_k}{\lambda_{\Sigma,1}u_1^2 + \cdots + \lambda_{\Sigma,p}u_p^2}\right)^2\right] \boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T$$

and

$$ASV(\hat{\boldsymbol{p}}_{SS,j}; F) = \sum_{\substack{k=1 \\ k\neq j}}^{p} \frac{4}{(\lambda_{S,j} - \lambda_{S,k})^2} E\left[\left(E_{\boldsymbol{z}_2}\left[\frac{(z_{1k} - z_{2k})(z_{1j} - z_{2j})}{||\boldsymbol{z}_1 - \boldsymbol{z}_2||^2}\right]\right)^2\right] \boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T,$$

where $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are independent observations from F_Λ, $\boldsymbol{u} = (u_1, \ldots, u_p)^T$ is uniformly distributed on the unit sphere and the $\lambda_{S,j}$'s are defined in (9).

The asymptotic distributions of $\widehat{SCov_2}$ and $\widehat{SSCov_2}$ under an elliptical F are derived for example by Tyler (1987) and Dümbgen (1998). The limiting normality of the corresponding eigenvector estimators thus follows.

Corollary 2. *Let F be an elliptical distribution with parameters $\boldsymbol{\mu}$ and Σ, and write $\hat{\boldsymbol{p}}_{V,j}$ for the eigenvector estimate based on any shape matrix estimator $\hat{V}$. Then $\sqrt{n}(\hat{\boldsymbol{p}}_{V,j} - \boldsymbol{\gamma}_j)$ has a limiting normal distribution with zero mean and asymptotic variance*

$$ASV(\hat{\boldsymbol{p}}_{V,j}; F) = \frac{E[\alpha_V^2(r)]}{p(p+2)} \sum_{\substack{k=1 \\ k\neq j}}^{p} \frac{\lambda_j\lambda_k}{(\lambda_j - \lambda_k)^2}\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T.$$

The $\alpha_V(r)$ functions for $SCov_2(F_x)$ and $SSCov_2(F_x)$ are given in (15).

To compare efficiency properties of estimators, we proceed as in Croux et al., (2002). The asymptotic relative efficiencies (ARE) of eigenvector estimators based on the scatter (or shape) estimator $\hat{C}_1$ relative to the eigenvector estimators based on $\hat{C}_2$ are then computed using

$$ARE(\hat{\boldsymbol{p}}_{C_1,j}, \hat{\boldsymbol{p}}_{C_2,j}; F) = \frac{Trace(ASV(\hat{\boldsymbol{p}}_{C_2,j}; F))}{Trace(ASV(\hat{\boldsymbol{p}}_{C_1,j}; F))}. \tag{16}$$

In Table 1, we list the asymptotic relative efficiencies for the first eigenvector estimates based on different spatial sign covariance matrices with respect to estimates based on sample covariance matrix. The efficiencies for $\widehat{SCov}$ and $\widehat{SSCov}$ are computed at different p-variate t-distribution cases with $\Sigma = diag(p, p-1, \ldots, 1)$ and selected values of degrees of freedom ν with $\nu = \infty$ referring to the multivariate normal case.

Notice that when efficiencies of eigenvector estimators based on two affine equivariant shape estimates, $\hat{V}_1$ and $\hat{V}_2$, are compared, (16) reduces to

$$ARE(\hat{\boldsymbol{p}}_{V_1,j}, \hat{\boldsymbol{p}}_{V_2,j}; F) = \frac{E[\alpha^2_{V_2}(r)]}{E[\alpha^2_{V_1}(r)]},$$

where functions α_{V_i}, given in (13), depend only on corresponding shape matrices as well as the underlying spherical distribution. The efficiencies of eigenvector estimators are therefore given by the efficiencies of corresponding shape matrices. When computing the efficiencies for $\widehat{SCov}_2$ and $\widehat{SSCov}_2$, we thus use spherical p-variate t-distributions with $\Sigma = I_p$. The results are listed in Table 1.

As seen in the table, the eigenvectors based on the symmetrised spatial sign covariance matrix $\widehat{SSCov}$ and Dümbgen's $\widehat{SSCov}_2$ are almost as efficient as those based on the classical sample covariance matrix in the multivariate normal case. For heavy-tailed distributions, the efficiencies are much higher. The eigenvectors based on the spatial sign covariance matrix $\widehat{SCov}$ and Tyler's shape estimate $\widehat{SCov}_2$ have much lower efficiencies than their symmetrised counterparts, but for heavy-tailed distributions, these estimators outperform the classical ones.

