

# Signed Peripherality Functions for Multivariate Statistical Inference

Subhabrata Majumdar, Snigdhasu Chatterjee

Abstract: (**will change**) 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 shapes of these influence functions indicate robustness of estimated 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. Through two real data examples, we also show the effectiveness of DCM-based PCA in analyzing high-dimensional data and outlier detection, and compare it with other methods of robust PCA.

Keywords: Multivariate ranking, data depth, dimension reduction, principal components analysis, robust statistics, functional data

# 1 Introduction

Consider a point  $\mathbf{x}$  from some  $\mathcal{X} \subseteq \mathbb{R}^p$ , and the spatial sign function  $\mathbf{S} : \mathcal{X} \times \mathbb{R}^p \rightarrow \mathbb{R}^p$ , defined as

$$\mathbf{S}(\mathbf{x}; \boldsymbol{\mu}_x) = \frac{\mathbf{x} - \boldsymbol{\mu}_x}{\|\mathbf{x} - \boldsymbol{\mu}_x\|} \mathbb{I}(\mathbf{x} \neq \boldsymbol{\mu}_x),$$

where  $\boldsymbol{\mu}_x \in \mathbb{R}^p$  is a fixed location parameter,  $\mathbb{I}(\cdot)$  is the 0/1 indicator function and  $\|\cdot\|$  is the euclidean norm. This is a generalization of the real-valued case of the indicator for whether a point  $x \in \mathbb{R}$  is to the right, left or at a scalar  $\mu_x \in \mathbb{R}$ , and was first introduced by Möttönen and Oja (1995). Given a data matrix  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ , a popular approach of using spatial signs in the literature has been to take spatial signs of the data points, i.e.  $\mathbf{S}(\mathbf{x}_i)$ , and use the matrix of signs as a proxy of  $\mathbf{X}$  to formulate robust versions of location and scale problems (Locantore et al., 1999; Oja, 2010; Wang et al., 2015).

The spatial sign maps all data points to the surface of a  $p$ -dimensional unit sphere (see Figure 1). While this transformation has its advantages (i.e. robustness), e.g. the eigenvectors remain the same while limiting the influence of outliers when the points are drawn from an elliptical distribution (Taskinen et al., 2012), valuable information is lost in the form of magnitudes of sample points. As a result, spatial sign-based procedures suffer from low efficiency. For example, eigenvector estimates obtained from the covariance matrix of spatial sign, which is called the Sign Covariance Matrix (SCM), are asymptotically inadmissible (Magyar and Tyler, 2014)- in the sense that there is an estimator (namely, Tyler’s M-estimate of scatter (Tyler, 1987)) that has uniformly lower asymptotic risk than the SCM.

To utilize the magnitude information of sample points, at the same time preserving the robustness property, we propose a general class weighted spatial sign functions. To this end, we first propose what we call a *peripherality function*.

**Definition 1.1.** For a set of probability measures  $\mathcal{M}$  taking values in  $\mathbb{R}^p$ , a *peripherality function*  $P : \mathbb{R}^p \times \mathcal{M} \rightarrow \mathbb{R}^+ \cup \{0\}$ , is a function that satisfies the following condition:

(P) For every probability measure  $\mathbb{F} \in \mathcal{M}$ , there exists a constant  $\boldsymbol{\mu}_F \in \mathbb{R}^p$  such that for any  $t \in [0, 1]$  and  $\mathbf{x} \in \mathbb{R}^p$ ,

$$P(\boldsymbol{\mu}_F; \mathbb{F}) \leq P(\boldsymbol{\mu}_F + t(\mathbf{x} - \boldsymbol{\mu}_F); \mathbb{F}).$$

That is, for every fixed  $\mathbb{F}$ , the peripherality function achieves a minimum at  $\boldsymbol{\mu}_F$ , and is non-decreasing in every direction away from  $\boldsymbol{\mu}_F$ .

The peripherality function quantifies the relative position of  $\mathbf{x}$  with respect to  $\boldsymbol{\mu}_F$ , and condition (P) formalizes a general notion of centrality. A whole class of such functions can be defined from *Data Depth* functions (Zuo and Serfling, 2000), which give the center-outward ranking of points in a multivariate dataset. Data depth has seen extensive use in the past decade, ranging from robust nonparametric classification (Dutta and Ghosh, 2012; Ghosh and Chaudhuri, 2005; Jornsten, 2004; Sguera et al., 2014) to parametric estimation of means (Zuo et al., 2004), covariance matrices (Zuo and Cui, 2005) and sparse

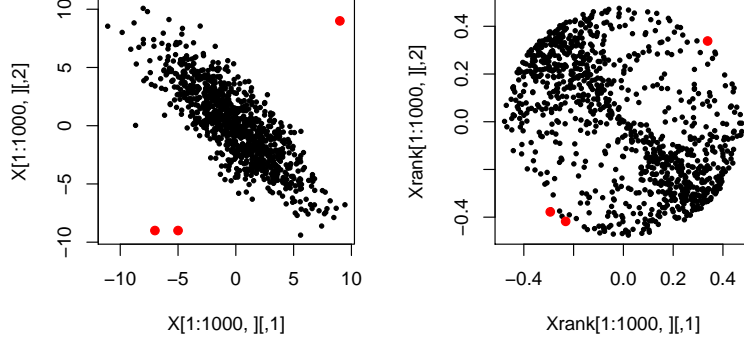


Figure 1: (Left) 1000 points randomly drawn from  $\mathcal{N}_2((0, 0)^T, \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix})$  and (Right) their signed peripherality functions based on halfspace depth

multivariate regression coefficients (Dutta and Genton, 2017; Majumdar and Chatterjee, 2018). Peripherality functions can be defined as some inverse ranking based on data depth, and the concept of *outlyingness* associated with data depth is essentially same as what we use in this paper. However, traditional depth functions require a wider array of theoretical conditions (Zuo and Serfling, 2000) in addition to our condition (P), and a sizeable portion of methods we discuss do not require the use of depth functions *per se* to be implemented.

Our weighted sign vector, which we call the *signed peripherality* function, is now defined as the product of the spatial sign and transformed peripherality functions:

$$\mathbf{R}(\mathbf{x}; \boldsymbol{\mu}_x, \mathbb{F}) = \mathbf{S}(\mathbf{x}, \boldsymbol{\mu}_x) \cdot \kappa_P(P(\mathbf{x}, \mathbb{F})), \quad (1.1)$$

where  $\kappa_P : \mathbb{R} \rightarrow \mathbb{R}$  is a monotone function. Its purpose is to either downweight or upweight the spatial sign of a point based on its peripherality with respect to a data cloud. As seen in the third panel of Figure 1,  $\mathbf{R}(\mathbf{x}; \boldsymbol{\mu}_x, \mathbb{F})$  maps points *inside* a ball of finite radius, thus limiting the influence of outlying points *and* preserving the magnitude information of data points. In this paper, we demonstrate certain interesting applications of this multivariate rank transformation in robust statistical inference. Notice that for the trivial choice  $P(\mathbf{x}; \mathbb{F}) = \|\mathbf{x} - \boldsymbol{\mu}_F\|$  and  $\boldsymbol{\mu}_x = \boldsymbol{\mu}_F, \kappa_P(x) = x$ , we get  $\mathbf{R}(\mathbf{x}; \boldsymbol{\mu}_x, \mathbb{F}) = \mathbf{x}$ . However, in this paper we illustrate how using the rank-transformed data in place of the original data points or their spatial signs leads to interesting robustness results for other choices of the peripherality functions.

The notion of multivariate ranks goes back to Puri and Sen (1971), where they take the vector consisting of marginal univariate ranks as multivariate rank vector. Subsequent definitions of multivariate ranks were proposed by Hallin and Paindaveine (2002); Möttönen and Oja (1995) and Chernozhukov et al. (2014). Compared to them, the advantages of our signed peripherality function are two-fold: interpretability and flexibility. Firstly, as evident in Figure 1, our rank transformation preserves the general shape of the data. In fact, if the map  $\|\mathbf{x}\| \mapsto P(\mathbf{x}, \mathbb{F})$  is one-one (given a fixed  $\mathbb{F}$ ), then it is possible

to get back  $\mathbf{x}$  from  $\mathbf{R}(\mathbf{x}; \boldsymbol{\mu}_x, \mathbb{F})$ . It also provides an intuitive extension to any spatial sign-based methodology. Secondly, there is considerable scope of generalization within the present framework. Given any general inner product space, the definition of signed peripherality remains the same when euclidean norm is replaced by the norm induced by the inner product. Even within  $\mathbb{R}^p$ , it is possible to generalize the ranking function in (1.1) by adding separate transformations on the sign and peripherality functions to obtain alternate representations of the data in lower or higher dimensional spaces. However, we limit our discussion in this paper to the original formulation in (1.1).

In this paper, we assume  $\mathbb{F}$  to be an elliptical distribution. Following Fang et al. (1990), elliptical distributions can be formally defined here using their characteristic function:

**Definition 1.2.** A  $p$ -dimensional random vector  $\mathbf{X}$  is said to elliptically distributed if and only if there exist a vector  $\boldsymbol{\mu} \in \mathbb{R}^p$ , a positive semi-definite matrix  $\Omega \equiv \Sigma^{-1} \in \mathbb{R}^{p \times p}$  and a function  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}$  such that the characteristic function  $\mathbf{t} \mapsto \phi_{\mathbf{X}-\boldsymbol{\mu}}(\mathbf{t})$  of  $\mathbf{X} - \boldsymbol{\mu}$  corresponds to  $\mathbf{t} \mapsto \phi(\mathbf{t}^T \Sigma \mathbf{t})$ ,  $\mathbf{t} \in \mathbb{R}^p$ .

The density function of an elliptically distributed random variable takes the form:

$$h(\mathbf{x}; \boldsymbol{\mu}, \Sigma) = |\Omega|^{1/2} g((\mathbf{x} - \boldsymbol{\mu})^T \Omega (\mathbf{x} - \boldsymbol{\mu})),$$

where  $g$  is a non-negative scalar-valued density function that is continuous and strictly increasing, and is called the *density generator* of the elliptical distribution. For ease of notation, we denote such a distribution by  $\mathcal{E}(\boldsymbol{\mu}, \Sigma, g)$ .

## Organization of paper.

### Notation.

## 2 Weighted sign vectors in location problem

In this section, we focus on the general situation of estimating or testing for the location parameter  $\boldsymbol{\mu} \equiv \boldsymbol{\mu}_F$  using the weighted sign vectors in (1.1). We impose the following very general conditions on the weights  $w(\mathbf{x}, \mathbb{F}) := \kappa_P(P(\mathbf{x}, \mathbb{F}))$ :

(A1) The weights are affine invariant, i.e.

$$w(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{A}\mathbb{F} + \mathbf{b}) = w(\mathbf{x}, \mathbb{F}),$$

for any  $\mathbf{b} \in \mathbb{R}^p$  and full rank matrix  $\mathbf{A} \in \mathbb{R}^{p \times p}$ .

(A2) Given a fixed  $\mathbb{F}$ , the weights are square-integrable:

$$\int [w(\mathbf{y}, \mathbb{F})]^2 d\mathbb{G}(\mathbf{y}) < \infty; \mathbb{G} \in \mathcal{M}.$$

Because of the affine invariance, it is possible to simplify the weight as a function of the norm of  $\mathbf{z} = \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ :  $w(\mathbf{x}, \mathbb{F}) = f(r)$ , with  $r = \|\mathbf{z}\|$ .

