

Robust estimation of principal components from depth-based multivariate rank covariance matrix

Subho Majumdar

Abstract: Analyzing principal components for multivariate data from its spatial sign covariance matrix (SCM) has been proposed as a computationally simple and robust alternative to normal PCA, but it suffers from poor efficiency properties and is actually inadmissible with respect to the maximum likelihood estimator. Here we use data depth-based spatial ranks in place of spatial signs to obtain the orthogonally equivariant Depth Covariance Matrix (DCM) and use its eigenvector estimates for PCA. We derive asymptotic properties of the sample DCM and influence functions of its eigenvectors. The shape of these influence functions indicate good efficiency properties compared to the SCM. Finite sample simulation studies show that principal components of the sample DCM are robust with respect to deviations from normality and infinitesimal contamination, as well as are more efficient than the SCM.

Keywords: Data depth; Principal components analysis; Robustness; Sign covariance matrix; Multivariate ranking

1 Introduction

2 Preliminaries

3 Depth-based rank covariance matrix

Consider a vector-valued random variable $\mathbf{X} \in \mathbb{R}^p$ that follows an elliptic distribution $|\Sigma|^{-1/2}f((\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}))$, with a mean vector $\boldsymbol{\mu}$ and covariance matrix Σ . Now, given a depth function $D_{\mathbf{X}}(\mathbf{x})$ (equivalently, an htped function $\tilde{D}_{\mathbf{X}}(\mathbf{x})$), transform the original random variable as: $\tilde{\mathbf{x}} = \tilde{D}_{\mathbf{X}}(\mathbf{x})\mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$, $\mathbf{S}(\cdot)$ being the spatial sign functional:

$$\mathbf{S}(\mathbf{x}) = \begin{cases} \mathbf{x}/\|\mathbf{x}\| & \text{if } \mathbf{x} \neq \mathbf{0} \\ \mathbf{0} & \text{if } \mathbf{x} = \mathbf{0} \end{cases}$$

The sign transformation is rotation equivariant, in the sense that $\mathbf{S}(P\mathbf{x}) = P\mathbf{x}/\|P\mathbf{x}\| = P\mathbf{x}/\|\mathbf{x}\| = P\mathbf{S}(\mathbf{x})$ for any orthogonal matrix P . The transformed random variable $\tilde{\mathbf{X}}$ can be seen as the multivariate rank corresponding to \mathbf{X} (e.g. Serfling (2006)). Now, consider the spectral decomposition of the covariance matrix: $\Sigma = \Gamma\Lambda\Gamma^T$, Γ being orthogonal and Λ diagonal with positive diagonal elements. Also normalize the original random variable as $\mathbf{z} = \Gamma^T\Lambda^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$. In this setup, we can represent the transformed random variable as

$$\begin{aligned} \tilde{\mathbf{x}} &= \tilde{D}_{\mathbf{X}}(\mathbf{x})\mathbf{S}(\mathbf{X} - \boldsymbol{\mu}) \\ &= \tilde{D}_{\Gamma\Lambda^{1/2}\mathbf{z} + \boldsymbol{\mu}}(\Gamma\Lambda^{1/2}\mathbf{z} + \boldsymbol{\mu})\mathbf{S}(\Gamma\Lambda^{1/2}\mathbf{z}) \\ &= \Gamma\tilde{D}_{\mathbf{Z}}(\mathbf{z})\mathbf{S}(\Lambda^{1/2}\mathbf{z}) \\ &= \Gamma\Lambda^{1/2}\tilde{D}_{\mathbf{Z}}(\mathbf{z})\mathbf{S}(\mathbf{z})\frac{\|\mathbf{z}\|}{\|\Lambda^{1/2}\mathbf{z}\|} \end{aligned}$$

$\tilde{D}_{\mathbf{z}}(\mathbf{z})$ is an odd function in \mathbf{z} because of affine invariance, as is $\|\mathbf{z}\|/\|\Lambda^{1/2}\mathbf{z}\|$. Since $\mathbf{S}(\mathbf{z})$ is odd in \mathbf{z} for circularly symmetric \mathbf{z} , it follows that $E(\tilde{\mathbf{X}}) = \mathbf{0}$, and consequently we obtain an expression for the covariance matrix of $\tilde{\mathbf{X}}$:

