

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

Subhabrata 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 robustness of principal components, and 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, as well as are more efficient than the SCM and its affine equivariant version, Tyler's shape matrix.

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

# 1 Introduction

In multivariate analysis, the study of principal components has been important because it provides a small number of uncorrelated variables from analyzing data on a potentially large number of variables, so that these new components explain most of the underlying variability in the original data. In case of multivariate normal distribution, the sample covariance matrix provides the most asymptotically efficient estimates of eigenvectors/ principal components, but it is extremely sensitive to outliers as well as relaxations of the normality assumption. To tackle this, several estimators of the population covariance or correlation matrix have been proposed which can be used for PCA. They can be roughly put into these categories: robust, breakdown point estimators that are computation-intensive (??); M-estimators that are calculated by simple iterative algorithms but not necessarily high-breakdown point (??); and symmetrised estimators that are highly efficient and robust to deviations from normality, but sensitive to outliers and computationally demanding (??).

When principal components are of interest, one can also estimate the population eigenvectors by analyzing the spatial sign of a multivariate vector: the vector divided by its magnitude, instead of the original data, and this yields computationally simple and high-breakdown estimates of principal components (??). Although the resulting Sign Covariance Matrix (SCM) is not affine equivariant, its orthogonal equivariance suffices for the purpose of PCA. However, the resulting estimates are not very efficient, and are in fact asymptotically inadmissible (?), in the sense that there is an estimator (Tyler's M-estimate of scatter, to be precise) that has uniformly lower asymptotic risk than the SCM.

In this paper we study the covariance matrix, paying special attention to its eigenvectors, of the multivariate rank vector that is obtained from the data-depth of a point and its spatial sign. The nonparametric concept of data-depth had first been proposed by ? when he introduced the halfspace depth. Given a dataset, the depth of a given point in the sample space measures how far inside the data cloud the point exists. An overview

of statistical depth functions can be found in (?). Depth-based methods have recently been popular for robust nonparametric classification (????). In parametric estimation, depth-weighted means (?) and covariance matrices (?) provide high-breakdown point as well as efficient estimators, although they do involve choice of a suitable weight function and tuning parameters.

The paper is arranged in the following fashion. Section 2 provides with the preliminary theoretical concepts and notations required for developments in the subsequent chapters. Section 3 introduces the Depth Covariance Matrix (DCM) and states some basic results relating to this. Section 4 provides asymptotic results regarding the sample DCM, calculated using data depths with respect to the empirical distribution function, as well as its eigenvectors and eigenvalues. Section 5 focuses solely on principal component estimation using the sample DCM. We obtain influence functions and asymptotic efficiencies for eigenvectors of the DCM. We also compare their finite sample efficiencies for several multinormal and multivariate  $t$ -distributions with those of the SCM, Tyler's scatter matrix and its depth-weighted version through a simulation study. Finally, we wrap up our discussion in Section 6 by giving a summary of our findings and providing some potential future areas of research.

## 2 Preliminaries

### 2.1 Spatial signs and sign covariance matrix

Given a vector  $\mathbf{x} \in \mathbb{R}^p$ , its spatial sign is defined as the vector valued function (?):

$$\mathbf{S}(\mathbf{x}) = \begin{cases} \mathbf{x} \|\mathbf{x}\|^{-1} & \text{if } \mathbf{x} \neq \mathbf{0} \\ \mathbf{0} & \text{if } \mathbf{x} = \mathbf{0} \end{cases}$$

When  $\mathbf{x}$  is a random vector 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  $\Sigma$ , the sign vectors  $\mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$  re-

side on the surface of a  $p$ -dimensional unit ball centered at  $\boldsymbol{\mu}$ . Denote by  $\Sigma_S(\mathbf{X}) = E\mathbf{S}(\mathbf{X} - \boldsymbol{\mu})\mathbf{S}(\mathbf{X} - \boldsymbol{\mu})^T$  the covariance matrix of spatial signs, or the *Sign Covariance Matrix* (SCM). The transformation  $\mathbf{x} \mapsto \mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$  keeps eigenvectors unchanged, and eigenvectors of the sample SCM  $\hat{\Sigma}_S = \sum_{i=1}^n \mathbf{S}(\mathbf{x}_i - \boldsymbol{\mu})\mathbf{S}(\mathbf{x}_i - \boldsymbol{\mu})^T/n$  are  $\sqrt{n}$ -consistent estimators of their population counterparts (?).

The sign transformation is rotation equivariant, i.e.  $\mathbf{S}(P(\mathbf{x} - \boldsymbol{\mu})) = P(\mathbf{x} - \boldsymbol{\mu})/\|P(\mathbf{x} - \boldsymbol{\mu})\| = P(\mathbf{x} - \boldsymbol{\mu})/\|\mathbf{x} - \boldsymbol{\mu}\| = P\mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$  for any orthogonal matrix  $P$ , and as a result the SCM is rotation equivariant too, in the sense that  $\Sigma_S(P\mathbf{X}) = P\Sigma_S(\mathbf{X})P^T$ . But this does not hold for any general non-singular matrix. An affine equivariant version of the sample SCM is obtained as the solution  $\hat{\Sigma}_T$  of the following equation:

$$\hat{\Sigma}_T(\mathbf{X}) = \frac{p}{n} \sum_{i=1}^n \frac{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T}{(\mathbf{x} - \boldsymbol{\mu})^T \hat{\Sigma}_T(\mathbf{X})^{-1} (\mathbf{x} - \boldsymbol{\mu})}$$

which turns out to be Tyler's M-estimator of scatter (?). In this context, one should note that for scatter matrices, affine equivariance will mean any affine transformation on the original random variable  $\mathbf{X} \mapsto \mathbf{X}^* = A\mathbf{X} + \mathbf{b}$  ( $A$  non-singular,  $\mathbf{b} \in \mathbb{R}^p$ ) being carried over to the covariance matrix estimate upto a scalar multiple:  $\hat{\Sigma}_T(\mathbf{X}^*) = k.A\hat{\Sigma}_T(\mathbf{X})A^T$  for some  $k > 0$ .