4. Conclusions

In this paper, orthogonally equivariant spatial sign covariance matrices as well as their affine equivariant counterparts were compared in a PCA framework. Robustness and efficiency studies showed that the eigenvectors based on different covariance matrix estimators are highly resistant to outliers, and, in particular, the eigenvectors based on symmetrised estimators have exellent asymptotic efficiencies. Estimates based on spatial signs are easy and fast to compute even for high-dimensional data thereby providing practical robust alternatives to classical sample covariance matrix based methods. Programs for computing estimators based on spatial sign vectors are available in the **R**-packages **SpatialNP** and **MNM** (Nordhausen and Oja, 2011).

Table 1: AREs of first eigenvector estimates based on (a) $\widehat{SSCov}$ and $\widehat{SCov}$ (between parenthesis) and (b) $\widehat{SSCov_2}$ and $\widehat{SCov_2}$ (between parenthesis) as compared to eigenvector estimates based on sample covariance matrix. The underlying distribution is the multivariate t-distribution with (a) $\Sigma = diag(p, p-1, \ldots, 1)$ and (b) $\Sigma = I_p$, and selected values of p and ν.

	(a)					(b)				
p	$\nu = 5$	$\nu = 6$	$\nu = 8$	$\nu = 15$	$\nu = \infty$	$\nu = 5$	$\nu = 6$	$\nu = 8$	$\nu = 15$	$\nu = \infty$
2	2.32	1.58	1.23	1.01	0.91	2.33	1.58	1.22	1.01	0.91
	(1.46)	(0.97)	(0.73)	(0.57)	(0.49)	(1.50)	(1.00)	(0.75)	(0.59)	(0.50)
4	2.41	1.66	1.27	1.05	0.93	2.45	1.66	1.28	1.05	0.93
	(1.95)	(1.29)	(0.97)	(0.76)	(0.64)	(2.00)	(1.33)	(1.00)	(0.79)	(0.67)
6	2.50	1.71	1.31	1.06	0.95	2.52	1.71	1.31	1.07	0.95
	(2.17)	(1.45)	(1.08)	(0.86)	(0.73)	(2.25)	(1.50)	(1.13)	(0.89)	(0.75)
8	2.52	1.72	1.34	1.07	0.95	2.57	1.74	1.33	1.09	0.96
	(2.32)	(1.54)	(1.16)	(0.91)	(0.77)	(2.40)	(1.60)	(1.20)	(0.95)	(0.80)
10	2.57	1.73	1.35	1.09	0.97	2.60	1.76	1.34	1.09	0.97
	(2.43)	(1.61)	(1.21)	(0.95)	(0.81)	(2.50)	(1.67)	(1.25)	(0.98)	(0.83)

Acknowledgements

The authors wish to thank the referee and the associate editor for their comments on the manuscript. The work of Sara Taskinen and Hannu Oja was supported by grants from the Academy of Finland.

References

Anderson, T.W., 1963. Asymptotic theory for principal component analysis. Ann. Math. Statist. 34, 122–148.

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

Brown, B.M., 1983. Statistical uses of the spatial median. J. Roy. Statist. Soc. Ser. B. 45, 25–30.

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., Ollila, E., and Oja, H., 2002. Sign and rank covariance matrices: statistical properties and application to principal components analysis, In:

Y. Dodge (Eds.), Statistical Data Analysis Based on the L1-Norm and Related Methods, Birkhauser, Basel, pp. 257–271.

Devlin, S.J., Gnanadesikan, R. and Kettenring, J.R., 1981. Robust estimation of dispersion matrices and principal components. J. Am. Statist. Assoc. 76, 354–362.

Dümbgen, L., 1998. On Tyler's M-functional of scatter in high dimension. Ann. Inst. Statist. Math. 50, 471–491.

Dümbgen, L. and Tyler, D., 2005. On the Breakdown Properties of Some Multivariate M-Functionals. Scand. J. Statist., 32, 247–264

Frahm, G. 2009. Asymptotic distributions of robust shape matrices and scales. J. Multivariate Analysis 100, 1329–1337.