Theorem 3.1. *Let the random variable $\mathbf{X} \in \mathbb{R}^p$ follow an elliptical distribution with center $\boldsymbol{\mu}$ and covariance matrix $\Sigma = \Gamma\Lambda\Gamma^T$, its spectral decomposition. Then, given a depth function $D_{\mathbf{X}}(\cdot)$ the covariance matrix of the transformed random variable $\tilde{\mathbf{X}}$ is*

$$\text{Cov}(\tilde{\mathbf{X}}) = \Gamma\Lambda_{D,S}\Gamma^T, \quad \text{with} \quad \Lambda_{D,S} = E \left[(\tilde{D}_{\mathbf{z}}(\mathbf{z}))^2 \frac{\Lambda^{1/2}\mathbf{z}\mathbf{z}^T\Lambda^{1/2}}{\mathbf{z}^T\Lambda\mathbf{z}} \right] \quad (1)$$

where $\mathbf{z} = (z_1, \dots, z_p)^T \sim N(\mathbf{0}, I_p)$ and $\Lambda_{D,S}$ a diagonal matrix with diagonal entries

$$\lambda_{D,S,i} = E_{\mathbf{z}} \left[\frac{(\tilde{D}_{\mathbf{z}}(\mathbf{z}))^2 \lambda_i z_i^2}{\sum_{j=1}^p \lambda_j z_j^2} \right]$$

The matrix of eigenvectors of the covariance matrix, Γ , remains unchanged in the transformation $\mathbf{X} \rightarrow \tilde{\mathbf{X}}$. As a result, the multivariate rank vectors can be used for robust principal component analysis, which will be outlined in the following sections. However, as one can see in the above expression, the diagonal entries of $\Lambda_{D,S}$ do not change if a scale change is done on all entries of Λ , meaning the $\Lambda_{D,S}$ matrices corresponding to F and cF for some $c \neq 0$ will be same. This is the reason for lack of affine equivariance of the DCM. Following the case of multivariate sign covariance matrices (Taskinen et al., 2012) one can get back the shape components, i.e. original *standardized* eigenvalues Λ^* from $\Lambda_{D,S}$ by an iterative algorithm:

1. Set $k = 0$, and start with an initial value $\Lambda^{*(0)}$.
2. Calculate the next iterate

$$\Lambda^{*(k+1)} = \left[E \left(\frac{(\tilde{D}_{\mathbf{z}}(\mathbf{z}))^2 \mathbf{z}\mathbf{z}^T}{\mathbf{z}^T \Lambda^{*(k)} \mathbf{z}} \right) \right]^{-1} \Lambda_{D,S}$$

and standardize its eigenvalues:

$$\Lambda^{*(k+1)} = \frac{\Lambda^{*(k+1)}}{\det(\Lambda^{*(k+1)})^{1/p}}$$

3. Stop if convergence criterion is satisfied. Otherwise set $k \rightarrow k + 1$ and go to step 2.

Unlike sign covariance matrices and symmetrized sign covariance matrices (Dümbgen, 1998), however, attempting to derive an affine equivariant counterpart (as opposed to only orthogonal equivariance) of the DCM through an iterative approach analogous to Tyler (1987) will not result in anything new. This is because Tyler’s scatter matrix Σ_T is defined as the implicit solution to the following equation:

$$\Sigma_T = \frac{\mathbf{x}\mathbf{x}^T}{\mathbf{x}^T \Sigma_T^{-1} \mathbf{x}} \quad (2)$$

and simply replacing \mathbf{x} by its multivariate rank counterpart $\tilde{\mathbf{x}}$ will not change the estimate Σ_T as \mathbf{x} and $\tilde{\mathbf{x}}$ have the same directions. Instead we consider a depth-weighted version of Tyler’s scatter matrix (i.e. weights $(\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2$ in right side of (2)) in the simulations in Section 5. The simulations show that it has slightly better finite-sample efficiency than Σ_T but has same asymptotic performance. We conjecture that its concentration properties can be obtained by taking an approach similar to Soloveychik and Wiesel (2014).

4 Asymptotic results

4.1 The sample DCM

Let us now consider n iid random draws from our elliptic distribution F , say $\mathbf{X}_1, \dots, \mathbf{X}_n$. For ease of notation, denote $SS(\mathbf{x}; \boldsymbol{\mu}) = \mathbf{S}(\mathbf{x} - \boldsymbol{\mu})\mathbf{S}(\mathbf{x} - \boldsymbol{\mu})^T$. Then, given the depth func-

tion and known location center $\boldsymbol{\mu}$, one can show that the sample DCM $\sum_{i=1}^n (\tilde{D}_{\mathbf{X}}(\mathbf{x}_i))^2 SS(\mathbf{x}_i; \boldsymbol{\mu})/n$ is a \sqrt{n} -consistent estimator of $E[(\tilde{D}_{\mathbf{X}}(\mathbf{X}))^2 SS(\mathbf{x}; \boldsymbol{\mu})]$ by a straightforward application of CLT. But in practice the population depth function $D_{\mathbf{X}}(\mathbf{x}) = D(\mathbf{x}, F)$ is estimated by the depth function based on the empirical distribution function, F_n . Denote this sample depth by $D_{\mathbf{X}}^n(\mathbf{x}) = D(\mathbf{x}, F_n)$. One also needs to replace the known location parameter $\boldsymbol{\mu}$ by some estimator $\hat{\boldsymbol{\mu}}_n$. Examples of robust estimators of location that are relevant here include the spatial median (Brown, 1983; Haldane, 1948), Oja median (Oja, 1983), projection median (Zuo, 2003) etc. Now, given $D_{\mathbf{X}}^n(\cdot)$ and $\hat{\boldsymbol{\mu}}_n$, to plug them into the sample DCM and still go through with the consistency we need the following result:

Lemma 4.1. *Consider a random variable $\mathbf{X} \in \mathbb{R}^p$ having a continuous and symmetric distribution with location center $\boldsymbol{\mu}$ such that $E\|\mathbf{x} - \boldsymbol{\mu}\|^{-3/2} < \infty$. Given n random samples from this distribution, suppose $\hat{\boldsymbol{\mu}}_n$ is an estimator of $\boldsymbol{\mu}$ so that $\sqrt{n}(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}) = O_P(1)$. Then with the above notations, and given that $D_{\mathbf{X}}^n(\mathbf{x}) \xrightarrow{P} D_{\mathbf{X}}(\mathbf{x})$,*

$$\sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n (\tilde{D}_{\mathbf{X}}^n(\mathbf{x}_i))^2 SS(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_n) - \frac{1}{n} \sum_{i=1}^n (\tilde{D}_{\mathbf{X}}(\mathbf{x}_i))^2 SS(\mathbf{x}_i; \boldsymbol{\mu}) \right] \xrightarrow{P} 0$$

Given this lemma, we can now state the result for consistency of the sample DCM:

Theorem 4.2. *Consider n iid samples from the distribution in Lemma 4.1. Then, given a depth function $D_{\mathbf{X}}(\cdot)$ and an estimate of center $\hat{\boldsymbol{\mu}}_n$ so that $\sqrt{n}(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}) = O_P(1)$,*

$$\sqrt{n} \left[\text{vec} \left\{ \frac{1}{n} \sum_{i=1}^n (\tilde{D}_{\mathbf{X}}^n(\mathbf{x}_i))^2 SS(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_n) \right\} - E \left[\text{vec} \left\{ (\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2 SS(\mathbf{x}; \boldsymbol{\mu}) \right\} \right] \right] \xrightarrow{D} N_{p^2}(\mathbf{0}, V_{D,S}(F))$$

$$\text{with } V_{D,S}(F) = \text{Var} \left[\text{vec} \left\{ (\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2 SS(\mathbf{x}; \boldsymbol{\mu}) \right\} \right]$$

In case F is elliptical, an elaborate form of the covariance matrix $V_{D,S}(F)$ explic-

itly specifying each of its elements (more directly those of its Γ^T -rotated version) can be obtained, which is given in Appendix A. This form is useful when deriving limiting distributions of eigenvectors and eigenvalues of the sample DCM.

4.2 Eigenvectors and eigenvalues

Since we are mainly interested in using the DCM for a robust version of principal components analysis, from now on we assume that the eigenvalues of Σ are distinct: $\lambda_1 > \lambda_2 > \dots > \lambda_p$ to obtain asymptotic distributions of principal components. In the case of eigenvalues with larger than 1 multiplicities, the limiting distributions of eigenprojection matrices can be obtained analogous to those of the sign covariance matrix (Magyar and Tyler, 2014).

We are going to derive the asymptotic joint distributions of eigenvectors and eigenvalues of the sample DCM. The following result allows us to get these, provided we know the limiting distribution of the sample DCM itself:

Theorem 4.3. *(Taskinen et al., 2012) Let F_Λ be an elliptical distribution with a diagonal covariance matrix Λ , and \hat{C} be any positive definite symmetric $p \times p$ matrix such that at F_Λ the limiting distribution of $\sqrt{n}\text{vec}(\hat{C} - \Lambda)$ 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 $\sqrt{n}\text{vec}(\hat{P} - I_p)$ and $\sqrt{n}\text{vec}(\hat{\Lambda} - \Lambda)$ are multivariate (singular) normal and*

$$\sqrt{n}\text{vec}(\hat{C} - \Lambda) = [(\Lambda \otimes I_p) - (I_p \otimes \Lambda)] \sqrt{n}\text{vec}(\hat{P} - I_p) + \sqrt{n}\text{vec}(\hat{\Lambda} - \Lambda) + o_P(1) \quad (3)$$

The first matrix picks only off-diagonal elements of the LHS and the second one only diagonal elements. We shall now use this as well as the form of the asymptotic covariance

matrix of the vec of sample DCM, i.e. $V_{D,S}(F)$ to obtain limiting variance and covariances of eigenvalues and eigenvectors.