## 2.2 Data depth and outlyingness

For any multivariate distribution  $F = F_{\mathbf{X}}$ , the depth of a point  $\mathbf{x} \in \mathbb{R}^p$ , say  $D(\mathbf{x}, F_{\mathbf{X}})$  is any real-valued function that provides a 'center outward ordering' of  $\mathbf{x}$  with respect to  $F$  (?). ? outlines the desirable properties of a statistical depth function:

(P1) Affine invariance:  $D(A\mathbf{x} + \mathbf{b}, F_{A\mathbf{X} + \mathbf{b}}) = D(\mathbf{x}, F_{\mathbf{X}})$

(P2) Maximality at center:  $D(\boldsymbol{\theta}, F_{\mathbf{X}}) = \sup_{\mathbf{x} \in \mathbb{R}^p} D(\mathbf{x}, F_{\mathbf{X}})$  for  $F_{\mathbf{X}}$  having center of symmetry  $\boldsymbol{\theta}$ . This point is called the *deepest point* of the distribution.

(P3) Monotonicity with respect to deepest point:  $D(\mathbf{x}; F_{\mathbf{X}}) \leq D(\boldsymbol{\theta} + a(\mathbf{x} - \boldsymbol{\theta}), F_{\mathbf{X}})$ ,  $\boldsymbol{\theta}$  being deepest point of  $F_{\mathbf{X}}$ .

(P4) Vanishing at infinity:  $D(\mathbf{x}; F_{\mathbf{X}}) \rightarrow \mathbf{0}$  as  $\|\mathbf{x}\| \rightarrow \infty$ .

In (P2) the types of symmetry considered can be central symmetry, angular symmetry and halfspace symmetry. Also for multimodal probability distributions, properties (P2) and (P3) are actually restrictive towards the formulation of a reasonable depth function. In our derivations that follow, we replace these two by a slightly weaker condition: (P2\*) The *existence* of a maximal point, i.e.  $\sup_{\mathbf{x} \in \mathbb{R}^p} D(\mathbf{x}, F_{\mathbf{X}}) < \infty$ . We denote this point by  $M_D(F)$ .

A real-valued function measuring the outlyingness of a point with respect to the data cloud can be seen as the opposite of what data depth does. Indeed, such functions have been used to define several depth functions, for example simplicial volume depth, projection depth and  $L^p$ -depth. Here we give a general definition of such functions as a transformation on any depth function:

**Definition 2.1.** Given a random variable  $\mathbf{X}$  following a probability distribution  $F$ , and a depth function  $D(\cdot, \cdot)$ , we define Htped of a point  $\mathbf{x}$  as:  $\tilde{D}(\mathbf{x}, F) = h(d_{\mathbf{x}})$  as any function of the data depth so that  $h(d_{\mathbf{x}})$  is bounded, monotonically decreasing in  $d_{\mathbf{x}}$  and  $\sup_{\mathbf{x}} \tilde{D}(\mathbf{x}, F) < \infty$ .

For a fixed depth function, there are several choices of a corresponding htped. We develop our theory assuming a general htped function, but for the plots and simulations, fix our htped as  $\tilde{D}(\mathbf{x}, F) = M_D(F) - D(\mathbf{x}, F)$ , i.e. simply subtract the depth of a point from the maximum possible depth over all points in sample space.

We will be using the following 3 measures of data-depth to obtain our DCMs and compare their performances:

- **Halfspace depth** (HD) (?) is defined as the minimum probability of all halfspaces containing a point. In our notations,

$$HD(\mathbf{x}, F) = \inf_{\mathbf{u} \in \mathbb{R}^p; \mathbf{u} \neq \mathbf{0}} P(\mathbf{u}^T \mathbf{X} \geq \mathbf{u}^T \mathbf{x})$$

- **Mahalanobis depth** (MhD) (?) is based on the Mahalanobis distance of  $\mathbf{x}$  to  $\boldsymbol{\mu}$  with respect to  $\Sigma$ :  $d_{\Sigma}(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}$ . It is defined as

$$MhD(\mathbf{X}, F) = \frac{1}{1 + d_{\Sigma}^2(\mathbf{x} - \boldsymbol{\mu})}$$

note here that  $d_{\Sigma}(\mathbf{x}, \boldsymbol{\mu})$  can be seen as a valid depth function of  $\mathbf{x}$  with respect to  $F$ .

- **Projection depth** (PD) (?) is another depth function based on an outlyingness function. Here that function is

$$O(\mathbf{x}, F) = \sup_{\|\mathbf{u}\|=1} \frac{|\mathbf{u}^T \mathbf{x} - m(\mathbf{u}^T \mathbf{X})|}{s(\mathbf{u}^T \mathbf{X})}$$

where  $m$  and  $s$  are some univariate measures location and scale, respectively. Given this the depth at  $\mathbf{x}$  is defined as  $PD(\mathbf{x}, F) = 1/(1 + O(\mathbf{x}, F))$ .

Computation-wise, MhD is easy to calculate since the sample mean and covariance matrix are generally used as estimates of  $\mu$  and  $\Sigma$ , respectively. However this makes MhD less robust with respect to outliers. PD is generally approximated by taking maximum over a number of random projections. There have been several approaches for calculating HD. A recent unpublished paper (?) provides a general algorithm that computes exact HD in  $O(n^{p-1} \log n)$  time. In this paper, we shall use inbuilt functions in the R package `fda.usc` for calculating the above depth functions.