A straightforward use of weighted signs in a location problem is to construct an outlier-robust alternative to the Hotelling's  $T^2$  test using their sample mean vector and covariance matrix. Formally, this means testing for  $H_0 : \boldsymbol{\mu} = \mathbf{0}_p$  vs.  $H_1 : \boldsymbol{\mu} \neq \mathbf{0}_p$  based on the test statistic:

$$T_{n,w} = n\bar{\mathbf{X}}_w^T(\mathbb{V}\mathbb{X}_w)^{-1}\bar{\mathbf{X}}_w,$$

(**define**  $\mathbb{X}_w$ ) with  $\bar{\mathbf{X}}_w = \sum_{i=1}^n \mathbf{X}_{w,i}/n$  and  $\mathbf{X}_{w,i} = w(\mathbf{X}_i)\mathbf{S}(\mathbf{X}_i)$  for  $i = 1, 2, \dots, n$ . However, the following holds true for this weighted sign test:

**Proposition 2.1.** *Consider  $n$  random variables  $Z = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^T$  distributed independently and identically as  $\mathcal{E}(\boldsymbol{\mu}, k\mathbf{I}_p, g)$ ;  $k \in \mathbb{R}$ , and the class of hypothesis tests defined above. Then, given any  $\alpha \in (0, 1)$ , local power at  $\boldsymbol{\mu} \neq \mathbf{0}_p$  for the level- $\alpha$  test based on  $T_{n,w}$  is maximum when  $w(\mathbf{Z}_1) = c$ , a constant independent of  $\mathbf{Z}_1$ .*

This essentially means that power-wise the (unweighted) spatial sign test (Oja, 2010) is optimal in the given class of hypothesis tests when the data comes from a spherically symmetric distribution. Our simulations show that this empirically holds for non-spherical but elliptic distributions as well.

## 2.1 The weighted spatial median

To explore usage of weighted spatial signs in the location problem that improve upon the state-of-the-art, we concentrate on the following optimization problem:

$$\boldsymbol{\mu}_w = \arg \min_{\boldsymbol{\mu}_0 \in \mathbb{R}^p} \mathbb{E}(w(\mathbf{X})|\mathbf{X} - \boldsymbol{\mu}_0|). \quad (2.1)$$

This can be seen as a generalization of the Fermat-Weber location problem (which has the spatial median (Brown, 1983; Chaudhuri, 1996) as the solution) using data-dependent weights. Using affine equivariant weights in (2.1) ensures that the weights are independent of  $\boldsymbol{\mu}_0$ , allowing the optimization problem to have a unique solution. We call this solution the *weighted spatial median* of  $\mathbb{F}$ , and denote it by  $\boldsymbol{\mu}_w$ . In a sample setup it is estimated by iteratively solving the equation  $\sum_{i=1}^n w(\mathbf{X}_i)\mathbf{S}(\mathbf{X}_i - \hat{\boldsymbol{\mu}}_w)/n = \mathbf{0}_p$ .

The sample weighted spatial median  $\hat{\boldsymbol{\mu}}_w$  is a  $\sqrt{n}$ -consistent estimator of  $\boldsymbol{\mu}_w$ , with the following result gives its asymptotic distribution:

**Theorem 2.2.** *Let  $\mathbf{A}_w, \mathbf{B}_w$  be two matrices, dependent on the weight function  $w$  such that*

$$\mathbf{A}_w = \mathbb{E} \left[ \frac{w(\boldsymbol{\epsilon})}{\|\boldsymbol{\epsilon}\|} \left( 1 - \frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T}{\|\boldsymbol{\epsilon}\|^2} \right) \right]; \quad \mathbf{B}_w = \mathbb{E} \left[ \frac{(w(\boldsymbol{\epsilon}))^2 \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T}{\|\boldsymbol{\epsilon}\|^2} \right]$$

where  $\boldsymbol{\epsilon} \sim \mathcal{E}(\mathbf{0}, \Sigma, g)$ . Then

$$\sqrt{n}(\hat{\boldsymbol{\mu}}_w - \boldsymbol{\mu}_w) \rightsquigarrow \mathcal{N}_p(\mathbf{0}, \mathbf{A}_w^{-1}\mathbf{B}_w\mathbf{A}_w^{-1}) \quad (2.2)$$

Setting  $w(\boldsymbol{\epsilon}) = 1$  above yields the asymptotic covariance matrix for the spatial median. Following this, the asymptotic relative efficiency (ARE) of  $\boldsymbol{\mu}_w$  corresponding to some non-uniform weight function with respect to the spatial median, say  $\boldsymbol{\mu}_s$  will be:

$$ARE(\boldsymbol{\mu}_w, \boldsymbol{\mu}_s) = \left[ \frac{\det(\mathbf{A}_w^{-1}\mathbf{B}_w\mathbf{A}_w^{-1})}{\det(\mathbf{A}_s^{-1}\mathbf{B}_s\mathbf{A}_s^{-1})} \right]^{1/p}$$

	$t_3$	$t_5$	$t_{10}$	$t_{20}$	Normal
$p = 5$	1.28	1.20	1.16	1.14	1.13
$p = 10$	1.15	1.10	1.07	1.07	1.06
$p = 20$	1.09	1.05	1.04	1.03	1.03
$p = 50$	1.05	1.02	1.01	1.01	1.01

Table 1: Table of  $ARE(\boldsymbol{\mu}_w; \boldsymbol{\mu}_s)$  for different spherical distributions

with  $\mathbf{A} = \mathbb{E}[1/\|\boldsymbol{\epsilon}\|(I_p - \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T/\|\boldsymbol{\epsilon}\|^2)]$  and  $\mathbf{B} = \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T/\|\boldsymbol{\epsilon}\|^2]$ . This is further simplified under spherical symmetry:

**Corollary 2.3.** *For the spherical distribution  $\mathcal{E}(\boldsymbol{\mu}, k\mathbf{I}_p, g)$ ;  $k \in \mathbb{R}$ ,  $\boldsymbol{\mu} \in \mathbb{R}^p$ , we have*

$$ARE(\boldsymbol{\mu}_w, \boldsymbol{\mu}_s) = \frac{\left[\mathbb{E}\left(\frac{f(r)}{r}\right)\right]^2}{\mathbb{E}f^2(r) \left[\mathbb{E}\left(\frac{1}{r}\right)\right]^2}$$

Table 1 summarizes the AREs for several families of elliptic distributions, numerically calculated using 10,000 random samples, and taking  $f(r) = 1/(1+r)$ . It is evident from the table that the weighted spatial median outperforms its unweighted counterpart for all data dimensions and distribution families. While the performance is much better for small values of  $p$ , weighting the signs seems to have less and less effect as  $p$  grows larger. Assuming a first-order autoregressive (AR(1)) covariance structure, i.e.  $\sigma_{ij} = \rho^{|i-j|}$  with  $\rho \in (0, 1)$ , results in largely similar ARE values as those obtained in table 1 that assume  $\Sigma = \mathbf{I}_p$ .

## 2.2 A high-dimensional test of location

It is possible to take an alternative approach to the location testing problem by using the covariance-type U-statistic  $C_{n,w} = \sum_{i=1}^n \sum_{j=1}^{i-1} \mathbf{X}_{w,i}^T \mathbf{X}_{w,j}$ . This class of test statistics are especially attractive since they are readily generalized to cover high-dimensional situations, i.e. when  $p > n$ . The Chen and Qin (CQ) high-dimensional test of location for multivariate normal  $\mathbf{X}_i$  (Chen and Qin, 2010) is a special case of this test that uses the statistic  $C_n = \sum_{i=1}^n \sum_{j=1}^{i-1} \mathbf{X}_i^T \mathbf{X}_j$ . The testing procedure of Wang et al. (2015) (from here on referred to as WPL test) shows that one can improve upon the power of the CQ test for non-gaussian elliptical distributions by using spatial signs  $\mathbf{S}(\mathbf{X}_i)$  in place of the actual variables.

Under mild regularity conditions in the lines of Wang et al. (2015), the following results hold for our generalized test statistic  $C_{n,w}$  under  $H_0$  as  $n, p \rightarrow \infty$ :

$$\frac{C_{n,w}}{\sqrt{\frac{n(n-1)}{2} \text{Tr}(\mathbf{B}_w^2)}} \rightsquigarrow N(0, 1), \quad (2.3)$$

and under contiguous alternatives  $H_1 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ ,

$$\frac{C_{n,w} - \frac{n(n-1)}{2} \boldsymbol{\mu}_0^T \mathbf{A}_w^2 \boldsymbol{\mu}_0 (1 + o(1))}{\sqrt{\frac{n(n-1)}{2} \text{Tr}(\mathbf{B}_w^2)}} \rightsquigarrow N(0, 1). \quad (2.4)$$

$\boldsymbol{\mu} = \text{rep}(.15, p)$				
$p$	$n$	CQ	WPL	$C_{n,w}$
500	20	0.051	0.376	0.418
500	50	0.060	0.832	0.866
1000	20	0.044	0.541	0.584
1000	50	0.039	0.973	0.987
$\boldsymbol{\mu} = \text{rep}(0, p)$				
$p$	$n$	CQ	WPL	$C_{n,w}$
500	20	0.049	0.061	0.063
500	50	0.039	0.061	0.064
1000	20	0.042	0.060	0.063
1000	50	0.043	0.050	0.050

Table 2: Table of empirical powers of level-0.05 tests for the Chen and Qin (CQ), WPL and  $C_{n,w}$  statistics

We provide the details behind deriving these two results in the supplementary material, which extend the results of Wang et al. (2015) in a weighted sign setup using modified regularity conditions.

Following this, the ARE of this test statistic with respect to its unweighted version, i.e. the WPL statistic, is expressed as:

$$ARE(C_{n,w}, \text{WPL}; \boldsymbol{\mu}_0) = \frac{\boldsymbol{\mu}_0^T \mathbf{A}_w^2 \boldsymbol{\mu}_0}{\boldsymbol{\mu}_0^T \mathbf{A}^2 \boldsymbol{\mu}_0} \sqrt{\frac{\text{Tr}(\mathbf{B}^2)}{\text{Tr}(\mathbf{B}_w^2)}} (1 + o(1)),$$

when  $\Sigma = k\mathbf{I}_p$ , this again simplifies to  $\mathbb{E}^2(f(r)/r)/[\mathbb{E}f^2(r).\mathbb{E}^2(1/r)]$ . Thus the ARE values are exactly same as those in Table 1, indicating that for large data dimensions the WPL test and that based on  $C_{n,w}$  are almost equivalent.

However, in a practical high-dimensional setup sample sizes are often small. Thus, comparing the the two tests with respect to their *finite sample* efficiencies instead gives a better idea of their practical utility. We do so in table 2, which lists empirical powers calculated from 1000 replications of each setup under an AR(1) covariance structure (with  $\rho = 0.8$ ). While under  $H_0 : \boldsymbol{\mu} = \mathbf{0}_p$  all tests have similar performance,  $C_{n,w}$  beats the other two under deviations from the null.