Gervini, D., 2008. Robust functional estimation using the spatial median and spherical principal components. Biometrika 95, 587–600.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.J., 1986. Robust Statistics: the Approach Based on Influence Functions, Wiley, New York.

Hettmansperger, T.P. and Randles, R.H., 2003. A practical affine equivariant multivariate median. Biometrika 89, 851–860.

Locantore, N., Marron, J.S., Simpson, D.G., Tripoli, N., Zhang, J.T. and Cohen, K.L., 1999. Robust principal components for functional data. Test 8, 1–73.

Marden, J.I., 1999. Some robust estimates of principal components. Statist. Probab. Lett. 43, 349–359.

Nordhausen, K. and Oja H., 2011. Multivariate L1 Methods: The package MNM. Journal of Statistical Software, 43, 1–28.

Oja, H. and Randles, R., 2004. Multivariate nonparametric tests. Statist. Sci. 19, 598–605.

Paindaveine, D., 2008. A canonical definition of shape. Statist. Probab. Lett.. 78, 2240–2247.

Sirkiä, S., Taskinen, S., and Oja, H., 2007. Symmetrised M-estimators of scatter. J. Multivariate Anal. 98, 1611–1629.

Tanaka, Y., 1988. Sensitivity analysis in principal component analysis: Influence on the subspace spanned by principal component. Comm. Statist., 17, 3157-3175.

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

Tyler, D.E., 1987. A distribution-free M-estimator of multivariate scatter. Ann. Statist. 15, 234–251.

Tyler, D.E., 2010. A note on multivariate location and scatter statistics for sparse data sets. Statist. Prob. Lett.. 80, 1409–1413.

Vardi, Y. and Zhang, C.-H., 2001. A Modified Weiszfeld Algorithm for the Fermat-Weber Location Problem. Math. Program. 90, 559–566.

Visuri, S., Koivunen, V. and Oja, H., 2000. Sign and rank covariane matrices. J. Statist. Plann. Inf. 91, 557–575.

Appendix: proofs of the results

Proof of Theorem 1. Croux et al. (2002) noted that the transformed observation $\boldsymbol{z} = \Gamma^T(\boldsymbol{x} - \boldsymbol{\mu})$ can be written as $\boldsymbol{z} = r\Lambda^{1/2}\boldsymbol{u}$, where $r = ||\Lambda^{-1/2}\boldsymbol{z}||$ and $\boldsymbol{u} = \Lambda^{-1/2}\boldsymbol{z}/r$. Then r and $\boldsymbol{u}$ are independent, $\boldsymbol{u}$ is uniformly distributed on the unit sphere, and $SCov(F_x) = \Gamma\Lambda_S\Gamma^T$, where Λ_S is defined in (8). Since the pairwise differences $\boldsymbol{z}_1 - \boldsymbol{z}_2 = \Gamma^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)$ follow a centered distribution with diagonal covariance matrix 2Λ, we may write $\boldsymbol{z}_1 - \boldsymbol{z}_2 = r\sqrt{2}\Lambda^{1/2}\boldsymbol{u}$, where $r = ||\sqrt{2}^{-1}\Lambda^{-1/2}(\boldsymbol{z}_1 - \boldsymbol{z}_2)||$ and $\boldsymbol{u} = \sqrt{2}^{-1}\Lambda^{-1/2}(\boldsymbol{z}_1 - \boldsymbol{z}_2)/\mathrm{r}$ are independent with $\boldsymbol{u}$ uniformly distributed on the unit sphere. Then

$$\begin{aligned} SSCov(F_x) &= E\left[\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{||\boldsymbol{x}_1 - \boldsymbol{x}_2||^2}\right] = \Gamma E\left[\frac{(\boldsymbol{z}_1 - \boldsymbol{z}_2)(\boldsymbol{z}_1 - \boldsymbol{z}_2)^T}{||\boldsymbol{z}_1 - \boldsymbol{z}_2||^2}\right]\Gamma^T \\ &= \Gamma E\left[\frac{\Lambda^{1/2}\boldsymbol{u}\boldsymbol{u}^T\Lambda^{1/2}}{\boldsymbol{u}^T\Lambda\boldsymbol{u}}\right]\Gamma^T = \Gamma\Lambda_S\Gamma^T. \end{aligned}$$

Proof of Theorem 2. For an elliptical distribution F, the influence function of $SCov(F_x)$ given in (6) is

$$IF(\boldsymbol{x}; SCov, F) = \frac{(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T}{||\boldsymbol{x}-\boldsymbol{\mu}||^2} - SCov(F)$$

(Croux et al., 2002). Then the influence function of $\boldsymbol{p}_{S,j}$ based on $SCov(F_x)$ can be easily derived using (12) and the fact that $SCov(F_x) = \Gamma\Lambda_S\Gamma^T$.