### 3 Depth-based rank covariance matrix

Consider a vector-valued random variable  $\mathbf{X} \in \mathbb{R}^p$ . Data depth is as much a property of the random variable as it is of the underlying distribution, so for ease of notation while working with transformed random variables, from now on we shall be using  $D_{\mathbf{X}}(\mathbf{x}) = D(\mathbf{x}, F)$  to denote the depth of a point  $\mathbf{x}$ . Now, given a depth function  $D_{\mathbf{X}}(\mathbf{x})$

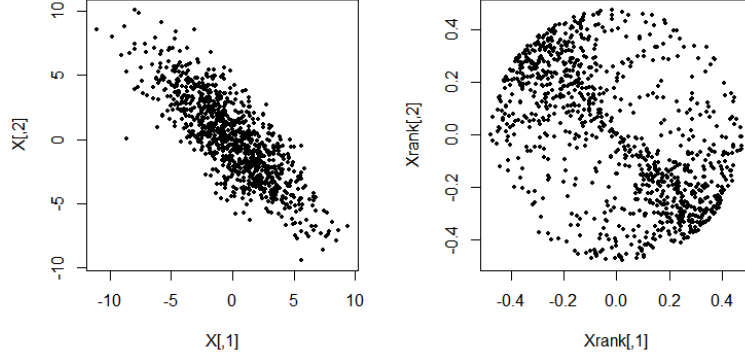


Figure 1: (Left) 1000 points randomly drawn from  $\mathcal{N}_2((0, 0)^T, \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix})$  and (Right) their multivariate ranks based on halfspace depth

(equivalently, an htped function  $\tilde{D}_{\mathbf{X}}(\mathbf{x}) = \tilde{D}(\mathbf{x}, F)$ ), 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. The transformed random variable  $\tilde{\mathbf{X}}$  can be seen as the multivariate rank corresponding to  $\mathbf{X}$  (e.g. ?). Figure 1 gives an idea of how  $\tilde{\mathbf{X}}$  is distributed when  $\mathbf{X}$  has a bivariate normal distribution. Compared to the spatial sign, which are distributed on the surface of  $p$ -dimensional unit ball centered at  $\boldsymbol{\mu}$ , these spatial ranks have the same direction as original data and reside *inside* the  $p$ -dimensional ball around  $\boldsymbol{\mu}$  that has radius  $M_D(F)$  (which, for the case of halfspace depth, equals 0.5).

Now consider the spectral decomposition for the covariance matrix of  $F$ :  $\Sigma = \Gamma\Lambda\Gamma^T$ ,  $\Gamma$  being orthogonal and  $\Lambda$  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

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

$\tilde{D}_{\mathbf{z}}(\mathbf{z})$  is an even 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*

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

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,  $\Gamma$ , 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  $\Lambda$ , 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 (?) one can get back the shape components, i.e. original *standardized* eigenvalues  $\Lambda^*$  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 (?), however, attempting to derive an affine equivariant counterpart (as opposed to only orthogonal equivariance) of the DCM through an iterative approach analogous to ? will not result in anything new. This is because Tyler's scatter matrix  $\Sigma_T$  is defined as the implicit solution to the following equation:

$$\Sigma_T = E \left[ \frac{\mathbf{x}\mathbf{x}^T}{\mathbf{x}^T \Sigma_T^{-1} \mathbf{x}} \right] \quad (3)$$

and simply replacing  $\mathbf{x}$  by its multivariate rank counterpart  $\tilde{\mathbf{x}}$  will not change the estimate  $\Sigma_T$  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 (3)) in the simulations in Section 5. The simulations show that it has slightly better finite-sample efficiency than  $\Sigma_T$  but has same asymptotic performance. We conjecture that its concentration properties can be obtained by taking an approach similar to ?.

## 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 function 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 (??), Oja median (?), projection median (?) 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  $V_{D,S}(F)$  explicitly specifying each of its elements (more directly those of its  $\Gamma^T$ -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  $\Sigma$  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 (?).

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.** (?) *Let  $F_\Lambda$  be an elliptical distribution with a diagonal covariance matrix  $\Lambda$ , and  $\hat{C}$  be any positive definite symmetric  $p \times p$  matrix such that at  $F_\Lambda$  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*

$$\sqrt{n}\text{vec}(\hat{C} - \Lambda) = [(\Lambda \otimes I_p) - (I_p \otimes \Lambda)] \sqrt{n}\text{vec}(\hat{P} - I_p) + \sqrt{n}\text{vec}(\hat{\Lambda} - \Lambda) + o_P(1) \quad (4)$$

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:

$$AVar(\mathbf{g}_i) = \sum_{k=1; k \neq i}^p \frac{1}{(\lambda_{D,s,k} - \lambda_{D,S,i})^2} E \left[ \frac{(\tilde{D}\mathbf{z}(\mathbf{z}))^4 \lambda_i \lambda_k z_i^2 z_k^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] \gamma_k \gamma_k^T \quad (5)$$

$$ACov(\mathbf{g}_i, \mathbf{g}_j) = -\frac{1}{(\lambda_{D,s,i} - \lambda_{D,S,j})^2} E \left[ \frac{(\tilde{D}\mathbf{z}(\mathbf{z}))^4 \lambda_i \lambda_j z_i^2 z_j^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] \gamma_j \gamma_i^T; \quad i \neq j \quad (6)$$

where  $\Gamma = (\gamma_1, \dots, \gamma_p)$ . 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:

$$AVar(l_i) = E \left[ \frac{(\tilde{D}\mathbf{z}(\mathbf{z}))^4 \lambda_i^2 z_i^4}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] - \lambda_{D,S,i}^2 \quad (7)$$

$$ACov(l_i, l_j) = E \left[ \frac{(\tilde{D}\mathbf{z}(\mathbf{z}))^4 \lambda_i \lambda_j z_i^2 z_j^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] - \lambda_{D,S,i} \lambda_{D,S,j}; \quad i \neq j \quad (8)$$