### 3 Inverse depth and estimation of scatter

We elaborate on the problem estimating the covariance matrix  $\Sigma$  and its eigenvectors and eigenvalues using signed peripherality functions in this section. In addition to conditions (P) and (A1), we assume the following conditions on these functions in this section:

(B1) The following holds:

$$P(\mathbf{x}; \mathbb{F}) \leq P(\boldsymbol{\mu}_F + t(\mathbf{x} - \boldsymbol{\mu}_F), \mathbb{F}),$$

for any  $\mathbf{x} \in \mathbb{R}^p$  and  $t \in (0, 1)$ .

(B2) There exists a positive constant  $M(P, \mathbb{F})$  such that  $P(\mathbf{x}; \mathbb{F}) \rightarrow M(P, \mathbb{F})$  as  $\|\mathbf{x}\| \rightarrow \infty$ .

Condition (A2) in Section 2 is now replaced by the stronger condition (B2). Note that all the current conditions are essentially the opposite of those imposed on a traditional depth function (Zuo and Serfling, 2000). Given a depth function, the peripherality functions in this section can be defined as a bounded monotonically decreasing function of it. Consequently, we use the term *inverse depth function* in place of peripherality functions in this section, denoting them by  $D^-(\mathbf{x}, \mathbb{F})$ , or equivalently  $D_{\mathbf{X}}^-(\mathbf{x})$ .

Given an inverse depth function  $D_{\mathbf{X}}^-(\mathbf{x})$  we transform the original random variable as:  $\tilde{\mathbf{X}} = D_{\mathbf{X}}^-(\mathbf{x})\mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$ . We attempt to use these vector as the proxy of original data for improved estimation of the components of the population covariance matrix  $\Sigma$  in this section. To this end, in Section 3.1 we give some results characterizing the population covariance matrix of these vectors, i.e.  $\mathbb{V}\tilde{\mathbf{X}}$ , and its affine equivariant counterpart.

### 3.1 Depth Covariance Matrix

Consider the spectral decomposition of  $\Sigma$ :  $\Sigma = \Gamma\Lambda\Gamma^T$ ,  $\Gamma$  being orthogonal and  $\Lambda$  diagonal with positive diagonal elements  $\lambda_1, \dots, \lambda_p$ . 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

$$\begin{aligned}\tilde{\mathbf{X}} &= D_{\mathbf{X}}^-(\mathbf{X})\mathbf{S}(\mathbf{x} - \boldsymbol{\mu}) \\ &= D_{\Gamma\Lambda^{1/2}\mathbf{Z} + \boldsymbol{\mu}}^-(\Gamma\Lambda^{1/2}\mathbf{Z} + \boldsymbol{\mu})\mathbf{S}(\Gamma\Lambda^{1/2}\mathbf{Z}) \\ &= \Gamma D_{\mathbf{Z}}^-(\mathbf{Z})\mathbf{S}(\Lambda^{1/2}\mathbf{Z}) \\ &= \Gamma\Lambda^{1/2}D_{\mathbf{Z}}^-(\mathbf{Z})\mathbf{S}(\mathbf{Z})\frac{\|\mathbf{Z}\|}{\|\Lambda^{1/2}\mathbf{Z}\|}.\end{aligned}\tag{3.1}$$

$D_{\mathbf{Z}}^-(\cdot)$  is an even function in its argument because of affine invariance, as is  $\|\mathbf{z}\|/\|\Lambda^{1/2}\mathbf{z}\|$ . Since the sign function  $\mathbf{S}(\cdot)$  is odd in its argument, it follows that  $\mathbb{E}\tilde{\mathbf{X}} = \mathbf{0}$ . Consequently, we obtain an expression for  $\mathbb{V}\tilde{\mathbf{X}}$ , which we call the *Depth Covariance Matrix* (DCM):

**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*

$$\mathbb{V}\tilde{\mathbf{X}} = \Gamma\Lambda_D\Gamma^T, \quad \text{with} \quad \Lambda_D = \mathbb{E}_{\mathbf{Z}} \left[ (D_{\mathbf{Z}}^-(\mathbf{z}))^2 \frac{\Lambda^{1/2}\mathbf{z}\mathbf{z}^T\Lambda^{1/2}}{\mathbf{z}^T\Lambda\mathbf{z}} \right], \tag{3.2}$$

where  $\mathbf{Z} = (Z_1, \dots, Z_p)^T \sim N(\mathbf{0}, \mathbf{I}_p)$ , so that  $\Lambda_D$  a diagonal matrix with diagonal entries

$$\lambda_{D,i} = \mathbb{E}_{\mathbf{Z}} \left[ \frac{(D_{\mathbf{Z}}^-(\mathbf{z}))^2 \lambda_i z_i^2}{\sum_{j=1}^p \lambda_j z_j^2} \right].$$



The matrix of eigenvectors  $\Gamma$  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 we are going to discuss shortly. However, as one can see in the above expression, the diagonal entries of  $\Lambda_D$  do not change if a scale change is done on all entries of  $\Lambda$ , meaning the  $\Lambda_D$  matrices corresponding to  $\mathbb{F}$  and  $c\mathbb{F}$  for some  $c \neq 0$  will be same. This is the reason the DCM is not equivariant under affine transformations.

We need to follow the general framework of M-estimation with data-dependent weights (Huber, 1981) to construct an affine equivariant counterpart of the DCM. Specifically, we implicitly define the Affine-equivariant Depth Covariance Matrix (ADCM) as

$$\Sigma_{Dw} = \frac{1}{\mathbb{V}\tilde{Z}_1} \mathbb{E}_{\mathbf{X}} \left[ \frac{(D_{\tilde{\mathbf{X}}}^-(\mathbf{x}))^2 (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T}{(\mathbf{x} - \boldsymbol{\mu})^T \Sigma_{Dw}^{-1} (\mathbf{x} - \boldsymbol{\mu})} \right] \quad (3.3)$$

Its affine equivariance follows from the fact that the weights  $(D_{\tilde{\mathbf{X}}}^-(\mathbf{x}))^2$  depend only on the standardized quantities  $\mathbf{z} \sim \mathcal{E}(\mathbf{0}, k\mathbf{I}_p, g)$ . We solve (3.3) iteratively by obtaining a sequence of positive definite matrices  $\Sigma_{Dw}^{(k)}$  until convergence:

$$\Sigma_{Dw}^{(k+1)} = \frac{1}{Var(\tilde{Z}_1)} E \left[ \frac{(D_{\tilde{\mathbf{X}}}^-(\mathbf{x}))^2 (\Sigma_{Dw}^{(k)})^{1/2} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T (\Sigma_{Dw}^{(k)})^{1/2}}{(\mathbf{x} - \boldsymbol{\mu})^T (\Sigma_{Dw}^{(k)})^{-1} (\mathbf{x} - \boldsymbol{\mu})} \right]$$

To ensure existence and uniqueness of this estimator, let us consider the class of scatter estimators  $\Sigma_M$  that are obtained as solutions of the following equation:

$$E_{\mathbf{Z}_M} \left[ u(\|\mathbf{z}_M\|) \frac{\mathbf{z}_M \mathbf{z}_M^T}{\|\mathbf{z}_M\|^2} - v(\|\mathbf{z}_M\|) I_p \right] = 0 \quad (3.4)$$

with  $\mathbf{z}_M = \Sigma_M^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ . Under the following assumptions on the scalar valued functions  $u$  and  $v$ , the above equation produces a unique solution (Huber, 1981):

- (M1) The function  $u(r)/r^2$  is monotone decreasing, and  $u(r) > 0$  for  $r > 0$ ;
- (M2) The function  $v(r)$  is monotone decreasing, and  $v(r) > 0$  for  $r > 0$ ;
- (M3) Both  $u(r)$  and  $v(r)$  are bounded and continuous;
- (M4)  $u(0)/v(0) < p$ ;
- (M5) For any hyperplane in the sample space  $\mathcal{X}$ , (i)  $P(H) = E_{\mathbf{X}} 1_{\mathbf{x} \in H} < 1 - pv(\infty)/u(\infty)$  and (ii)  $P(H) \leq 1/p$ .

In our case we take  $v(r) = Var(\tilde{Z}_1)$ , i.e. a constant, thus (M2) and (M3) are trivially satisfied. As for  $u$ , we notice that most well-known depth functions can be expressed as simple functions of the norm of the standardized random variable. For example,  $PD_{\mathbf{Z}}(\mathbf{z}) = (1 - G(\|\mathbf{z}\|))$ ;  $MhD_{\mathbf{Z}}(\mathbf{z}) = (1 + \|\mathbf{z}\|^2)^{-1}$ ;  $HSD_{\mathbf{Z}}(\mathbf{z}) = (1 + \|\mathbf{z}\|)^{-1}$  etc., so that we can take as  $u$  square of the corresponding peripherality functions:

$$u_{PD}(r) = G^2(r); \quad u_{MhD}(r) = \frac{r^4}{(1 + r^2)^2}; \quad u_{HSD}(r) = \frac{r^2}{(1 + r/G^{-1}(0.75))^2}$$

It is easy to verify that the above choices of  $u$  satisfy (M1) and (M3). To check (M4) and (M5), first notice that  $\mathbf{Z}$  has a spherically symmetric distribution, so that its norm and sign are independent. Since  $D_{\mathbf{Z}}(\mathbf{z})$  depends only on  $\|\mathbf{z}\|$ , we have

$$\text{Var}(\tilde{Z}_1) = \text{Var}\left(D_{\mathbf{Z}}^-(\mathbf{Z}) \frac{Z_1}{\|\mathbf{Z}\|}\right) = \text{Var}(D_{\mathbf{Z}}^-(\mathbf{Z}))\text{Var}(S_1(\mathbf{Z})) = \frac{1}{p}\text{Var}(D_{\mathbf{Z}}^-(\mathbf{Z}))$$

as  $\text{Cov}(\mathbf{S}(\mathbf{Z})) = \text{Cov}((S_1(\mathbf{Z}), S_2(\mathbf{Z}), \dots, S_p(\mathbf{Z}))^T) = I_p/p$ . Now for MhD and HSD  $u(\infty) = 1, u(0) = 0$ , so (M4) and (M5) are immediate. To achieve this for PD, we only need to replace  $u_{PD}(r)$  with  $u_{PD}^*(r) = G^2(r) - 1/4$ .

### 3.2 Calculating the sample DCM and ADCM

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 vectorized form of  $\sqrt{n}$ -times the sample DCM:  $\sum_{i=1}^n (D_{\mathbf{X}}^-(\mathbf{x}_i))^2 SS(\mathbf{x}_i; \boldsymbol{\mu}) / \sqrt{n}$  has an asymptotic multivariate normal distribution with mean  $\sqrt{n} \cdot \text{vec}(E[(D_{\mathbf{X}}^-(\mathbf{X}))^2 SS(\mathbf{x}; \boldsymbol{\mu})])$  and a certain covariance matrix by straightforward application of the central limit theorem (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  $D_{\mathbf{X}}^n(\mathbf{x}) = D(\mathbf{x}, F_n)$ . Here we make the following assumption regarding how it approximates  $D_{\mathbf{X}}(\mathbf{x})$ :

**(D5) Uniform convergence:**  $\sup_{\mathbf{x} \in \mathbb{R}^p} |D_{\mathbf{X}}^n(\mathbf{x}) - D_{\mathbf{X}}(\mathbf{x})| \rightarrow 0$  as  $n \rightarrow \infty$ .