The influence function of $SSCov(F_x)$ may be derived by applying functional (7) to $F_\epsilon = (1-\epsilon)F + \epsilon\Delta_{\boldsymbol{x}}$ and taking the derivative with respect to ϵ at $\epsilon = 0$. This yields

$$IF(\boldsymbol{x}; SSCov, F) = 2\,E_{\boldsymbol{x}_2}\left[\frac{(\boldsymbol{x}_1-\boldsymbol{x}_2)(\boldsymbol{x}_1-\boldsymbol{x}_2)^T}{||\boldsymbol{x}_1-\boldsymbol{x}_2||^2}\right] - 2\,SSCov(F_x).$$

Then using the above result and $SSCov(F_x) = \Gamma\Lambda_{SS}\Gamma^T$ we have that

$$\boldsymbol{\gamma}_k^T\, IF(\boldsymbol{x}; SSCov, F)\,\boldsymbol{\gamma}_j = \boldsymbol{\gamma}_k^T\left[E_{\boldsymbol{x}_2}\left[\frac{(\boldsymbol{x}_1-\boldsymbol{x}_2)(\boldsymbol{x}_1-\boldsymbol{x}_2)^T}{||\boldsymbol{x}_1-\boldsymbol{x}_2||^2}\right] - SSCov(F_x)\right]\boldsymbol{\gamma}_j$$
$$= E_{\boldsymbol{x}_2}\left[\frac{\boldsymbol{\gamma}_k^T(\boldsymbol{x}_1-\boldsymbol{x}_2)(\boldsymbol{x}_1-\boldsymbol{x}_2)^T\boldsymbol{\gamma}_j}{||\boldsymbol{x}_1-\boldsymbol{x}_2||^2}\right] - \lambda_{S,j}\boldsymbol{\gamma}_k^T\boldsymbol{\gamma}_j = E_{\boldsymbol{z}_2}\left[\frac{(z_{1k}-z_{2k})(z_{1k}-z_{2k})}{||\boldsymbol{z}_1-\boldsymbol{z}_2||^2}\right].$$

The result then follows from (12).

Proof of Theorem 3. Let $\hat{C}$ be any positive definite symmetric matrix such that the limiting distribution of $\sqrt{n}\,vec(\hat{C}-\Lambda)$ is p^2-variate normal with mean zero, and decompose $\hat{C} = \hat{P}\hat{\Lambda}\hat{P}^T$. As the transformation $\hat{C} \to (\hat{\Lambda}, \hat{P})$ is continuously differentiable in a neighborhood of (Λ, I_p), $\hat{P}$ and $\hat{\Lambda}$ are $\sqrt{n}$-consistent estimators of I_p and Λ, see e.g Theorem 13.5.1 in Anderson (1984). Further, since $\hat{C}\hat{P} = \hat{P}\hat{\Lambda}$, we may write

$$\sqrt{n}\,vec(\hat{C}\hat{P}-\Lambda) = \sqrt{n}\,vec((\hat{C}-\Lambda)\hat{P} + \Lambda(\hat{P}-I_p))$$
$$\sqrt{n}\,vec(\hat{P}\hat{\Lambda}-\Lambda) = \sqrt{n}\,vec((\hat{P}-I_p)\hat{\Lambda} + (\hat{\Lambda}-\Lambda)).$$

Thus

$$\sqrt{n}\,vec\left((\hat{C}-\Lambda)\hat{P} + \Lambda(\hat{P}-I_p)\right) = \sqrt{n}\,vec\left((\hat{P}-I_p)\hat{\Lambda} + (\hat{\Lambda}-\Lambda)\right),$$

and by Slutsky's Theorem and the properties of *vec*-operator, we have that

$$\sqrt{n}\,vec(\hat{C}-\Lambda) = ((\Lambda\otimes I_p) - (I_p\otimes\Lambda))\,\sqrt{n}\,vec(\hat{P}-I_p) + \sqrt{n}\,vec(\hat{\Lambda}-\Lambda) + o_P(1).$$