## 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 distribution (?). 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. ??). 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  $F_\epsilon$  is  $F$  with an additional mass of  $\epsilon$  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 ?, we now get the influence function of the  $i^{\text{th}}$  column of  $\hat{\Gamma}_D = (\hat{\gamma}_{D,1}, \dots, \hat{\gamma}_{D,p})$ ;  $i = 1, \dots, p$ :

$$\begin{aligned} IF(\mathbf{x}_0; \hat{\gamma}_{D,i}, F) &= \sum_{k=1; k \neq i}^p \frac{1}{\lambda_{D,S,i} - \lambda_{D,S,k}} \left\{ \gamma_k^T IF(\mathbf{x}_0; Cov(\tilde{\mathbf{X}}), \gamma_i) \right\} \gamma_k \\ &= \sum_{k=1; k \neq i}^p \frac{1}{\lambda_{D,S,i} - \lambda_{D,S,k}} \left\{ \gamma_k^T (\tilde{D}_{\mathbf{X}}(\mathbf{x}_0))^2 SS(\mathbf{x}_0; \boldsymbol{\mu}) \gamma_i - \lambda_{D,S,i} \gamma_k^T \gamma_i \right\} \gamma_k \\ &= \sum_{k=1; k \neq i}^p \frac{\sqrt{\lambda_i \lambda_k} z_{0i} z_{0k}}{\lambda_{D,S,i} - \lambda_{D,S,k}} \cdot \frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}_0))^2}{\mathbf{z}_0^T \Lambda \mathbf{z}_0} \gamma_k \end{aligned} \quad (9)$$

where  $\Gamma^T \Lambda^{-1/2}(\mathbf{x}_0 - \boldsymbol{\mu}) = \mathbf{z}_0 = (z_{01}, \dots, z_{0p})^T$ . Clearly this influence function will be bounded, which indicates good robustness properties of principal components. Moreover, since the htped function takes small values for points close to the center of the distribution, it does not suffer from the inlier effect that is typical of the SCM and Tyler's shape matrix. The influence function for the  $i^{\text{th}}$  eigenvector estimates of these two matrices (say  $\hat{\gamma}_{S,i}$

and  $\hat{\gamma}_{T,i}$ , respectively) are as follows:

$$IF(\mathbf{x}_0; \hat{\gamma}_{S,i}, F) = \sum_{k=1; k \neq i}^p \frac{\sqrt{\lambda_i \lambda_k}}{\lambda_{S,i} - \lambda_{S,k}} \cdot \frac{z_{0i} z_{0k}}{\mathbf{z}_0^T \Lambda \mathbf{z}_0} \gamma_k, \text{ with } \lambda_{S,i} = E\mathbf{Z} \left( \frac{\lambda_i z_i^2}{\sum_{j=1}^p \lambda_j z_j^2} \right)$$

$$IF(\mathbf{x}_0; \hat{\gamma}_{T,i}, F) = (p+2) \sum_{k=1; k \neq i}^p \frac{\sqrt{\lambda_i \lambda_k}}{\lambda_i - \lambda_k} \cdot \frac{z_{0i} z_{0k}}{\mathbf{z}_0^T \mathbf{z}_0} \gamma_k$$

for  $i = 1, 2$ . In Figure 2 we consider first eigenvectors of different scatter estimates for the  $\mathcal{N}_2((0, 0)^T, \text{diag}(2, 1))$  and plot norms of these influence functions for different values of  $\mathbf{x}_0$ . The plots for SCM and Tyler's shape matrix demonstrate the 'inlier effect', i.e. points close to symmetry center and the center itself having high influence. The influence function for the sample covariance matrix is obtained by replacing  $(p+2)$  by  $\|\mathbf{z}_0\|^2$  in the expression of  $IF(\mathbf{x}_0; \hat{\gamma}_{T,i}, F)$  above, hence is unbounded and the corresponding eigenvector estimators are not robust. In comparison, all three DCMs considered here have a bounded influence function as well as small values of the influence function at 'deep' points.

## 5.2 Asymptotic and finite-sample efficiencies

Suppose  $\hat{\Sigma}$  is a  $\sqrt{n}$ -consistent estimator of the population covariance matrix  $\Sigma$ , 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 ?)

$$AVar(\sqrt{n} \hat{\gamma}_i) = \sum_{k=1; k \neq i}^p \frac{\lambda_i \lambda_k}{(\lambda_i - \lambda_k)^2} \gamma_k \gamma_k^T \quad (10)$$

the asymptotic relative efficiencies of eigenvectors from the sample DCM with respect to the sample covariance matrix can now be derived using (10) above and (5) from Corol-