Corollary 4.4. *Consider the sample DCM $\hat{S}^D(F) = \sum_{i=1}^n (\tilde{D}_{\mathbf{X}}^n(\mathbf{x}_i))^2 SS(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_{\mathbf{n}})/n$ and its spectral decomposition $\hat{S}^D(F) = \hat{\Gamma}_D \hat{\Lambda}_D \hat{\Gamma}_D^T$. Then the matrices $G = \sqrt{n}(\hat{\Gamma}_D - \Gamma)$ and $L = \sqrt{n}(\hat{\Lambda}_D - \Lambda_{D,S})$ have independent distributions. $\text{vec}(G)$ asymptotically has a p^2 -variate normal distribution with mean $\mathbf{0}_{p^2}$, and the asymptotic variance and covariance of different columns of $G = (\mathbf{g}_1, \dots, \mathbf{g}_p)$ are as follows:*

$$AVar(\mathbf{g}_i) = \sum_{k=1; k \neq i}^p \frac{1}{(\lambda_{D,s,k} - \lambda_{D,S,i})^2} E \left[\frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 \lambda_i \lambda_k z_i^2 z_k^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k^T \quad (4)$$

$$ACov(\mathbf{g}_i, \mathbf{g}_j) = -\frac{1}{(\lambda_{D,s,i} - \lambda_{D,S,j})^2} E \left[\frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 \lambda_i \lambda_j z_i^2 z_j^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] \boldsymbol{\gamma}_j \boldsymbol{\gamma}_i^T; \quad i \neq j \quad (5)$$

where $\Gamma = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)$. The vector consisting of diagonal elements of L , say $\mathbf{l} = (l_1, \dots, l_p)^T$ asymptotically has a p -variate normal distribution with mean $\mathbf{0}_p$ and variance-covariance elements:

$$AVar(l_i) = E \left[\frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 \lambda_i^2 z_i^4}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] - \lambda_{D,S,i}^2 \quad (6)$$

$$ACov(l_i, l_j) = E \left[\frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 \lambda_i \lambda_j z_i^2 z_j^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] - \lambda_{D,S,i} \lambda_{D,S,j}; \quad i \neq j \quad (7)$$

5 Robustness and efficiency properties

In this section, we first obtain the influence functions of the DCM as well as its eigenvectors and eigenvalues, which are essential to understand how much influence a sample point, especially an infinitesimal contamination, has on any functional on the distribu-

tion (Hampel et al., 1986). We shall also derive the asymptotic efficiencies of individual principal components with respect to those of the original covariance matrix and sign covariance matrix. Unlike affine equivariant estimators of shape, the ARE of eigenvectors (with respect to any other affine equivariant estimator) can not be simplified as a ratio of two scalar quantities dependent on only the distribution of $\|\mathbf{z}\|$ (e.g. Ollilia et al. (2003); Taskinen et al. (2012)), but actually depend on the (standardized) eigenvalues. Finite sample efficiency of the DCM estimates with respect to infinitesimal contamination and heavy-tailed distributions shall also be demonstrated by a simulation study.

5.1 Influence functions

Given any probability distribution F , the influence function of any point \mathbf{x}_0 in the sample space \mathcal{X} for some functional $T(F)$ on the distribution is defined as

$$IF(\mathbf{x}_0; T, F) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} (T(F_\epsilon) - T(F))$$

where F_ϵ is F with an additional mass of ϵ at \mathbf{x}_0 , i.e. $F_\epsilon = (1 - \epsilon)F + \epsilon\Delta_{\mathbf{x}_0}$; $\Delta_{\mathbf{x}_0}$ being the distribution with point mass at \mathbf{x}_0 . When $T(F) = E_F g$ for some F -integrable function g , $IF(\mathbf{x}_0; T, F) = g(\mathbf{x}_0) - T(F)$. It now follows that for the DCM,

$$IF(\mathbf{x}_0; Cov(\tilde{\mathbf{X}}), F) = (\tilde{D}_{\mathbf{X}}(\mathbf{x}_0))^2 SS(\mathbf{x}_0; \boldsymbol{\mu}) - Cov(\tilde{\mathbf{X}})$$