As $\hat{\Lambda}-\Lambda$ is diagonal and $(\Lambda\otimes I_p)-(I_p\otimes\Lambda)$ picks only the off-diagonal elements of $\hat{P}-I_p$, the limiting normality of $\sqrt{n}\,vec(\hat{P}-I_p)$ and $\sqrt{n}\,vec(\hat{\Lambda}-\Lambda)$ follows.

Proof of Corollary 1. Consider first the eigenvector estimators based on $\widehat{SCov}$. For simplicity, write $\hat{S}=\widehat{SCov}$. The central limit theorem implies that

$$\sqrt{n}\,vec(\hat{S}-\Lambda_S)\xrightarrow{d}N(0,ASV(\hat{S},F_\Lambda)),$$

where Λ_S is given in (6) and

$$\begin{aligned}ASV(\hat{S},F_\Lambda)&=E\left[vec\left(\frac{\boldsymbol{z}\boldsymbol{z}^T}{||\boldsymbol{z}||^2}-\Lambda_S\right)vec^T\left(\frac{\boldsymbol{z}\boldsymbol{z}^T}{||\boldsymbol{z}||^2}-\Lambda_S\right)\right]\\&=E\left[vec\left(\frac{\Lambda^{1/2}\boldsymbol{u}\boldsymbol{u}^T\Lambda^{1/2}}{\boldsymbol{u}^T\Lambda\boldsymbol{u}}-\Lambda_S\right)vec^T\left(\frac{\Lambda^{1/2}\boldsymbol{u}\boldsymbol{u}^T\Lambda^{1/2}}{\boldsymbol{u}^T\Lambda\boldsymbol{u}}-\Lambda_S\right)\right].\end{aligned}$$

In particular, the asymptotic variance of the off-diagonal element of $\hat{S}$ simplifies to

$$ASV(\hat{S}_{jk},F_\Lambda)=E\left[\left(\frac{\sqrt{\lambda_{\Sigma,j}\lambda_{\Sigma,k}}u_ju_k}{\lambda_{\Sigma,1}u_1^2+\cdots+\lambda_{\Sigma,p}u_p^2}\right)^2\right]. \tag{17}$$

Next decompose $\hat{S}=\hat{P}_S\hat{\Lambda}_S\hat{P}_S^T$ and note that at F_Λ, $\hat{P}_S$ and $\hat{\Lambda}_S$ are $\sqrt{n}$-consistent estimators of I_p and Λ_S. The limiting normality of eigenvector estimators then follows from Theorem 3 and for the off-diagonal elements of $\hat{P}_S$, we have that

$$\sqrt{n}\,\hat{p}_{jk}=(\lambda_{S,j}-\lambda_{S,k})^{-1}\sqrt{n}\,\hat{S}_{jk},\quad j\neq k.$$

The asymptotic variance at elliptical F follows from the orthogonal equivariance property of $\hat{P}_S$. Note first that,

$$\sqrt{n}(\hat{\boldsymbol{p}}'_{S,j}-\boldsymbol{\gamma}_j)=\sqrt{n}\,\Gamma\,(\hat{\boldsymbol{p}}_{S,j}-\boldsymbol{e}_j)=\sum_{\substack{k=1\\k\neq j}}^{p}\boldsymbol{\gamma}_j\sqrt{n}\,\hat{p}_{jk}+\boldsymbol{\gamma}_j\sqrt{n}\,(\hat{p}_{jj}-1),$$

where $\hat{\boldsymbol{p}}'_{S,j}$ denotes the eigenvector estimator computed at an elliptical F and $\boldsymbol{e}_j$ is the j-th column of I_p. Since $\sqrt{n}\,(\hat{p}_{jj}-1)=o_P(1)$, the asymptotic variance reduces to

$$ASV(\hat{\boldsymbol{p}}'_{S,j};F)=\sum_{\substack{k=1\\k\neq j}}^{p}(\lambda_{S,j}-\lambda_{S,k})^{-2}ASV(\hat{S}_{jk},F_\Lambda)\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T,$$

where $ASV(\hat{S}_{jk}, F_\Lambda)$ is given in (17).