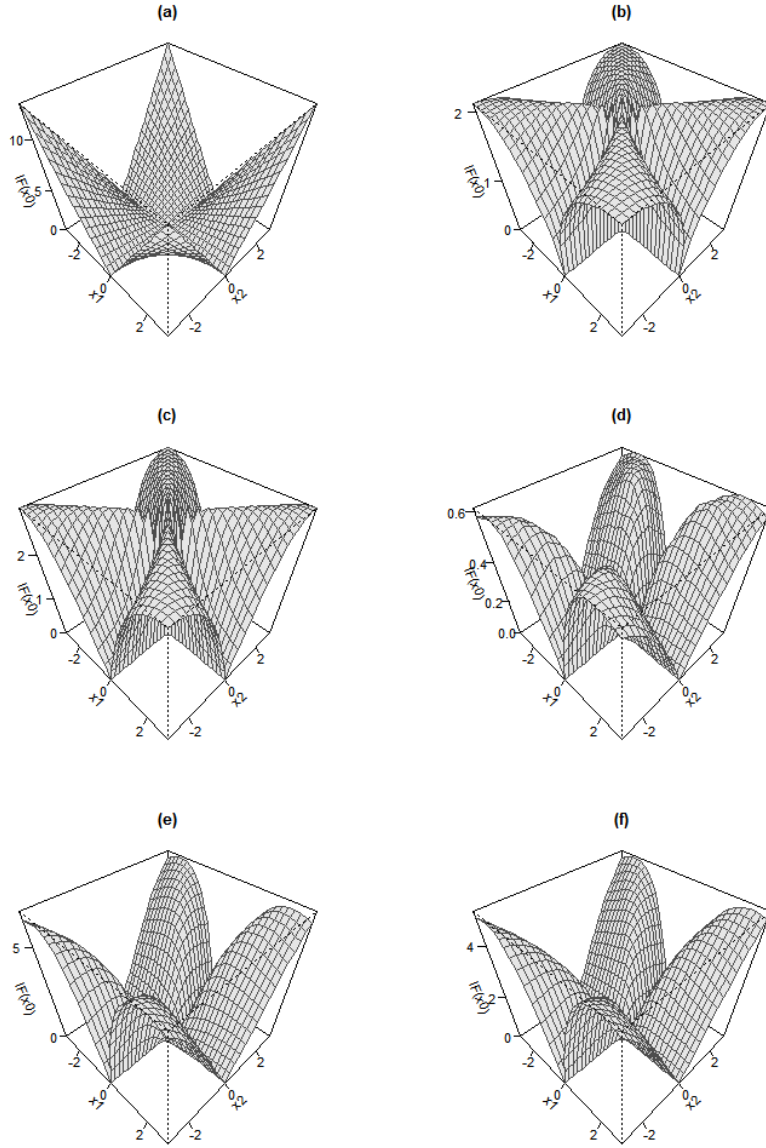


Figure 2: Plot of the norm of influence function for first eigenvector of (a) sample covariance matrix, (b) SCM, (c) Tyler's scatter matrix and DCMs for (d) Halfspace depth, (e) Mahalanobis depth, (f) Projection depth for a bivariate normal distribution with  $\mu = \mathbf{0}$ ,  $\Sigma = \text{diag}(2, 1)$

lary 4.4:

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

For 2 dimensions, this expression can be somewhat simplified. Suppose the two eigenvalues are  $\lambda$  and  $\rho\lambda$ . In that case the eigenvalues of the DCM are

$$\lambda_{D,S,1} = E \left( \frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 z_1^2}{z_1^2 + \rho z_2^2} \right), \quad \lambda_{D,S,2} = E \left( \frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 \rho z_2^2}{z_1^2 + \rho z_2^2} \right)$$

and by simple algebra we get

$$ARE(\hat{\gamma}_1^D, \hat{\gamma}_1; F) = ARE(\hat{\gamma}_2^D, \hat{\gamma}_2; F) = \frac{1}{(1 - \rho)^2} \frac{E^2 \left( \frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 (z_1^2 - \rho z_2^2)}{(z_1^2 + \rho z_2^2)} \right)}{E \left( \frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 z_1^2 z_2^2}{(z_1^2 + \rho z_2^2)^2} \right)}$$

For  $\rho = 0.5$ , Table 1 below considers 6 different elliptic distributions (namely, bivariate  $t$  with  $\text{df} = 5, 6, 10, 15, 25$  and bivariate normal), and summarizes the ARE for first eigenvector of the SCM, Tyler's scatter matrix and DCM for 3 choices of depth function (HSD-CM, MhD-CM, PD-CM: columns 3-5), as well as their depth-weighted versions mentioned at the end of section 3 (HSD-wCM, MhD-wCM, PD-wCM: columns 6-8). The SCM and Tyler's M-estimator perform better than the sample covariance matrix only for a bivariate  $t$ -distribution with  $\text{df} = 5$ . Estimates based on depth-based covariance matrices have much better performances for all distributions, and continue to be competitive of the sample covariance matrix estimates as the base distribution approaches normality, especially those based on projection depth. Interestingly, depth-weighted versions seem to have better performances than their corresponding DCMs, more so for heavy-tailed distributions.

We now obtain finite sample efficiencies of the three DCMs as well as their depth-weighted affine equivariant counterparts by a simulation study, and compare them with



	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
Bivariate $t_5$	1.46	1.50	1.97	1.54	1.99	2.09	1.57	2.11
Bivariate $t_6$	0.97	1.00	1.37	1.12	1.40	1.45	1.19	1.49
Bivariate $t_{10}$	0.65	0.67	0.96	0.88	1.02	1.02	0.94	1.10
Bivariate $t_{15}$	0.57	0.59	0.87	0.82	0.94	0.92	0.87	1.00
Bivariate $t_{25}$	0.54	0.55	0.79	0.75	0.88	0.85	0.81	0.95
BVN	0.49	0.50	0.73	0.71	0.82	0.77	0.75	0.88

Table 1: Asymptotic efficiencies relative to sample covariance matrix for  $p = 2$

the same from the SCM and Tyler's scatter matrix. We consider the same 6 elliptical distributions considered in ARE calculations above, and from every distribution draw 10000 samples each for sample sizes  $n = 20, 50, 100, 300, 500$ . All distributions are centered at  $\mathbf{0}_p$ , and have covariance matrix  $\Sigma = \text{diag}(p, p-1, \dots, 1)$ . We consider 3 choices of  $p$ : 2, 3 and 4.