The assumption that empirical depths converge uniformly at all points  $\mathbf{x}$  to their population versions holds under very mild conditions for several well known depth functions: for example projection depth (Zuo, 2003) and simplicial depth (Dümbgen, 1992). 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 CLT we need the following result:

**Lemma 3.2.** *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 the assumption (D5) we have*

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

Following this, we are now in a position to state the result for consistency of the sample DCM:

**Theorem 3.3.** Consider  $n$  iid samples from the distribution in Lemma 3.2. Then, given a depth function  $D_{\mathbf{X}}(\cdot)$  and an estimate of center  $\hat{\boldsymbol{\mu}}_n$  so that  $\sqrt{n}(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}) = O_P(1)$ ,

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

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

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.

In contrast to the DCM, the issue of estimating  $\boldsymbol{\mu}$  to plug into the ADCM is easily handled by simultaneously solving for the location and scatter functionals  $(\boldsymbol{\mu}_{Dw}, \Sigma_{Dw})$ :

$$E \left[ \frac{\Sigma_{Dw}^{-1/2}(\mathbf{x} - \boldsymbol{\mu}_{Dw})}{\|\Sigma_{Dw}^{-1/2}(\mathbf{x} - \boldsymbol{\mu}_{Dw})\|} \right] = \mathbf{0}_p \quad (3.5)$$

$$E \left[ \frac{(D_{\mathbf{X}}^-(\mathbf{x}))^2 \Sigma_{Dw}^{-1/2}(\mathbf{x} - \boldsymbol{\mu}_{Dw})(\mathbf{x} - \boldsymbol{\mu}_{Dw})^T \Sigma_{Dw}^{-1/2}}{(\mathbf{x} - \boldsymbol{\mu}_{Dw})^T \Sigma_{Dw}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{Dw})} \right] = \text{Var}(\tilde{Z}_1) I_p \quad (3.6)$$

In the framework of (3.3), for any fixed  $\Sigma_M$  there exists a unique and fixed solution of the location problem  $E_{\mathbf{Z}_M}(w(\|\mathbf{z}_M\|)\mathbf{z}_M) = \mathbf{0}_p$  under the following condition:

(M6) The function  $w(r)r$  is monotone increasing for  $r > 0$ .

This condition is easy to verify for our choice of the weights:  $w(\|\mathbf{z}_M\|) = D_{\mathbf{Z}_M}^-(\mathbf{z}_M)/\|\mathbf{z}_M\|$ . Uniqueness of simultaneous fixed point solutions of 3.5 and 3.6 is guaranteed when  $\mathbf{X}$  has a symmetric distribution (Huber, 1981).

In practice it is difficult to calculate the scale multiple  $\text{Var}(\tilde{Z}_1)$  analytically for known depth functions and an arbitrary  $F$ . Here we instead obtain its standardized version  $\Sigma_{Dw}^* = \Sigma_{Dw}/\text{Var}(\tilde{Z}_1)$  (so that the determinant equals 1), alongwith  $\boldsymbol{\mu}_{Dw}$  using the following iterative algorithm:

1. Start from some initial estimates  $(\boldsymbol{\mu}_{Dw}^{(0)}, \Sigma_{Dw,(0)})$ . Set  $t = 0$ ;
2. Calculate the standardized observations  $\mathbf{z}_i^{(t)} = \Sigma_{Dw,(t)}^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu}_{Dw}^{(t)})$ ;
3. Update the location estimate:

$$\boldsymbol{\mu}_{Dw}^{(t+1)} = \frac{\sum_{i=1}^n \tilde{\mathbf{x}}_i / \|\mathbf{z}_i^{(t)}\|}{\sum_{i=1}^n 1 / \|\mathbf{z}_i^{(t)}\|}$$

4. Update the scatter estimate:

$$\Sigma_{Dw}^{*(t+1)} = \frac{1}{n} \sum_{i=1}^n \frac{(D_{\mathbf{X}}^-(\mathbf{x}_i))^2 (\mathbf{x}_i - \boldsymbol{\mu}_{Dw}^{(t+1)})(\mathbf{x}_i - \boldsymbol{\mu}_{Dw}^{(t+1)})^T}{\|\mathbf{z}_i^{(t)}\|^2}; \quad \Sigma_{Dw}^{*(t+1)} \leftarrow \frac{\Sigma_{Dw}^{*(t+1)}}{\det(\Sigma_{Dw}^{*(t+1)})^{1/p}}$$

5. Continue until convergence.

### 3.3 Robust PCA using eigenvectors

Since we are mainly interested in using the DCM for robust 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 its eigenvectors. In case any of the eigenvalues have multiplicity larger than 1, limiting distributions of the corresponding eigenprojection matrices can be obtained analogous to those of the sign covariance matrix (Magyar and Tyler, 2014).

#### 3.3.1 Influence functions

We start with deriving the influence functions for eigenvectors of the DCM and ADCM. This will help in demonstrating the robustness of their estimates, as well as deriving the asymptotic distributions of their sample counterparts. 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 (Hampel et al., 1986). 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,

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

Following Croux and Haesbroeck (2000), we now get the influence function of the  $i^{\text{th}}$  eigenvector of  $Cov(\tilde{\mathbf{X}})$ , say  $\boldsymbol{\gamma}_D = (\gamma_{D,1}, \dots, \gamma_{D,p})$ ;  $i = 1, \dots, p$ :

$$\begin{aligned} IF(\mathbf{x}_0; \boldsymbol{\gamma}_{D,i}, F) &= \sum_{k=1; k \neq i}^p \frac{1}{\lambda_{D,S,i} - \lambda_{D,S,k}} \left\{ \boldsymbol{\gamma}_k^T IF(\mathbf{x}_0; Cov(\tilde{\mathbf{X}}), \boldsymbol{\gamma}_i) \right\} \boldsymbol{\gamma}_k \\ &= \sum_{k=1; k \neq i}^p \frac{1}{\lambda_{D,S,i} - \lambda_{D,S,k}} \left\{ \boldsymbol{\gamma}_k^T (D_{\tilde{\mathbf{X}}}^-(\mathbf{x}_0))^2 SS(\mathbf{x}_0; \boldsymbol{\mu}) \boldsymbol{\gamma}_i - \lambda_{D,S,i} \boldsymbol{\gamma}_k^T \boldsymbol{\gamma}_i \right\} \boldsymbol{\gamma}_k \\ &= \sum_{k=1; k \neq i}^p \frac{\sqrt{\lambda_i \lambda_k} z_{0i} z_{0k}}{\lambda_{D,S,i} - \lambda_{D,S,k}} \cdot \frac{(D_{\tilde{\mathbf{Z}}}^-(\mathbf{z}_0))^2}{\mathbf{z}_0^T \Lambda \mathbf{z}_0} \boldsymbol{\gamma}_k \end{aligned} \quad (3.7)$$

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.

For the ADCM, we first notice that the influence function of any affine equivariant estimate of scatter can be expressed as

$$IF(\mathbf{x}_0, C, F) = \alpha_C(\|\mathbf{z}_0\|) \frac{\mathbf{z}_0 \mathbf{z}_0^T}{\mathbf{z}_0^T \mathbf{z}_0} - \beta_C(\|\mathbf{z}_0\|) C$$

for scalar valued functions  $\alpha_C, \beta_C$  (Hampel et al., 1986). Following this, the influence function of an eigenvector  $\gamma_{C,i}$  of  $C$  is derived:

$$IF(\mathbf{x}_0, \gamma_{C,i}, F) = \alpha_C(\|\mathbf{z}_0\|) \sum_{k=1, k \neq i}^p \frac{\sqrt{\lambda_i \lambda_k}}{\lambda_i - \lambda_k} \cdot \frac{z_{0i} z_{0k}}{\mathbf{z}_0^T \mathbf{z}_0} \gamma_k$$

When  $C = \Sigma_M$ , i.e. the solution to (3.3), then Huber (1981) shows that

$$\alpha_C(\|\mathbf{z}_0\|) = \frac{p(p+2)u(\|\mathbf{z}_0\|)}{E_{F_0} [pu(\|\mathbf{y}\|) + u'(\|\mathbf{y}\|)\|\mathbf{y}\|]}$$

Setting  $u(\|\mathbf{z}_0\|) = (D_{\mathbf{Z}}^-(\mathbf{z}_0))^2$  ensures that the influence function of eigenvectors of the ADCM is bounded as well as increasing in magnitude with  $\|\mathbf{z}_0\|$ .

In Figure 2 we consider first eigenvectors of our scatter estimates, as well as two well-known robust estimates of scatter: the Sign Covariance Matrix (SCM) and Tyler's shape matrix, for the  $\mathcal{N}_2((0,0)^T, \text{diag}(2,1))$  and plot norms of these influence functions for different values of  $\mathbf{x}_0$ . Influence function for the  $i^{\text{th}}$  eigenvectors of these two matrices (say  $\gamma_{S,i}$  and  $\gamma_{T,i}$ , respectively) are as follows:

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

Their corresponding plots demonstrate the 'inlier effect', i.e. points close to symmetry center and the center itself having high influence, which results in loss of efficiency. The influence function for the sample covariance matrix is obtained by replacing  $(p+2)$  by  $\|\mathbf{z}_0\|^2$  in the expression of  $IF(\mathbf{x}_0; \gamma_{T,i}, F)$  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.

### 3.3.2 Asymptotic and finite-sample efficiencies

Suppose  $\hat{C}$  is a  $\sqrt{n}$ -consistent estimator of a scatter functional  $C$ . Then the asymptotic variance of its eigenvectors are Anderson (2003)

$$AVar(\sqrt{n}\hat{\gamma}_{c,i}) = \sum_{k=1; k \neq i}^p \frac{\lambda_i \lambda_k}{(\lambda_i - \lambda_k)^2} \gamma_k \gamma_k^T \quad (3.8)$$

The asymptotic relative efficiencies of eigenvectors from the sample DCM with respect to the sample covariance matrix can now be derived using (3.8) above and (B.2) from