The limiting normality of eigenvector estimatestimators based on $\widehat{SSCov}$ can be proved as above. The limiting normality of $\widehat{SSCov}$ follows from the U-statistics theory. Write $\hat{S} = \widehat{SSCov}$. Then $vec(\hat{S}) = \text{ave}\{h(\boldsymbol{z}_i, \boldsymbol{z}_j)\}$ with the kernel function

$$h(\boldsymbol{z}_1, \boldsymbol{z}_2) = vec\{S(\boldsymbol{z}_1 - \boldsymbol{z}_2)\boldsymbol{S}(\boldsymbol{z}_1 - \boldsymbol{z}_2)^T\}.$$

The limiting distribution is then given by

$$\sqrt{n}\, vec(\hat{S} - \boldsymbol{\theta}) \xrightarrow{d} N(0, ASV(\hat{S}, F_\Lambda)),$$

where $\boldsymbol{\theta} = E[h(\boldsymbol{z}_1, \boldsymbol{z}_2)] = vec(\Lambda_{SS})$ and $ASV(\hat{S}, F_\Lambda) = 4E[h_1(\boldsymbol{z}_1)h_1^T(\boldsymbol{z}_1)]$, where $h_1(\boldsymbol{z}_1) = E_{\boldsymbol{z}_2}[h(\boldsymbol{z}_1, \boldsymbol{z}_2)] - \boldsymbol{\theta}$. The asymptotic variance of the off-diagonal element of $\hat{S}$ is then given by

$$ASV(\hat{S}_{ij}, F_\Lambda) = 4\, E\left[\left(E_{\boldsymbol{z}_2}\left[\frac{(z_{1i} - z_{2i})(z_{1j} - z_{2j})}{||\boldsymbol{z}_1 - \boldsymbol{z}_2||^2}\right]\right)^2\right].$$

Proof of Corollary 2. Write first $\hat{V}'$ for the affine equivariant spatial sign covariance matrix estimate computed on spherical F_0. Then

$$\sqrt{n}\, vec(\hat{V}' - I_p) \xrightarrow{d} N(0, ASV(\hat{V}', F_0)),$$

with

$$ASV(\hat{V}', F_0) = ASV(\hat{V}'_{jk}, F_0)\left(I_{p^2} + K_{p,p} - 2p^{-1}vec(I_p)vec^T(I_p)\right),$$

where $K_{p,p}$ denotes a commutation matrix, that is, a $p^2 \times p^2$ block matrix with (i, j)-block being equal to a $p \times p$ matrix that has 1 at entry (j, i) and zero elsewhere. The asymptotic variances of any off-diagonal element satisfy

$$ASV(\hat{V}'_{jk}, F_0) = \frac{E[\alpha_C^2(r)]}{p(p+2)}.$$

Then due to affine equivariance of $\hat{V}'$ we have that at F_Λ,

$$\sqrt{n}\, vec(\hat{V} - \Lambda_\Sigma) = (\Lambda_\Sigma^{1/2} \otimes \Lambda_\Sigma^{1/2})\sqrt{n}\, vec(\hat{V}' - I_p).$$

The limiting normality at F_Λ therefore follows and

$$ASV(\hat{V}_{jk}, F_\Lambda) = \lambda_{\Sigma,j}\lambda_{\Sigma,k} ASV(\hat{V}'_{jk}, F_0).$$

Next decompose $\hat{V} = \hat{P}_V \hat{\Lambda}_V \hat{P}_V^T$ and note that at F_Λ, $\hat{P}_V$ and $\hat{\Lambda}_V$ are $\sqrt{n}$-consistent estimators of I_p and Λ_Σ. The limiting normality of eigenvector estimators then follows from Theorem 3 and the asymptotic variance may be computed as in the proof of Corollary 1 and it reduces to

$$ASV(\hat{\boldsymbol{p}}_{V,j}; F) = \sum_{\substack{k=1 \\ k\neq j}}^{p} (\lambda_{\Sigma,j} - \lambda_{\Sigma,k})^{-2} ASV(\hat{V}_{jk}, F_\Lambda) \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k^T.$$