We use the concept of principal angles (?) to find out error estimates for the first eigenvector of a scatter matrix. In our case, the first eigenvector will be

$$\gamma_1 = (1, \overbrace{0, \dots, 0}^{p-1})^T$$

For an estimate of the eigenvector, say  $\hat{\gamma}_1$ , error in prediction is measured by the smallest angle between the two lines, i.e.  $\cos^{-1} |\gamma_1^T \hat{\gamma}_1|$ . A smaller absolute value of this angle is equivalent to better prediction. We repeat this 10000 times and calculate the **Mean Squared Prediction Angle**:

$$MSPA(\hat{\gamma}_1) = \frac{1}{10000} \sum_{m=1}^{10000} \left( \cos^{-1} \left| \gamma_1^T \hat{\gamma}_1^{(m)} \right| \right)^2$$

Finally, the finite sample efficiency of some eigenvector estimate  $\hat{\gamma}_1^E$  relative to that obtained from the sample covariance matrix, say  $\hat{\gamma}_1^{Cov}$  is obtained as:

$$FSE(\hat{\gamma}_1^E, \hat{\gamma}_1^{Cov}) = \frac{MSPA(\hat{\gamma}_1^{Cov})}{MSPA(\hat{\gamma}_1^E)}$$

Tables 2, 3 and 4 give FSE values for  $p = 2, 3, 4$ , respectively. In general, all the effi-

ciencies increase as the dimension  $p$  goes up. DCM-based estimators (columns 3-5 in each table) outperform SCM and Tyler’s scatter matrix, and among the 3 depths considered, projection depth seems to give the best results. Its finite sample performances are better than Tyler’s and Huber’s M-estimators of scatter as well as their symmetrized counterparts (see Table 4 in ?, and quite close to the affine equivariant spatial sign covariance matrix (see Table 2 in ?) For  $p = 2$ ,  $n = 300, 500$  the first 5 columns of Table 2 approximate the asymptotic efficiencies in Table 1 well, except for the multivariate  $t$ -distribution with  $df = 5$ . Finally, the depth-weighted iterated versions of these 3 SCMs (columns 6-8 in each table) seem to further better the performance of their corresponding orthogonal equivariant counterparts.

## 6 Conclusion

In the above sections we introduce a covariance matrix based on depth-based multivariate ranks that keeps the eigenvectors of the actual population unchanged for elliptical distributions. We provide asymptotic results for the sample DCM, its eigenvalues and eigenvectors. Bounded influence functions as well as simulation studies suggest that the eigenvector estimates obtained from the DCM are highly robust, yet do not lose much in terms of efficiency. Thus it provides a plausible alternative to existing approaches of robust PCA that are based on estimation of covariance matrices (for example SCM, Tyler’s scatter matrix, Dümbgen’s symmetrized shape matrix).

An immediate extension of this would be to study the depth-weighted iterated scatter matrices, i.e. matrices  $\Sigma_{Dw}$  that are solution to the following type of equations:

$$\Sigma_{Dw} = E \left[ \frac{(\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2 \mathbf{x} \mathbf{x}^T}{\mathbf{x}^T \Sigma_{Dw}^{-1} \mathbf{x}} \right]$$

as they seem to perform better than those obtained from the corresponding DCMs. Unlike the DCM, these matrices will possess the affine equivariance property. It is possible to develop tests for central and elliptic symmetry based on the decomposition of the

$F = \text{Bivariate } t_5$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.80	0.83	0.95	0.95	0.89	1.00	0.96	0.89
$n=50$	0.86	0.90	1.25	1.10	1.21	1.32	1.13	1.25
$n=100$	1.02	1.04	1.58	1.20	1.54	1.67	1.24	1.63
$n=300$	1.24	1.28	1.81	1.36	1.82	1.93	1.44	1.95
$n=500$	1.25	1.29	1.80	1.33	1.84	1.91	1.39	1.97
$F = \text{Bivariate } t_6$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.77	0.79	0.92	0.92	0.86	0.96	0.92	0.85
$n=50$	0.76	0.78	1.11	1.00	1.08	1.17	1.03	1.13
$n=100$	0.78	0.79	1.27	1.06	1.33	1.35	1.11	1.41
$n=300$	0.88	0.91	1.29	1.09	1.35	1.38	1.15	1.45
$n=500$	0.93	0.96	1.37	1.13	1.40	1.44	1.19	1.48
$F = \text{Bivariate } t_{10}$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.70	0.72	0.83	0.84	0.77	0.89	0.87	0.79
$n=50$	0.58	0.60	0.90	0.84	0.86	0.95	0.88	0.91
$n=100$	0.57	0.59	0.92	0.87	0.97	0.98	0.90	1.03
$n=300$	0.62	0.64	0.93	0.85	0.99	0.99	0.91	1.06
$n=500$	0.62	0.65	0.93	0.86	1.00	1.00	0.92	1.08
$F = \text{Bivariate } t_{15}$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.63	0.66	0.76	0.78	0.72	0.81	0.81	0.73
$n=50$	0.52	0.52	0.79	0.75	0.80	0.84	0.79	0.85
$n=100$	0.51	0.52	0.83	0.77	0.88	0.88	0.81	0.94
$n=300$	0.55	0.56	0.84	0.79	0.91	0.89	0.84	0.98
$n=500$	0.56	0.59	0.85	0.80	0.93	0.91	0.86	0.99
$F = \text{Bivariate } t_{25}$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.63	0.65	0.77	0.79	0.74	0.80	0.81	0.74
$n=50$	0.49	0.50	0.73	0.71	0.76	0.78	0.75	0.80
$n=100$	0.45	0.46	0.73	0.69	0.81	0.78	0.73	0.87
$n=300$	0.51	0.52	0.78	0.75	0.87	0.83	0.79	0.94
$n=500$	0.53	0.55	0.79	0.75	0.87	0.84	0.80	0.94
$F = \text{BVN}$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.56	0.60	0.69	0.71	0.67	0.73	0.74	0.68
$n=50$	0.42	0.43	0.66	0.66	0.70	0.71	0.69	0.75
$n=100$	0.42	0.43	0.69	0.66	0.77	0.74	0.71	0.83
$n=300$	0.47	0.49	0.71	0.69	0.82	0.76	0.73	0.88
$n=500$	0.48	0.50	0.73	0.71	0.83	0.78	0.76	0.89