Following Croux and Haesbroeck (2000), we now get the influence function of the i^{th}

column of $\hat{\Gamma}_D = (\hat{\gamma}_{D,1}, \dots, \hat{\gamma}_{D,p})$; $i = 1, \dots, p$:

$$\begin{aligned}
IF(\mathbf{x}_0; \hat{\gamma}_{D,i}, F) &= \sum_{k=1; k \neq i}^p \frac{1}{\lambda_{D,S,i} - \lambda_{D,S,k}} \left\{ \gamma_k^T IF(\mathbf{x}_0; Cov(\tilde{\mathbf{X}}), \gamma_i) \right\} \gamma_k \\
&= \sum_{k=1; k \neq i}^p \frac{1}{\lambda_{D,S,i} - \lambda_{D,S,k}} \left\{ \gamma_k^T (\tilde{D}_{\mathbf{x}}(\mathbf{x}_0))^2 SS(\mathbf{x}_0; \boldsymbol{\mu}) \gamma_i - \lambda_{D,S,i} \gamma_k^T \gamma_i \right\} \gamma_k \\
&= \sum_{k=1; k \neq i}^p \frac{\sqrt{\lambda_i \lambda_k} z_{0i} z_{0k}}{\lambda_{D,S,i} - \lambda_{D,S,k}} \cdot \frac{(\tilde{D}_{\mathbf{z}}(\mathbf{z}_0))^2}{\mathbf{z}_0^T \Lambda \mathbf{z}_0} \gamma_k
\end{aligned}$$

where $\Gamma^T \Lambda^{-1/2}(\mathbf{x}_0 - \boldsymbol{\mu}) = \mathbf{z}_0 = (z_{01}, \dots, z_{0p})^T$.

5.2 Asymptotic and finite-sample efficiencies

Suppose $\hat{\Sigma}$ is a \sqrt{n} -consistent estimator of the population covariance matrix Σ , which permits a spectral decomposition $\hat{\Sigma} = \hat{\Gamma} \hat{\Lambda} \hat{\Gamma}^T$, where $\hat{\Gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_p)$. Then the asymptotic variance of the eigenvectors are (see Theorem 13.5.1 in Anderson (3rd ed. 2003))

$$AVar(\sqrt{n} \hat{\gamma}_i) = \sum_{k=1; k \neq i}^p \frac{\lambda_i \lambda_k}{(\lambda_i - \lambda_k)^2} \gamma_k \gamma_k^T \quad (8)$$

the asymptotic relative efficiencies of eigenvectors from the sample DCM with respect to the sample covariance matrix can now be derived using (8) above and (4) from Corollary 4.4:

$$\begin{aligned}
ARE(\hat{\gamma}_i^D, \hat{\gamma}_i; F) &= \frac{Trace(AVar(\sqrt{n} \hat{\gamma}_i^D))}{Trace(AVar(\sqrt{n} \hat{\gamma}_i))} \\
&= \left[\sum_{k=1; k \neq i}^p \frac{\lambda_i \lambda_k}{(\lambda_i - \lambda_k)^2} \right]^{-1} \left[\sum_{k=1; k \neq i}^p \frac{\lambda_i \lambda_k}{(\lambda_{D,S,i} - \lambda_{D,S,k})^2} E \left(\frac{(\tilde{D}_{\mathbf{z}}(\mathbf{z}))^4 z_i^2 z_k^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right) \right]
\end{aligned}$$

Appendix A: Form of $V_{D,S}(F)$

First observe that for F having covariance matrix $\Sigma = \Gamma\Lambda\Gamma^T$,

$$V_{D,S}(F) = (\Gamma \otimes \Gamma)V_{D,S}(F_\Lambda)(\Gamma \otimes \Gamma)^T$$

Where F_Λ has the same elliptic distribution as F , but with covariance matrix Λ . Now,

$$\begin{aligned} V_{D,S}(F_\Lambda) &= E \left[\text{vec} \left\{ \frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 \Lambda^{1/2} \mathbf{z} \mathbf{z}^T \Lambda^{1/2}}{\mathbf{z}^T \Lambda \mathbf{z}} - \Lambda_{D,S} \right\} \text{vec}^T \left\{ \frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 \Lambda^{1/2} \mathbf{z} \mathbf{z}^T \Lambda^{1/2}}{\mathbf{z}^T \Lambda \mathbf{z}} - \Lambda_{D,S} \right\} \right] \\ &= E \left[\text{vec} \left\{ (\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 SS(\Lambda^{1/2} \mathbf{z}; \mathbf{0}) \right\} \text{vec}^T \left\{ (\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 SS(\Lambda^{1/2} \mathbf{z}; \mathbf{0}) \right\} \right] \\ &\quad - \text{vec}(\Lambda_{D,S}) \text{vec}^T(\Lambda_{D,S}) \end{aligned}$$