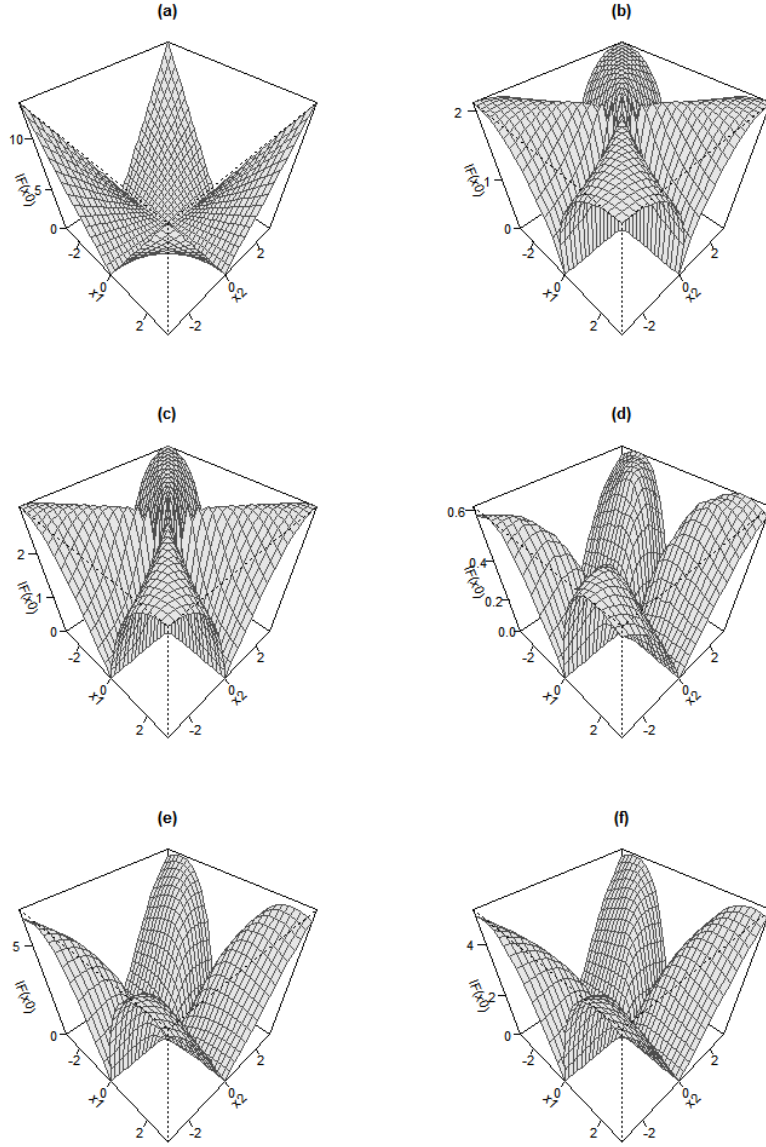


Figure 2: Plot of the norm of influence function for first eigenvector of (a) sample covariance matrix, (b) SCM, (c) Tyler's scatter matrix and DCMs for (d) Halfspace depth, (e) Mahalanobis depth, (f) Projection depth for a bivariate normal distribution with  $\boldsymbol{\mu} = \mathbf{0}$ ,  $\Sigma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

Corollary B.2:

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

Obtaining ARE of the ADCM is, in comparison to DCM, more straightforward. The asymptotic covariance matrix of an eigenvector of the affine equivariant scatter functional  $C$  is given by:

$$AVar(\sqrt{n}\hat{\gamma}_{C,j}) = ASV(C_{12}, F_0) \sum_{k=1, k \neq i}^p \frac{\lambda_i \lambda_k}{\lambda_i - \lambda_k} \cdot \gamma_i \gamma_k^T$$

where  $ASV(C_{12}, F_0)$  is the asymptotic variance of an off-diagonal element of  $C$  when the underlying distribution is  $F_0$ . Following [Croux and Haesbroeck \(2000\)](#) this equals

$$ASV(C_{12}, F_0) = E_{F_0} [\alpha_c(\|\mathbf{z}\|)^2 (S_1(\mathbf{z})S_2(\mathbf{z}))^2] = E_{F_0} \alpha_C(\|\mathbf{z}\|)^2 \cdot E_{F_0} (S_1(\mathbf{z})S_2(\mathbf{z}))^2$$

again using the fact that  $\|\mathbf{Z}\|$  and  $\mathbf{S}(\mathbf{Z})$  are independent with  $\mathbf{Z} \sim F_0$ . It now follows that

$$ARE(\hat{\gamma}_{\Sigma_M, i}, \hat{\gamma}_{Cov, i}; F) = \frac{E_{F_0} \alpha_{Cov}(\|\mathbf{z}\|)^2}{E_{F_0} \alpha_C(\|\mathbf{z}\|)^2} = \frac{E_{F_0} \|\mathbf{z}\|^4 \cdot [E_{F_0}(pu\|\mathbf{z}\| + u'(\|\mathbf{z}\|)\|\mathbf{z}\|)]^2}{E_{F_0}(u(\|\mathbf{z}\|))^2} \quad (3.9)$$

Table 3 considers 6 different elliptic distributions (namely, bivariate  $t$  with  $\text{df} = 5, 6, 10, 15, 25$  and bivariate normal) and summarizes ARE for first eigenvectors for ADCMs corresponding to projection depth (PD-ACM) and halfspace depth (HSD-ACM). Due to difficulty of analytically obtain the AREs, we calculate them using Monte-Carlo simulation of  $10^6$  samples and subsequent numerical integration. The ADCM seems to be particularly efficient in lower dimensions for distributions with heavier tails ( $t_5$  and  $t_6$ ), while for distributions with lighter tails, the AREs increase with data dimension. At higher values of  $p$  the ADCM is almost as efficient as the sample covarnace matrix when the data comes from multivariate normal distribution.

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 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 ([Miao and Ben-Israel, 1992](#)) to find out error estimates for the first eigenvector of a scatter matrix. In our case, the first eigenvector will be

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

Distribution	PD-ACM				HSD-ACM			
	$p = 2$	$p = 5$	$p = 10$	$p = 20$	$p = 2$	$p = 5$	$p = 10$	$p = 20$
$t_5$	4.73	3.99	3.46	3.26	4.18	3.63	3.36	3.15
$t_6$	2.97	3.28	2.49	2.36	2.59	2.45	2.37	2.32
$t_{10}$	1.45	1.47	1.49	1.52	1.30	1.37	1.43	1.49
$t_{15}$	1.15	1.19	1.23	1.27	1.01	1.10	1.17	1.24
$t_{25}$	0.97	1.02	1.07	1.11	0.85	0.94	1.02	1.08
MVN	0.77	0.84	0.89	0.93	0.68	0.77	0.84	0.91

Table 3: Table of AREs of the ADCM for different choices of  $p$  and data-generating distributions, and two choices of depth functions

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.  $\cos^{-1} |\hat{\gamma}_1^T \hat{\gamma}_1|$ . 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 4, 5 and 6 give FSE values for  $p = 2, 3, 4$ , respectively. In general, all the efficiencies 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 Sirkiä et al. (2007), and quite close to the affine equivariant spatial sign covariance matrix (see Table 2 in Ollila et al. (2003)). 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.

### 3.4 Robust estimation of eigenvalues, and a plug-in estimator of $\Sigma$

As we have seen in theorem 3.1, eigenvalues of the DCM are not same as the population eigenvalues, whereas the ADCM only gives back standardized eigenvalues. However, it is possible to robustly estimate the original eigenvalues by working with the individual columns of the robust score matrix. We do this using the following steps:

1. Randomly divide the sample indices  $\{1, 2, \dots, n\}$  into  $k$  disjoint groups  $\{G_1, \dots, G_k\}$  of size  $\lfloor n/k \rfloor$  each;



2. Assume the data is centered. Transform the data matrix:  $S = \hat{\Gamma}_D^T X$ ;
3. Calculate coordinate-wise variances for each group of indices  $G_j$ :

$$\hat{\lambda}_{i,j} = \frac{1}{|G_j|} \sum_{l \in G_j} (s_{li} - \bar{s}_{G_j,i})^2; \quad i = 1, \dots, p; j = 1, \dots, k$$

where  $\bar{s}_{G_j} = (\bar{s}_{G_j,1}, \dots, \bar{s}_{G_j,p})^T$  is the vector of column-wise means of  $S_{G_j}$ , the sub-matrix of  $S$  with row indices in  $G_j$ .

4. Obtain estimates of eigenvalues by taking coordinate-wise medians of these variances:

$$\hat{\lambda}_i = \text{median}(\hat{\lambda}_{i,1}, \dots, \hat{\lambda}_{i,k}); \quad i = 1, \dots, p$$

The number of subgroups used to calculate this median-of-small-variances estimator can be determined following (Minsker, 2015). After this, we construct a consistent plug-in estimator of the population covariance matrix  $\Sigma$ :

**Theorem 3.4.** *Consider the estimates  $\hat{\lambda}_i$  obtained from the above algorithm, and the matrix of eigenvectors  $\hat{\Gamma}_D$  estimated using the sample DCM. Define  $\hat{\Sigma} = \hat{\Gamma}_D \hat{\Lambda} \hat{\Gamma}_D^T$ ;  $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_p)$ . Then as  $n \rightarrow \infty$ ,*

$$\|\hat{\Sigma} - \Sigma\|_F \xrightarrow{P} 0$$

$\|\cdot\|_F$  being the Frobenius norm.

(put in 1 or 2 sentences?)

## 4 Robust PCA and supervised models

In the presence of a vector of univariate responses, say  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$ , there is substantial literature devoted to utilizing the subspace generated by the basis of  $\text{Cov}(\mathbf{X})$  in modelling  $E(Y|\mathbf{X})$ . This ranges from the simple Principal Components Regression (PCR) to Partial Least Squares (PLS) and Envelope methods (Cook et al., 2010). Here we concentrate on robust inference using Sufficient Dimension Reduction (SDR) (Adraghi and Cook, 2009), mainly because it provides a general framework for reducing dimensionality of data directly using top eigenvectors of the covariance matrix of  $X$  (albeit in a different manner than PCR) or an appropriate affine transformation of it.

SDR attempts to find out a linear transformation  $R$  on  $\mathbf{X}$  such that  $E(Y|\mathbf{X}) = E(Y|R(\mathbf{X}))$ . Assuming that  $R(\mathbf{X})$  takes values in  $\mathbb{R}^d$ ,  $d \leq \min(n, p)$ , this can be achieved through an inverse regression model:

$$\mathbf{X}_y = \bar{\boldsymbol{\mu}} + \Gamma \mathbf{v}_y + \boldsymbol{\epsilon} \quad (4.1)$$

where  $\mathbf{X}_y = \mathbf{X}|Y = y$ ,  $\bar{\boldsymbol{\mu}} = E\mathbf{X}$ ,  $\Gamma$  is a  $p \times d$  semi-orthogonal basis for  $\mathcal{S}_\Gamma$ , the spanning subspace of  $\{E\mathbf{X}_y - \bar{\boldsymbol{\mu}} | y \in \mathcal{S}_Y\}$  ( $\mathcal{S}_Y$  is sample space of  $Y$ ) and  $\mathbf{v}_y = (\Gamma^T \Gamma)^{-1} \Gamma^T (E\mathbf{X}_y - \bar{\boldsymbol{\mu}}) \in \mathbb{R}^d$ . The random error term  $\boldsymbol{\epsilon}$  follows a multivariate normal distribution with mean  $\mathbf{0}_p$

and covariance matrix  $\Delta$ . This formulation is straightforward to implement when  $Y$  is categorical, while for continuous responses, the vector  $\mathbf{y}$  is divided into a number of slices.