Table 2: Finite sample efficiencies of several scatter matrices:  $p = 2$

3-variate $t_5$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.96	0.97	1.06	1.03	0.99	1.07	1.06	0.97
$n=50$	1.07	1.08	1.28	1.20	1.18	1.33	1.23	1.20
$n=100$	1.12	1.15	1.49	1.31	1.40	1.57	1.38	1.48
$n=300$	1.49	1.54	2.09	1.82	2.07	2.19	1.93	2.18
$n=500$	1.60	1.66	2.18	1.87	2.21	2.27	1.95	2.30
3-variate $t_6$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.90	0.92	1.00	0.99	0.95	1.02	1.01	0.94
$n=50$	0.95	0.96	1.16	1.09	1.09	1.21	1.14	1.11
$n=100$	0.98	0.99	1.32	1.22	1.25	1.38	1.27	1.29
$n=300$	1.10	1.14	1.57	1.40	1.58	1.62	1.47	1.64
$n=500$	1.17	1.20	1.57	1.43	1.60	1.63	1.51	1.67
3-variate $t_{10}$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.87	0.88	0.95	0.94	0.90	0.97	0.98	0.89
$n=50$	0.77	0.79	0.96	0.92	0.94	0.99	0.96	0.95
$n=100$	0.75	0.76	1.02	0.95	1.01	1.06	1.00	1.05
$n=300$	0.73	0.75	1.03	0.98	1.10	1.08	1.03	1.15
$n=500$	0.73	0.76	1.02	0.98	1.09	1.06	1.02	1.14
3-variate $t_{15}$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.84	0.86	0.92	0.92	0.89	0.94	0.94	0.87
$n=50$	0.75	0.76	0.92	0.90	0.90	0.96	0.94	0.93
$n=100$	0.66	0.67	0.91	0.87	0.95	0.96	0.92	1.00
$n=300$	0.61	0.64	0.90	0.87	1.00	0.93	0.91	1.04
$n=500$	0.65	0.67	0.89	0.87	0.99	0.93	0.91	1.03
3-variate $t_{25}$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.78	0.79	0.87	0.89	0.87	0.89	0.92	0.86
$n=50$	0.70	0.71	0.88	0.86	0.88	0.91	0.90	0.90
$n=100$	0.61	0.63	0.86	0.83	0.89	0.90	0.88	0.94
$n=300$	0.58	0.59	0.83	0.80	0.92	0.87	0.85	0.98
$n=500$	0.62	0.64	0.83	0.82	0.94	0.88	0.87	0.99
3-variate Normal	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.76	0.78	0.85	0.87	0.84	0.87	0.90	0.83
$n=50$	0.66	0.67	0.82	0.81	0.84	0.86	0.86	0.86
$n=100$	0.56	0.58	0.77	0.75	0.83	0.82	0.79	0.87
$n=300$	0.53	0.55	0.75	0.74	0.85	0.79	0.78	0.90
$n=500$	0.56	0.58	0.76	0.76	0.87	0.80	0.80	0.92