The matrix $\text{vec}(\Lambda_{D,S}) \text{vec}^T(\Lambda_{D,S})$ consists of elements $\lambda_i \lambda_j$ at $(i, j)^{\text{th}}$ position of the $(i, j)^{\text{th}}$ block, and 0 otherwise. These positions correspond to variance and covariance components of on-diagonal elements. For the expectation matrix, all its elements are of the form $E[\sqrt{\lambda_a \lambda_b \lambda_c \lambda_d} z_a z_b z_c z_d \cdot (\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 / (\mathbf{z}^T \Lambda \mathbf{z})^2]$, with $1 \leq a, b, c, d \leq p$. Since $(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 / (\mathbf{z}^T \Lambda \mathbf{z})^2$ is even in \mathbf{z} , which has a circularly symmetric distribution, all such expectations will be 0 unless $a = b = c = d$, or they are pairwise equal. Following a similar derivation for spatial sign covariance matrices in Magyar and Tyler (2014), we collect the non-zero elements and write the matrix of expectations:

$$(I_{p^2} + K_{p,p}) \left\{ \sum_{a=1}^p \sum_{b=1}^p \gamma_{ab}^D (\mathbf{e}_a \mathbf{e}_a^T \otimes \mathbf{e}_b \mathbf{e}_b^T) - \sum_{a=1}^p \gamma_{aa}^D (\mathbf{e}_a \mathbf{e}_a^T \otimes \mathbf{e}_a \mathbf{e}_a^T) \right\} + \sum_{a=1}^p \sum_{b=1}^p \gamma_{ab}^D (\mathbf{e}_a \mathbf{e}_b^T \otimes \mathbf{e}_a \mathbf{e}_b^T)$$

where $I_k = (\mathbf{e}_1, \dots, \mathbf{e}_k)$, $K_{m,n} = \sum_{i=1}^m \sum_{j=1}^n J_{ij} \otimes J_{ij}^T$ with J_{ij} the $m \times n$ matrix having 1 as $(i, j)^{\text{th}}$ element and 0 elsewhere, and $\gamma_{mn}^D = E[\lambda_m \lambda_n z_m^2 z_n^2 \cdot (\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 / (\mathbf{z}^T \Lambda \mathbf{z})^2]$; $1 \leq m, n \leq p$.

Putting everything together, denote $\hat{S}^D(F_\Lambda) = \sum_{i=1}^n (\tilde{D}_{\mathbf{Z}}^n(\mathbf{z}_i))^2 SS(\Lambda^{1/2} \mathbf{z}_i; \hat{\boldsymbol{\mu}}_n)/n$. Then the different types of elements in the matrix $V_{D,S}(F_\Lambda)$ are as given below ($1 \leq a, b, c, d \leq p$):

- Variance of on-diagonal elements

$$AVar(\sqrt{n}\hat{S}_{aa}^D(F_\Lambda)) = E \left[\frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 \lambda_a^2 z_a^4}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] - \lambda_{D,S,a}^2$$

- Variance of off-diagonal elements ($a \neq b$)

$$AVar(\sqrt{n}\hat{S}_{ab}^D(F_\Lambda)) = E \left[\frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 \lambda_a \lambda_b z_a^2 z_b^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right]$$

- Covariance of two on-diagonal elements ($a \neq b$)

$$ACov(\sqrt{n}\hat{S}_{aa}^D(F_\Lambda), \sqrt{n}\hat{S}_{bb}^D(F_\Lambda)) = E \left[\frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 \lambda_a \lambda_b z_a^2 z_b^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] - \lambda_{D,S,a} \lambda_{D,S,b}$$

- Covariance of two off-diagonal elements ($a \neq b \neq c \neq d$)

$$ACov(\sqrt{n}\hat{S}_{ab}^D(F_\Lambda), \sqrt{n}\hat{S}_{cd}^D(F_\Lambda)) = 0$$

- Covariance of one off-diagonal and one on-diagonal element ($a \neq b \neq c$)

$$ACov(\sqrt{n}\hat{S}_{ab}^D(F_\Lambda), \sqrt{n}\hat{S}_{cc}^D(F_\Lambda)) = 0$$

Appendix B: Proofs

Proof of Theorem 3.1.

□

Proof of Lemma 4.1. For two positive definite matrices A, B , we denote by $A > B$ that $A - B$ is positive definite. Also, denote

$$S_n = \sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n \left| (\tilde{D}_{\mathbf{X}}^n(\mathbf{x}_i))^2 - (\tilde{D}_{\mathbf{X}}(\mathbf{x}_i))^2 \right| SS(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_n) \right]$$

Now, convergence in probability of $\tilde{D}_{\mathbf{X}}^n(\cdot)$ to $\tilde{D}_{\mathbf{X}}(\cdot)$ (hence $(\tilde{D}_{\mathbf{X}}^n(\cdot))^2$ to $(\tilde{D}_{\mathbf{X}}(\cdot))^2$) implies that given $\epsilon, \delta > 0$ there exists $N \in \mathbb{N}$ so that

$$\begin{aligned} P \left[|(\tilde{D}_{\mathbf{X}}^{n_1}(\mathbf{x}))^2 - (\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2| > \epsilon \right] &< \delta \\ \Rightarrow \quad |(\tilde{D}_{\mathbf{X}}^{n_1}(\mathbf{x}))^2 - (\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2| &< \epsilon(1 - \delta) + M\delta \\ &< \epsilon + M\delta \end{aligned} \tag{9}$$

for all $n_1 \geq N$, where M is the largest value the square of a depth function can take. Consequently for all such n_1 ,

$$S_{n_1} < (\epsilon + M\delta) \cdot \sqrt{n_1} \left[\frac{1}{n_1} \sum_{i=1}^{n_1} SS(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_{n_1}) \right]$$