Under this model the minimal sufficient transformation is  $R(\mathbf{X}) = \Gamma^T \Delta^{-1} \mathbf{X}$ . The simplest case of this model is when  $\Delta = \sigma^2 I_p$ , for which the maximum likelihood estimator of  $R(\mathbf{X})$  turns out to be the first  $d$  PCs of  $Cov(\mathbf{X})$ . Taking  $\hat{E}\mathbf{X}_y = \bar{\mathbf{X}}_y$  and  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$ , one can now estimate  $\sigma^2$  as:  $\hat{\sigma}^2 = \sum_{i=1}^p s_{ii}/p$ , where  $s_{ii}$  is the  $i^{\text{th}}$  diagonal element of  $\hat{Cov}_Y(\mathbf{X}_Y - \bar{\mathbf{X}} - \hat{\Gamma}\hat{\mathbf{v}}_Y)$ . Following this, predictions for a new observation  $\mathbf{x}$  is obtained as a weighted sum of the responses:

$$\hat{E}(Y|\mathbf{X} = \mathbf{x}) = \frac{\sum_{i=1}^n w_i Y_i}{\sum_{i=1}^n w_i}; \quad w_i = \exp \left[ -\frac{1}{\hat{\sigma}^2} \|\hat{\Gamma}^T(\mathbf{x} - \mathbf{X}_i)\|^2 \right]$$

We formulate a robust version of the above procedure by estimating the quantities  $\Gamma, \bar{\boldsymbol{\mu}}, \boldsymbol{\mu}_y, \sigma^2$  by robust methods. Specifically, we take:

- $\tilde{\Gamma}$  = first  $d$  eigenvectors of the sample DCM;
- $\tilde{\boldsymbol{\mu}}$  = spatial median of the rows of  $X$ ;
- $\tilde{\boldsymbol{\mu}}_y$  = spatial median of the rows of  $(X|Y = y)$ , for all  $y \in S_Y$ ;
- $\tilde{\sigma}^2 = \sum_{i=1}^p [\widehat{\text{MAD}}_Y(X_{Y,i} - \tilde{\mu}_i - \tilde{\gamma}_i^T \tilde{\mathbf{v}}_Y)]^2/p$ , with  $\tilde{\Gamma} = (\tilde{\gamma}_1, \dots, \tilde{\gamma}_p)^T$ .

The following simulation study using the same setup as in (Adraghi and Cook, 2009) compares the performance of our robust SDR with the original method with or without the presence of bad leverage points in the covariate matrix  $X$ . For a fixed dimension  $p$ , we take  $n = 200, d = 1$ , generate the responses  $Y$  as independent standard normal, and the predictors as  $\mathbf{X}_Y = \boldsymbol{\gamma}^* v_Y^* + \boldsymbol{\epsilon}$ , with  $\boldsymbol{\gamma}_{p \times 1}^* = (1, \dots, 1)^T$ ,  $v_Y = Y + Y^2 + Y^3$  and  $Var(\boldsymbol{\epsilon}) = 25I_p$ . We measure performance of both SDR models by their mean squared prediction error on another set of 200 observations  $(Y^*, \mathbf{X}^*)$  generated similarly, and taking the average of these errors on 100 such training-test pair of datasets. Finally we repeat the whole setup for different choices of  $p = 5, 10, 25, 50, 75, 100, 125, 150$ .

Panel (a) of figure 3 compares prediction errors using robust and maximum likelihood SDR estimates when  $X$  contains no outliers, and the two methods are virtually indistinguishable. We now introduce outliers in each of the 100 datasets by adding 100 to first  $p/5$  coordinates of the first 10 observations in  $X$ , and repeat the analysis. Panel (b) of the figure shows that although our robust method performs slightly worse than the case when there were no outliers, it remains more accurate in predicting our of sample observations for all values of  $p$ .

## 5 Robust inference with functional data

### (Some technical notations)

We use the approach of Boente and Salibián-Barrera (2015) for performing robust PCA on functional data. Given data on  $n$  functions, say  $f_1, f_2, \dots, f_n \in L^2[0, 1]$ , each observed at a set of common design points  $\{t_1, \dots, t_m\}$ , we model each function as a linear

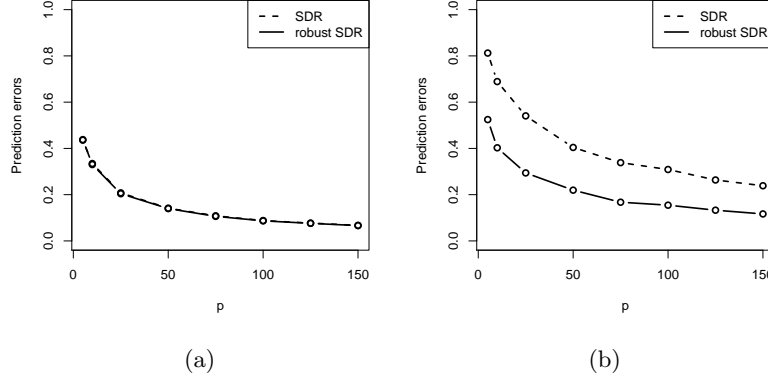


Figure 3: Average prediction errors for two methods of SDR (a) in absence and (b) in presence of outliers

combination of  $p$  mutually orthogonal B-spline basis functions  $\delta_1, \dots, \delta_p$ . Following this, we map data for each of the functions onto the coordinate system formed by the spline basis:

$$\tilde{x}_{ij} = \sum_{l=2}^m f_i(t_l) \delta_j(t_l) (t_l - t_{l-1}); \quad 1 \leq i \leq n, 1 \leq j \leq p \quad (5.1)$$

We now do depth-based PCA on the transformed  $n \times p$  data matrix  $\tilde{X}$ , and obtain the rank- $q$  approximation ( $q \leq p$ ) of the  $i^{\text{th}}$  observation using the robust  $p \times q$  loading matrix  $\tilde{P}$  and robust  $q \times 1$  score vector  $\tilde{s}_i$ :

$$\hat{\mathbf{x}}_i = \tilde{\boldsymbol{\mu}} + \tilde{P} \tilde{s}_i$$

with  $\tilde{\boldsymbol{\mu}}$  being the spatial median of  $\tilde{X}$ . Then we transform this approximation back to the original coordinates:  $\hat{f}_i(t_l) = \sum_{j=1}^p \hat{x}_{ij} \delta_j(t_l)$ .

Detection of anomalous observations is of importance in real-life problems involving functional data analysis. We now demonstrate the utility of our robust method for detecting functional outliers through two data examples.

#### (SD and OD definition, cutoffs... from previous manuscript)

We first look into the El-Niño data, which is part of a larger dataset on potential factors behind El-Niño oscillations in the tropical pacific available in <http://www.cpc.ncep.noaa.gov/data/indices/>. This records monthly average Sea Surface Temperatures from June 1970 to May 2004, and the yearly oscillations follow more or less the same pattern (see panel a of figure 4). Using a cubic spline basis with knots at alternate months starting in June gives a close approximation of the yearly time series data (panel b), and performing depth-based PCA with  $q = 1$  results in two points having their SD and OD larger than cutoff (panel c). These points correspond to the time periods June 1982 to May 1983 and June 1997 to May 1998 are marked by black curves in panels a and b), and pinpoint the two seasons with strongest El-Niño events.

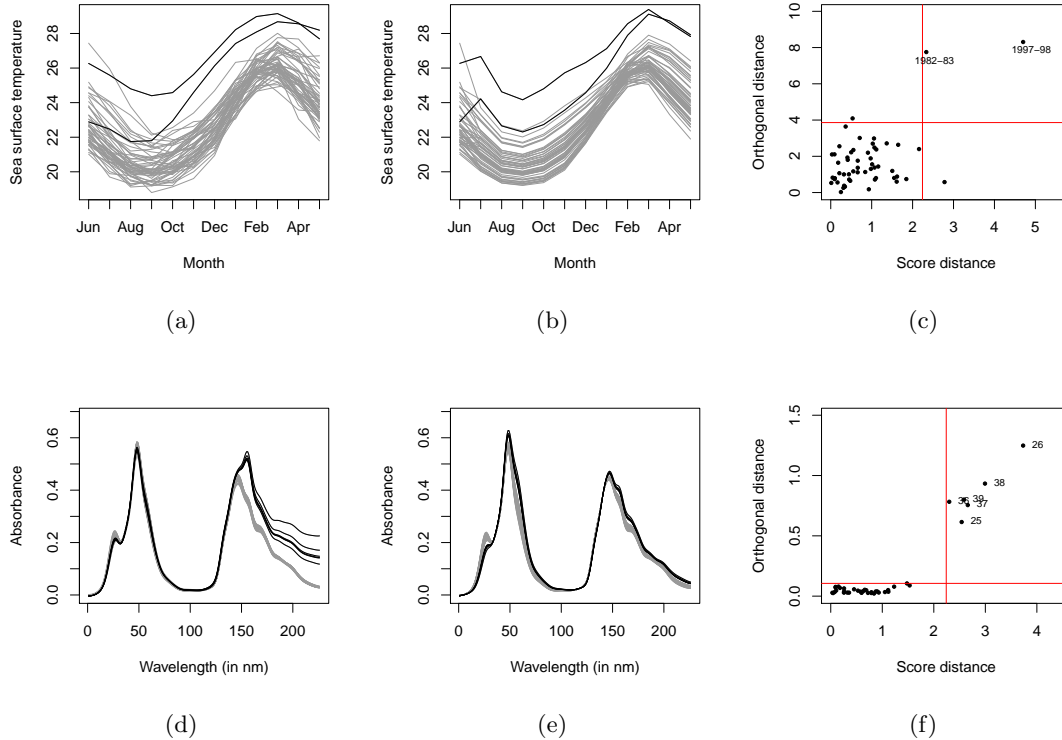


Figure 4: Actual sample curves, their spline approximations and diagnostic plots respectively for El-Niño (a-c) and Octane (d-f) datasets

Our second application is on the Octane data, which consists of 226 variables and 39 observations (Esbensen et al., 1994). Each sample is a gasoline compound with a certain octane number, and has its NIR absorbance spectra measured in 2 nm intervals between 1100 - 1550 nm. There are 6 outliers here: compounds 25, 26 and 36-39, which contain alcohol. We use the same basis structure as the one in El-Niño data here, and again the top robust PC turns out to be sufficient in identifying all 6 outliers (panels d, e and f of figure 4).

## 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 ro-

bust PCA that are based on estimation of covariance matrices (for example SCM, Tyler's scatter matrix, Dümbgen's symmetrized shape matrix).

(some text)

## 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,

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

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 ( $a \neq b \neq c$ )

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

## B Asymptotics of eigenvectors and eigenvalues

The following result allows us to obtain asymptotic joint distributions of eigenvectors and eigenvalues of the sample DCM, provided we know the limiting distribution of the sample DCM itself:

**Theorem B.1.** (*Taskinen et al., 2012*) 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

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

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 B.2.** 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. The random variable  $\text{vec}(G)$  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:

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

$$ACov(\mathbf{g}_i, \mathbf{g}_j) = -\frac{1}{(\lambda_{D,S,i} - \lambda_{D,S,j})^2} E \left[ \frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^4 \lambda_i \lambda_j z_i^2 z_j^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] \boldsymbol{\gamma}_j \boldsymbol{\gamma}_i^T; \quad i \neq j \quad (\text{B.3})$$

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:

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

$$ACov(l_i, l_j) = E \left[ \frac{(\tilde{D}\mathbf{z}(\mathbf{z}))^4 \lambda_i \lambda_j z_i^2 z_j^2}{(\mathbf{z}^T \Lambda \mathbf{z})^2} \right] - \lambda_{D,S,i} \lambda_{D,S,j}; \quad i \neq j \quad (\text{B.5})$$

*Proof of Corollary B.2.* In spirit, this corollary is similar to Theorem 13.5.1 in [Anderson \(2003\)](#). Due to the decomposition (B.1) 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

$$\sqrt{n}(\hat{\gamma}_{D,i} - \gamma_i) = \sqrt{n} \Gamma(\hat{\gamma}_{D,i}(F_\Lambda) - \mathbf{e}_i) = \sqrt{n} \left[ \sum_{k=1; k \neq i}^p \hat{\gamma}_{D,ik}(F_\Lambda) \gamma_k + (\hat{\gamma}_{D,ii}(F_\Lambda) - 1) \gamma_i \right] \quad (\text{B.6})$$

$\sqrt{n}(\hat{\gamma}_{D,ii}(F_\Lambda) - 1) = o_P(1)$  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 (B.6) 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 B.1 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.  $\square$

## C Proofs

*Proof of Proposition 2.1.* Under contiguous alternatives  $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ , the weighted sign test statistic  $T_{n,w}$  has mean  $E(w(\mathbf{Z})\mathbf{S}(\mathbf{Z}))$ . For spherically symmetric  $\mathbf{Z}$ ,  $w(\mathbf{Z})$  depends on  $\mathbf{Z}$  only through its norm. Since  $\|\mathbf{Z}\|$  and  $\mathbf{S}(\mathbf{Z})$  are independent, we get  $E(w(\mathbf{Z})\mathbf{S}(\mathbf{Z})) = Ew(\mathbf{Z}).E\mathbf{S}(\mathbf{Z})$ . The same kind of decomposition holds for  $Cov(w(\mathbf{Z})\mathbf{S}(\mathbf{Z}))$ .