Table 3: Finite sample efficiencies of several scatter matrices:  $p = 3$

4-variate $t_5$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	1.04	1.02	1.10	1.07	1.02	1.09	1.07	0.98
$n=50$	1.08	1.08	1.16	1.16	1.13	1.19	1.19	1.13
$n=100$	1.31	1.31	1.42	1.38	1.36	1.46	1.44	1.36
$n=300$	1.46	1.54	1.81	1.76	1.95	1.88	1.88	1.95
$n=500$	1.92	1.93	2.23	2.03	2.31	2.35	2.19	2.39
4-variate $t_6$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	1.00	1.05	1.03	1.05	1.00	1.04	1.04	0.95
$n=50$	1.03	1.01	1.13	1.12	1.11	1.19	1.17	1.10
$n=100$	1.08	1.12	1.25	1.23	1.27	1.24	1.25	1.22
$n=300$	1.34	1.36	1.64	1.52	1.60	1.67	1.61	1.68
$n=500$	1.26	1.34	1.55	1.49	1.60	1.65	1.61	1.69
4-variate $t_{10}$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.90	0.89	0.95	0.98	0.98	0.96	1.01	0.95
$n=50$	0.90	0.91	1.01	0.98	0.98	1.03	1.04	0.99
$n=100$	0.87	0.87	0.93	0.95	1.01	0.99	1.01	1.05
$n=300$	0.87	0.87	1.09	1.09	1.17	1.14	1.16	1.23
$n=500$	0.88	0.92	1.10	1.10	1.23	1.19	1.22	1.29
4-variate $t_{15}$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.92	0.90	0.94	0.94	0.96	0.95	0.97	0.89
$n=50$	0.82	0.83	0.88	0.91	0.93	0.88	0.93	0.93
$n=100$	0.84	0.87	0.92	0.95	1.00	0.93	0.96	1.00
$n=300$	0.73	0.75	0.96	0.99	1.10	1.00	1.06	1.12
$n=500$	0.73	0.76	0.95	0.96	1.06	0.94	0.97	1.06
4-variate $t_{25}$	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.89	0.92	0.92	0.92	0.90	0.96	0.95	0.89
$n=50$	0.82	0.84	0.89	0.90	0.91	0.93	0.96	0.92
$n=100$	0.77	0.76	0.90	0.90	0.96	0.94	0.98	1.04
$n=300$	0.73	0.77	0.93	0.91	0.98	1.00	0.98	1.03
$n=500$	0.67	0.71	0.83	0.83	0.96	0.88	0.90	1.00
4-variate Normal	SCM	Tyler	HSD-CM	MhD-CM	PD-CM	HSD-wCM	MhD-wCM	PD-wCM
$n=20$	0.82	0.84	0.87	0.90	0.91	0.89	0.93	0.89
$n=50$	0.80	0.81	0.87	0.88	0.88	0.88	0.92	0.88
$n=100$	0.68	0.71	0.80	0.85	0.91	0.82	0.86	0.92
$n=300$	0.61	0.63	0.82	0.85	0.93	0.86	0.91	0.96
$n=500$	0.60	0.64	0.77	0.80	0.90	0.82	0.86	0.96

Table 4: Finite sample efficiencies of several scatter matrices:  $p = 4$

multivariate rank vector in 1. The result in Theorem 3.1 is based on the fact that  $E(\tilde{\mathbf{X}}) = \mathbf{0}$ , and it holds for any centrally symmetric underlying distribution. Moreover, the depth of a point in the standardized scale (i.e.  $\mathbf{z}$ -scale) does not depend on the direction of  $\mathbf{z}$ . This is not possible without circular symmetry of  $\mathbf{z}$ , so any test of independence between  $D_{\mathbf{Z}}(\mathbf{z})$  and  $\mathbf{S}(\mathbf{z})$  can be seen as a test of ellipticity for the original random variable  $\mathbf{X}$ . Finally the applicability of this procedure in high-dimensional and functional data remains to be seen.

## 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  $F_{\Lambda}$  has the same elliptic distribution as  $F$ , but with covariance matrix  $\Lambda$ . Now,

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

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  $E[\sqrt{\lambda_a \lambda_b \lambda_c \lambda_d} z_a z_b z_c z_d (\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 / (\mathbf{z}^T \Lambda \mathbf{z})^2]$ , 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 ?, 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  $\gamma_{mn}^D = E[\lambda_m \lambda_n z_m^2 z_n^2 (\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 / (\mathbf{z}^T \Lambda \mathbf{z})^2]; 1 \leq m, n \leq p$ .

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.** The proof follows directly from writing out the expression of  $Cov(\tilde{\mathbf{X}})$ :

$$\begin{aligned} Cov(\tilde{\mathbf{X}}) &= E(\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T) - E(\tilde{\mathbf{X}})E(\tilde{\mathbf{X}})^T \\ &= \Gamma.E \left[ (\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 \frac{\|\mathbf{z}\|^2}{\|\Lambda^{1/2}\mathbf{z}\|} \Lambda^{1/2} \mathbf{S}(\mathbf{z}) \mathbf{S}(\mathbf{z})^T \Lambda^{1/2} \right] \Gamma^T - \mathbf{0}_p \mathbf{0}_p^T \\ &= \Gamma.E \left[ (\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 \frac{\Lambda^{1/2} \mathbf{z} \mathbf{z}^T \Lambda^{1/2}}{\mathbf{z}^T \Lambda \mathbf{z}} \right] \Gamma^T \end{aligned}$$

□

**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

$$\begin{aligned} P \left[ |(\tilde{D}_{\mathbf{X}}^{n_1}(\mathbf{x}))^2 - (\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2| > \epsilon \right] &< \delta \\ \Rightarrow |(\tilde{D}_{\mathbf{X}}^{n_1}(\mathbf{x}))^2 - (\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2| &< \epsilon(1 - \delta) + M\delta \\ &< \epsilon + M\delta \end{aligned} \tag{11}$$

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

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

$$C_k = \sqrt{N_k} \left[ \frac{1}{N_k} \sum_{i=1}^{N_k} SS(\mathbf{x}_i; \boldsymbol{\mu}) \right]$$

where  $N_k \in \mathbb{N}$  gives the relation (11) 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 with unknown location parameter  $\hat{\boldsymbol{\mu}}_n$  has the same asymptotic distribution as the same with known location  $\boldsymbol{\mu}$  (?), hence  $B_k \xrightarrow{P} 0$ . Now by Slutsky's theorem,  $A_k(B_k + C_k) \xrightarrow{P} 0$ . Finally, for any  $\epsilon_1 > 0$ ,  $P(S_{N_k} > \epsilon_1) < P(A_k B_k > \epsilon_1)$ , 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:

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

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

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 ?  $\square$

**Proof of Corollary 4.4.** In spirit, this corollary is similar to Theorem 13.5.1 in ?, and indeed, ? used this theorem to prove Theorem 4.3. Due to the decomposition (4) we

have, for the distribution  $F_\Lambda$ , 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

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

$\sqrt{n}(\hat{\gamma}_{D,ii}(F_\Lambda) - 1) = o_P(1)$  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  $i \neq j$ ,

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

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