We now construct a sequence of positive definite matrices $\{A_k(B_k + C_k) : k \in \mathbb{N}\}$ so that

$$\begin{aligned} A_k &= \frac{1 + M}{k}, \quad B_k = \sqrt{N_k} \left[\frac{1}{N_k} \sum_{i=1}^{N_k} \{SS(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_{N_k}) - SS(\mathbf{x}_i; \boldsymbol{\mu})\} \right] \\ C_k &= \sqrt{N_k} \left[\frac{1}{N_k} \sum_{i=1}^{N_k} SS(\mathbf{x}_i; \boldsymbol{\mu}) \right] \end{aligned}$$

where $N_k \in \mathbb{N}$ gives the relation (9) in place of N when we take $\epsilon = \delta = 1/k$. Under conditions $E\|\mathbf{x} - \boldsymbol{\mu}\|^{-3/2} < \infty$ and $\sqrt{n}(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}) = O_P(1)$, the sample sign covariance matrix converges in probability to its population version (Dürre et al., 2014), hence $B_k \xrightarrow{P} 0$. Now by Slutsky's theorem, $A_k(B_k + C_k) \xrightarrow{P} 0$. Finally, for any $\epsilon_1 > 0$, $P(S_{N_k} > \epsilon_1) < P(A_k B_k >$

ϵ_1), hence the subsequence $\{S_{N_k}\} \xrightarrow{P} 0$. Since the main sequence $\{S_k\}$ is bounded below by 0, we have the needed. \square

Proof of Theorem 4.2. The quantity in the statement of the theorem can be broken down as:

$$\begin{aligned} & \sqrt{n} \left[\text{vec} \left\{ \frac{1}{n} \sum_{i=1}^n (\tilde{D}_{\mathbf{x}}^n(\mathbf{x}_i))^2 SS(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_n) \right\} - \text{vec} \left\{ \frac{1}{n} \sum_{i=1}^n (\tilde{D}_{\mathbf{x}}(\mathbf{x}_i))^2 SS(\mathbf{x}_i; \boldsymbol{\mu}) \right\} \right] + \\ & \sqrt{n} \left[\text{vec} \left\{ \frac{1}{n} \sum_{i=1}^n (\tilde{D}_{\mathbf{x}}(\mathbf{x}_i))^2 SS(\mathbf{x}_i; \boldsymbol{\mu}) \right\} - E \left[\text{vec} \left\{ (\tilde{D}_{\mathbf{x}}(\mathbf{x}))^2 SS(\mathbf{x}; \boldsymbol{\mu}) \right\} \right] \right] \end{aligned}$$

The first part goes to 0 in probability by Lemma 4.1, and applying Slutsky's theorem we get the required convergence. \square

Proof of Theorem 4.3. See Taskinen et al. (2012) \square

Proof of Corollary 4.4. In spirit, this corollary is similar to Theorem 13.5.1 in Anderson (3rd ed. 2003), and indeed, Taskinen et al. (2012) used this theorem to prove Theorem 4.3. Due to the decomposition (3) we have, for the distribution F_Λ , the following relation between any off-diagonal element of $\hat{S}^D(F_\Lambda)$ and the corresponding element in the estimate of eigenvectors $\hat{\Gamma}_D(F_\Lambda)$:

$$\sqrt{n} \hat{\gamma}_{D,ij}(F_\Lambda) = \sqrt{n} \frac{\hat{S}_{ij}^D(F_\Lambda)}{\lambda_{D,S,i} - \lambda_{D,S,j}}; \quad i \neq j$$

So that for eigenvector estimates of the original F we have

$$\sqrt{n}(\hat{\gamma}_{D,i} - \boldsymbol{\gamma}_i) = \sqrt{n}(\hat{\gamma}_{D,i}(F_\Lambda) - \mathbf{e}_i) = \sqrt{n} \left[\sum_{k=1; k \neq i}^p \hat{\gamma}_{D,ik}(F_\Lambda) \boldsymbol{\gamma}_k + (\hat{\gamma}_{D,ii}(F_\Lambda) - 1) \boldsymbol{\gamma}_i \right] \quad (\text{A1})$$