We can now simplify the approximate local power  $\beta_{n,w}$  of the level- $\alpha$  ( $0 < \alpha < 1$ ) test based on  $T_{n,w}$ :

$$\begin{aligned} \beta_{n,w} &= K_p (\chi_{p,\alpha}^2 + n(E(w(\mathbf{Z})\mathbf{S}(\mathbf{Z})))^T \\ &\quad [E(w^2(\mathbf{Z})\mathbf{S}(\mathbf{Z})\mathbf{S}(\mathbf{Z})^T)]^{-1}(E(w(\mathbf{Z})\mathbf{S}(\mathbf{Z}))) \\ &= K_p \left( \chi_{p,\alpha}^2 + \frac{E^2 w(\mathbf{Z})}{E w^2(\mathbf{Z})} . E\mathbf{S}(\mathbf{Z})^T [Cov(\mathbf{S}(\mathbf{Z}))^{-1} E\mathbf{S}(\mathbf{Z})] \right) \end{aligned}$$

where  $K_p$  and  $\chi_{p,\alpha}^2$  are distribution function and upper- $\alpha$  cutoff of a  $\chi_p^2$  distribution, respectively. Since  $E^2 w(\mathbf{Z}) \leq Ew(\mathbf{Z})$ ,  $\beta_{n,w}$  the largest possible value of  $\beta_{n,w}$  is  $K_p(\chi_{p,\alpha}^2 + E\mathbf{S}(\mathbf{Z})^T [Cov(\mathbf{S}(\mathbf{Z}))^{-1} E\mathbf{S}(\mathbf{Z})])$ , the approximate power of the unweighted sign test statistic. Equality is of course achieved when  $w(\mathbf{Z})$  is a constant independent of  $\mathbf{Z}$ .  $\square$

*Sketch of proofs for equations 2.3 and 2.4.* A first step to obtain asymptotic normality for the high-dimensional location test statistic  $C_{n,w}$  is obtaining an equivalent result of Lemma 2.1 in Wang et al. (2015):

**Lemma C.1.** *Under the conditions*

$$(C1) Tr(\Sigma^4) = o(Tr^2(\Sigma^2)),$$

$$(C2) Tr^4(\Sigma) / Tr^2(\Sigma^2) \exp[-Tr^2(\Sigma) / 128p\lambda_{\max}^2(\Sigma)] = o(1)$$

when  $H_0$  is true we have

$$E[(\boldsymbol{\epsilon}_{w1}^T \boldsymbol{\epsilon}_{w2})^4] = O(1)E^2[(\boldsymbol{\epsilon}_{w1}^T \boldsymbol{\epsilon}_{w2})^2] \quad (C.1)$$

$$E[(\boldsymbol{\epsilon}_{w1}^T B_w \boldsymbol{\epsilon}_{w1})^2] = O(1)E^2[(\boldsymbol{\epsilon}_{w1}^T B_w \boldsymbol{\epsilon}_{w1})^2] \quad (C.2)$$

$$E[(\boldsymbol{\epsilon}_{w1}^T B_w \boldsymbol{\epsilon}_{w2})^2] = o(1)E^2[(\boldsymbol{\epsilon}_{w1}^T B_w \boldsymbol{\epsilon}_{w1})^2] \quad (C.3)$$

with  $\boldsymbol{\epsilon} \sim \mathcal{E}(\mathbf{0}_p, \Lambda, G)$  and  $\boldsymbol{\epsilon}_w = w(\boldsymbol{\epsilon})\mathbf{S}(\boldsymbol{\epsilon})$ .

A proof of this lemma is derived using results in section 3 of [El Karoui \(2009\)](#), noticing that any-scalar valued 1-Lipschitz function of  $\epsilon_w$  is a  $M_w$ -Lipschitz function of  $\mathbf{S}(\epsilon)$ , with  $M_w = \sup_{\epsilon} w(\epsilon)$ . Same steps as in the proof of Theorem 2.2 in [Wang et al. \(2015\)](#) follow now, using the lemma above in place of Lemma 2.1 therein, to establish asymptotic normality of  $C_{n,w}$  under  $H_0$ .

To derive the asymptotic distribution under contiguous alternatives we need the conditions (C3)-(C6) in [Wang et al. \(2015\)](#), as well as slightly modified versions of Lemmas A.4 and A.5:

**Lemma C.2.** *Given that condition (C3) holds, we have  $\lambda_{\max}(B_w) \leq 2 \frac{\lambda_{\max}}{\text{Tr}(\Sigma)}(1 + o(1))$ .*

**Lemma C.3.** *Define  $D_w = E \left[ \frac{w^2(\epsilon)}{\|\epsilon\|^2} (I_p - \mathbf{S}(\epsilon)\mathbf{S}(\epsilon)^T) \right]$ . Then  $\lambda_{\max}(A_w) \leq E(w(\epsilon)/\|\epsilon\|)$  and  $\lambda_{\max}(D_w) \leq E(w(\epsilon)/\|\epsilon\|)^2$ . Further, if (C3) and (C4) hold then  $\lambda_{\min}(A_w) \geq E(w(\epsilon)/\|\epsilon\|)(1 + o(1))/\sqrt{3}$ .*

The proof now exactly follows steps in the proof of theorem 2.3 in [Wang et al. \(2015\)](#), replacing vector signs by weighted signs, using the fact that  $w(\epsilon)$  is bounded above by  $M_w$  while applying conditions (C5)-(C6) and lemmas A.1, A.2, A.3, and finally using the above two lemmas in place of lemmas A.4 and A.5 respectively.  $\square$

*Proof of Theorem 3.1.* The proof follows directly from writing out the expression of  $\text{Cov}(\tilde{\mathbf{X}})$ :

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

$\square$

*Proof of Lemma 3.2.* 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 due to the assumption of uniform convergence, given  $\epsilon > 0$  we can find  $N \in \mathbb{N}$  such that

$$|(\tilde{D}_{\mathbf{X}}^{n_1}(\mathbf{x}_i))^2 - (\tilde{D}_{\mathbf{X}}(\mathbf{x}_i))^2| < \epsilon \quad (\text{C.4})$$

for all  $n_1 \geq N; i = 1, 2, \dots, n_1$ . This implies

$$\begin{aligned} S_{n_1} &< \epsilon \sqrt{n_1} \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} SS(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_{n_1}) \right] \\ &= \epsilon \sqrt{n_1} \left[ \frac{1}{n_1} \sum_{i=1}^{n_1} \{SS(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_{n_1}) - SS(\mathbf{x}_i; \boldsymbol{\mu})\} + \frac{1}{n_1} \sum_{i=1}^{n_1} SS(\mathbf{x}_i; \boldsymbol{\mu}) \right] \quad (\text{C.5}) \end{aligned}$$

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

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

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

where  $N_k \in \mathbb{N}$  gives the relation (C.4) in place of  $N$  when we take  $\epsilon = 1/k$ . Under conditions  $E\|\mathbf{x} - \boldsymbol{\mu}\|^{-3/2} < \infty$  and  $\sqrt{n}(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}) = O_P(1)$ , the sample SCM with unknown location parameter  $\hat{\boldsymbol{\mu}}_n$  has the same asymptotic distribution as the SCM with known location  $\boldsymbol{\mu}$  (Dürre et al., 2014), hence  $B_k = o_P(1)$ , thus  $A_k(B_k + C_k) \xrightarrow{P} 0$ .

Now (C.5) implies that for any  $\epsilon_1 > 0$ ,  $S_{N_k} > \epsilon_1 \Rightarrow A_k(B_k + C_k) > \epsilon_1$ , which means  $P(S_{N_k} > \epsilon_1) < P(A_k(B_k + C_k) > \epsilon_1)$ . Hence the subsequence  $\{S_{N_k}\} \xrightarrow{P} 0$ . Since the main sequence  $\{S_k\}$  is bounded below by 0, this implies  $\{S_k\} \xrightarrow{P} 0$ . Finally, we have that

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

$$S_n + \sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^n \{SS(\mathbf{x}_i; \hat{\boldsymbol{\mu}}_n) - SS(\mathbf{x}_i; \boldsymbol{\mu})\} \right] \quad (\text{C.6})$$

Since the second summand on the right hand side is  $o_P(1)$  due to Dürre et al. (2014) as mentioned before, we have the needed.  $\square$

*Proof of Theorem 3.3.* 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 3.2, and applying Slutsky's theorem we get the required convergence.  $\square$

*Proof of Theorem 3.4.* We are going to prove the following:

1.  $\|\hat{\Gamma}_D - \Gamma\|_F \xrightarrow{P} 0$ , and
2.  $\|\hat{\Lambda} - \Lambda\|_F \xrightarrow{P} 0$

as  $n \rightarrow \infty$ . For (1), we notice  $\sqrt{n} \text{vec}(\hat{\Gamma}_D - \Gamma)$  asymptotically has a (singular) multivariate normal distribution following Corollary B.2, so that  $\|\hat{\Gamma}_D - \Gamma\|_F = O_P(1/\sqrt{n})$  using Prokhorov's theorem.

It is now enough to prove convergence in probability of the individual eigenvalue estimates  $\hat{\lambda}_i; i = 1, \dots, p$ . For this, define estimates  $\tilde{\lambda}_i$  as median-of-small-variances estimator of the *true* score vectors  $\Gamma^T X$ . For this we have

$$|\tilde{\lambda}_i - \lambda_i| \xrightarrow{P} 0 \quad (\text{C.7})$$

using Theorem 3.1 of [Minsker \(2015\)](#), with  $\mu = \lambda_i$ . Now  $\hat{\lambda}_i = \text{med}_j(\text{Var}(X_{G_j}^T \hat{\gamma}_{D,i}))$  and  $\tilde{\lambda}_i = \text{med}_j(\text{Var}(X_{G_j}^T \gamma_i))$ , so that

$$\begin{aligned} |\hat{\lambda}_i - \tilde{\lambda}_i| &\leq \text{med}_j \left[ \text{Var}(X_{G_j}^T (\hat{\gamma}_{D,i} - \gamma_i)) \right] \\ &\leq \|\hat{\gamma}_{D,i} - \gamma_i\|^2 \text{med}_j [\text{Tr}(\text{Cov}(X_{G_j}))] \end{aligned}$$

using Cauchy-Schwarz inequality. Combining the facts  $\|\hat{\gamma}_{D,i} - \gamma_i\| = O_P(1/\sqrt{n})$  and  $\text{med}_j[\text{Tr}(\text{Cov}(X_{G_j}))] \xrightarrow{P} \text{Tr}(\Sigma)$  ([Minsker, 2015](#)) with [\(C.7\)](#), we get the needed.  $\square$

## D Introduction

(needs rewriting)

In multivariate analysis, the study of principal components is important since it provides a small number of uncorrelated variables from a potentially larger 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 address this issue, several robust estimators of the population covariance or correlation matrix have been proposed which can be used for Principal Components Analysis (PCA). They can be roughly put into these categories: robust, high breakdown point estimators that are computation-intensive (Maronna et al., 1976; Rousseeuw, 1985); M-estimators that are calculated by simple iterative algorithms but do not necessarily possess high breakdown point (Huber, 1977; Tyler, 1987); and symmetrised estimators that are highly efficient and robust to deviations from normality, but sensitive to outliers and computationally demanding (Dümbgen, 1998; Sirkiä et al., 2007).