$\sqrt{n}(\hat{\gamma}_{D,ii}(F_\Lambda) - 1) = o_P(1)$ and $ACov(\sqrt{n}\hat{S}_{ik}^D(F_\Lambda), \sqrt{n}\hat{S}_{il}^D(F_\Lambda)) = 0$ for $k \neq l$, so the above equation implies

$$AVar(\mathbf{g}_i) = AVar(\sqrt{n}(\hat{\gamma}_{D,i} - \gamma_i)) = \sum_{k=1; k \neq i}^p \frac{AVar(\sqrt{n}\hat{S}_{ik}^D(F_\Lambda))}{(\lambda_{D,s,i} - \lambda_{D,S,k})^2} \gamma_k \gamma_k^T$$

For the covariance terms, from (A1) we get, for $i \neq j$,

$$\begin{aligned} ACov(\mathbf{g}_i, \mathbf{g}_j) &= ACov(\sqrt{n}(\hat{\gamma}_{D,i} - \gamma_i), \sqrt{n}(\hat{\gamma}_{D,j} - \gamma_j)) \\ &= ACov\left(\sum_{k=1; k \neq i}^p \sqrt{n}\hat{\gamma}_{D,ik}(F_\Lambda)\gamma_k, \sum_{k=1; k \neq j}^p \sqrt{n}\hat{\gamma}_{D,jk}(F_\Lambda)\gamma_k\right) \\ &= ACov(\sqrt{n}\hat{\gamma}_{D,ij}(F_\Lambda)\gamma_j, \sqrt{n}\hat{\gamma}_{D,ji}(F_\Lambda)\gamma_i) \\ &= -\frac{AVar(\sqrt{n}\hat{S}_{ij}^D(\Lambda))}{(\lambda_{D,s,i} - \lambda_{D,S,j})^2} \gamma_j \gamma_i^T \end{aligned}$$

The exact forms given in the statement of the corollary now follows from the Form of $V_{D,S}$ in Appendix A.

For the on-diagonal elements of $\hat{S}^D(F_\Lambda)$ Theorem 4.3 gives us $\sqrt{n}\hat{\lambda}_{D,s,i}(F_\Lambda) = \sqrt{n}\hat{S}_{ii}^D(F_\Lambda)$ for $i = 1, \dots, p$. Hence

$$\begin{aligned} AVar(l_i) &= AVar(\sqrt{n}\hat{\lambda}_{D,s,i} - \sqrt{n}\lambda_{D,S,i}) \\ &= AVar(\sqrt{n}\hat{\lambda}_{D,s,i}(F_\Lambda) - \sqrt{n}\lambda_{D,S,i}(F_\Lambda)) \\ &= AVar(\sqrt{n}S_{ii}^D(F_\Lambda)) \end{aligned}$$

A similar derivation gives the expression for $AVar(l_i, l_j); i \neq j$. Finally, since the asymptotic covariance between an on-diagonal and an off-diagonal element of $\hat{S}^D(F_\Lambda)$, it follows that the elements of G and diagonal elements of L are independent. \square

References

T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, Hoboken, NJ, 3rd ed. 2003.

- B.M. Brown. Statistical Use of the Spatial Median. *J. Royal Stat. Soc. Ser. B*, 45:25–30, 1983.
- C. Croux and G. Haesbroeck. Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*, 87:603–618, 2000.
- L. Dümbgen. On Tyler’s M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.*, 50:471–491, 1998.
- A. Dürre, D. Vogel, and D.E. Tyler. The spatial sign covariance matrix with unknown location. *J. Mult. Anal.*, 130:107–117, 2014.
- J.B.S. Haldane. Note on the Median of a Multivariate Distribution. *Biometrika*, 35: 414–415, 1948.
- F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Staehel. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York, NY, 1986.
- A.F. Magyar and D.E. Tyler. The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions. *Biometrika*, 101:673–688, 2014.
- H. Oja. Descriptive Statistics for Multivariate Distributions. *Statist. and Prob. Lett.*, 1: 327–332, 1983.
- E. Ollilia, H. Oja, and C. Croux. The affine equivariant sign covariance matrix: asymptotic behavior and efficiencies. *J. Mult. Anal.*, 87:328–355, 2003.
- R. Serfling. Depth Functions in Nonparametric Multivariate Inference. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 72, pages 1–16, 2006.

- I. Soloveychik and A. Wiesel. Performance Analysis of Tyler’s Covariance Estimator. *IEEE Trans. Sig. Process.*, 63-2:418–426, 2014.
- S. Taskinen, I. Koch, and H. Oja. Robustifying principal component analysis with spatial sign vectors. *Statist. and Prob. Lett.*, 82:765–774, 2012.
- D.E. Tyler. A distribution-free M-estimator of multivariate scatter. *Ann. Statist.*, 15: 234–251, 1987.
- Y. Zuo. Projection-based depth functions and associated medians. *Ann. Statist.*, 31: 1460–1490, 2003.