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. The covariance matrix of these sign vectors, namely Sign Covariance Matrix (SCM) has the same set of eigenvectors as the covariance matrix of the original population, thus the multivariate sign transformation yields computationally simple and high-breakdown estimates of principal components (Locantore et al., 1999; Visuri et al., 2000). Although the 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 (Magyar and Tyler, 2014), 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.

The nonparametric concept of data-depth had first been proposed by Tukey (1975) 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 (Zuo and Serfling, 2000). Depth-based methods have recently been popular for robust nonparametric classification (Dutta and Ghosh, 2012; Ghosh and Chaudhuri, 2005; Jornsten, 2004; Sguera et al., 2014). In parametric estimation, depth-weighted means (Zuo et al., 2004) and covariance matrices (Zuo and Cui, 2005) provide high-breakdown point as well as efficient estimators, although they do involve choice of a suitable weight function and tuning parameters. In this paper we study the covariance matrix of the multivariate rank vector that is obtained from the data-depth of a point and its spatial sign, paying special attention to its eigenvectors. Specifically, we develop a robust version of principal components analysis for elliptically symmetric distributions based on the eigenvectors of this covariance matrix, and compare it with normal PCA and spherical PCA, i.e. PCA based on eigenvectors of the SCM.

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

(Möttönen and Oja, 1995):

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

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  $\mathbf{S}(\mathbf{x}-\boldsymbol{\mu})$  reside 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 of population covariance matrix 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 (Taskinen et al., 2012).

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  $\Sigma_S(P\mathbf{X}) = P\Sigma_S(\mathbf{X})P^T$ . This is not necessarily true in general if  $P$  is replaced by any non-singular matrix. An affine equivariant version of the sample SCM is obtained as the solution  $\hat{\Sigma}_T$  of the following equation:

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

which turns out to be Tyler's M-estimator of scatter (Tyler, 1987). 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  $k > 0$ .

For any multivariate distribution  $F = F_{\mathbf{X}}$  belonging to a set of distributions  $\mathcal{F}$ , 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  $F$  (Zuo and Serfling, 2000). Liu (1990) outlines the desirable properties of a statistical depth function:

- (D1) *Affine invariance*:  $D(A\mathbf{x} + \mathbf{b}, F_{A\mathbf{X}+\mathbf{b}}) = D(\mathbf{x}, F_{\mathbf{X}})$ ;
- (D2) *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.;
- (D3) *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  $F_{\mathbf{X}}$ .;
- (D4) *Vanishing at infinity*:  $D(\mathbf{x}; F_{\mathbf{X}}) \rightarrow 0$  as  $\|\mathbf{x}\| \rightarrow \infty$ .

In (D2) the types of symmetry considered can be central symmetry, angular symmetry and halfspace symmetry. Also for multimodal probability distributions, i.e. distributions with multiple local maxima in their probability density functions, properties (D2) and (D3) are actually restrictive towards the formulation of a reasonable depth function that captures the shape of the data cloud. In our derivations that follow, we replace these two by a weaker condition:

(D2\*) *Existence of a maximal point*: The maximum depth over all distributions  $F$  and points  $\mathbf{x}$  is bounded above, i.e.  $\sup_{F \in \mathcal{F}} \sup_{\mathbf{x} \in \mathbb{R}^p} D(\mathbf{x}, F) < \infty$ . We denote this point by  $M_D(F_{\mathbf{X}})$ .

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

- **Halfspace depth** (HD) (Tukey, 1975) 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) (Liu et al., 1999) 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 htped function of  $\mathbf{x}$  with respect to  $F$ .

- **Projection depth** (PD) (Zuo, 2003) 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 (Rainer and Mozharovskiy, 2014) 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.

## References

- Adraghi, K. P. and Cook, R. D. (2009). Sufficient dimension reduction and prediction in regression. *Phil. Trans. R. Soc. A*, 367:4385–4405.
- Anderson, T. (3rd ed. 2003). *An Introduction to Multivariate Statistical Analysis*. Wiley, Hoboken, NJ.

- Boente, G. and Salibián-Barrera, M. (2015). S-Estimators for Functional Principal Component Analysis. *J. Amer. Statist. Assoc.*, 110:1100–1111.
- Brown, B. (1983). Statistical Use of the Spatial Median. *J. Royal Statist. Soc. B*, 45:25–30.
- Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.*, 91:862–872.
- Chen, S. X. and Qin, Y. L. (2010). A Two-sample Test for High-dimensional Data with Application to Gene-Set Testing. *Ann. Statist.*, 38:808–835.
- Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. (2014). Monge-Kantorovich Depths, Quantiles, Ranks and Signs. <http://arxiv.org/abs/1412.8434>.
- Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Biometrika*, 20:927–1010.
- 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.
- Dümbgen, L. (1992). Limit theorems for the simplicial depth. *Statist. Probab. Lett.*, 14:119–128.
- Dümbgen, L. (1998). On Tyler’s M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.*, 50:471–491.
- Dürre, A., Vogel, D., and Tyler, D. (2014). The spatial sign covariance matrix with unknown location. *J. Mult. Anal.*, 130:107–117.
- Dutta, S. and Genton, M. G. (2017). Depthweighted robust multivariate regression with application to sparse data. *Can. J. Stat.*, 45(2):164–184.
- Dutta, S. and Ghosh, A. (2012). On robust classification using projection depth. *Ann. Inst. Stat. Math.*, 64-3:657–676.
- El Karoui, N. (2009). Concentration of Measure and Spectra of Random Matrices: with Applications to Correlation Matrices, Elliptical Distributions and Beyond. *Ann. Applied Probab.*, 19:2362–2405.
- Esbensen, K. H., Schönkopf, S., and Midtgaard, T. (1994). *Multivariate Analysis in Practice*. CAMO, Trondheim, Germany.
- Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*. Monographs on Statistics and Applied Probability 36. Chapman and Hall Ltd., London, United Kingdom.
- Ghosh, A. and Chaudhuri, P. (2005). On Maximum Depth and Related Classifiers. *Scand. J. Statist.*, 32:327–350.



- Haldane, J. (1948). Note on the Median of a Multivariate Distribution. *Biometrika*, 35:414–415.
- Hallin, M. and Paindaveine, D. (2002). Optimal tests for multivariate location based on interdirections and pseudo-Mahalanobis ranks. *Ann. Statist.*, 30:1103–1133.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Staehl, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley.
- Huber, P. (1977). *Robust Statistical Procedures*. Society for Industrial and Applied Mathematics, Philadelphia.
- Huber, P. J. (1981). *Robust Statistics*. Wiley series in probability and mathematical statistics. Wiley.
- Jornsten, R. (2004). Clustering and classification based on the  $l_1$  depth. *J. Mult. Anal.*, 90-1:67–89.
- Liu, R. (1990). On a notion of data depth based on random simplices. *Ann. of Statist.*, 18:405–414.
- Liu, R., Parelius, J., and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Ann. of Statist.*, 27:783–858.
- Locantore, N., Marron, J., Simpson, D., Tripoli, N., Zhang, J., and Cohen, K. (1999). Robust principal components of functional data. *TEST*, 8:1–73.
- Magyar, A. and Tyler, D. (2014). The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions. *Biometrika*, 101:673–688.
- Majumdar, S. and Chatterjee, S. (2018). Non-convex penalized multitask regression using data depth-based penalties. *Stat*, 7:e174.
- Maronna, R., Staehl, W., and Yohai, V. (1976). Bias-Robust Estimators of Multivariate Scatter Based on Projections. *J. Mult. Anal.*, 42:141–161.
- Miao, J. and Ben-Israel, A. (1992). On principal angles between subspaces in  $\mathbb{R}^n$ . *Lin. Algeb. Applic.*, 171:81–98.
- Minsker, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli*, 21:2308–2335.
- Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametric Stat.*
- Oja, H. (1983). Descriptive Statistics for Multivariate Distributions. *Statist. and Prob. Lett.*, 1:327–332.
- Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. Lecture Notes in Statistics. Springer.

- Ollilia, E., Oja, H., and Croux, C. (2003). The affine equivariant sign covariance matrix: asymptotic behavior and efficiencies. *J. Mult. Anal.*, 87:328–355.
- Puri, M. L. and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis*. Wiley, New York, NY.
- Rainer, D. and Mozharovskyi, P. (2014). Exact computation of the halfspace depth. <http://arxiv.org/abs/1411.6927>.
- Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications, Volume B (Proc. 4th Pannonian Symp. Math. Statist., Bad Tatzmannsdorf, 1983)*, pages 283–97, Dordrecht. D. Reidel.
- Serfling, R. (2006). Depth Functions in Nonparametric Multivariate Inference. In *DI-MACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 72, pages 1–16.
- Sguera, C., Galeano, P., and Lillo, R. (2014). Spatial depth-based classification for functional data. *TEST*, 23-4:725–750.
- Sirkiä, S., Taskinen, S., and Oja, H. (2007). Symmetrised M-estimators of scatter. *J. Mult. Anal.*, 98:1611–1629.
- Taskinen, S., Koch, I., and Oja, H. (2012). Robustifying principal component analysis with spatial sign vectors. *Statist. and Prob. Lett.*, 82:765–774.
- Tukey, J. (1975). Mathematics and picturing data. In James, R., editor, *Proceedings of the International Congress on Mathematics*, volume 2, pages 523–531.
- Tyler, D. (1987). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.*, 15:234–251.
- Visuri, S., Koivunen, V., and Oja, H. (2000). Sign and rank covariance matrices. *J. Statist. Plan. Inf.*, 91:557–575.
- Wang, L., Peng, B., and Li, R. (2015). A High-Dimensional Nonparametric Multivariate Test for Mean Vector. *J. Amer. Statist. Assoc.*, 110:1658–1669.
- Zuo, Y. (2003). Projection-based depth functions and associated medians. *Ann. Statist.*, 31:1460–1490.
- Zuo, Y. and Cui, M. (2005). Depth weighted scatter estimators. *Ann. Statist.*, 33-1:381–413.
- Zuo, Y., Cui, M., and He, X. (2004). On the Staehl-Donoho estimator and depth-weighted means of multivariate data. *Ann. Statist.*, 32-1:167–188.
- Zuo, Y. and Serfling, R. (2000). General notions of statistical depth functions. *Ann. Statist.*, 28-2:461–482.

$F = \text{Bivariate } t_5$	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 4: 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 5: 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 6: Finite sample efficiencies of several scatter matrices:  $p = 4$