Electronic Journal of Statistics
Vol. 0 (0000)
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

On Weighted Multivariate Sign Functions

Subhabrata Majumdar* **and Snigdhansu Chatterjee†**

School of Statistics,
University of Minnesota,
313 Ford Hall,
224 Church Street S.E.,
Minneapolis 55455, USA
e-mail: subho@research.att.com; chatt019@umn.edu

Abstract: Multivariate sign functions are often used for robust estimation and inference. We propose using data dependent weights in association with such functions. The proposed weighted sign functions retain desirable robustness properties, while significantly improving efficiency in estimation and inference compared to unweighted multivariate sign-based methods. Using weighted signs, we demonstrate methods of robust location estimation and robust principal component analysis. We extend the scope of using robust multivariate methods to include robust sufficient dimension reduction and functional outlier detection. Several numerical studies and real data applications demonstrate the efficacy of the proposed methodology.

MSC 2010 subject classifications: Primary 62G35; secondary 62H25, 62H20, 62G99.
Keywords and phrases: Multivariate sign, Principal component analysis, Data depth, Sufficient dimension reduction.

Contents

*Currently at AT&T Labs Research
†Corresponding author

1. Introduction

Given a point μ in a normed linear space $\mathcal{X}$ with norm denoted by $|\cdot|$, the *generalized sign function* $S : \mathcal{X} \times \mathcal{X} \mapsto \mathcal{X}$ with center μ is defined as

$$S(x;\mu) = \begin{cases} |x-\mu|^{-1}(x-\mu) & \text{if } x \neq \mu, \\ 0 & \text{if } x = \mu. \end{cases} \tag{1}$$

This is a functional and multivariate generalization of the real-valued *sign function*, that takes the values one, negative one or zero if the point $x \in \mathbb{R}$ is to the right, left or equal $\mu \in \mathbb{R}$ respectively. This generalized sign function was introduced by [25] for $\mathcal{X} = \mathbb{R}^p$, the p-dimensional real Euclidean space.

The function S maps μ to the origin and all other points of $\mathcal{X}$ to the unit sphere $\mathcal{S}_{0;1} = \{x \in \mathcal{X} : |x| = 1\}$. Given a dataset $\{X_i \in \mathbb{R}^p : i = 1, \ldots, n\}$, that we collect together in the $n \times p$ matrix $\mathbf{X} = (X_1 : \ldots : X_n)^T$, an approach for robust estimation and inference in multivariate data starts by evaluating (1) on each observation, thus defining $S_i = S(X_i; \mu_x)$ with respect to some suitable center $\mu_x \in \mathbb{R}^p$, and then using these for robust location and scale estimation and inference, including inference for μ_x [20, 27, 33]. Suppose $\mathbf{S} = (S_1 : \ldots : S_n)^T \in \mathbb{R}^{n \times p}$. If the data $\{X_i \in \mathbb{R}^p : i = 1, \ldots, n\}$ are independent, identically distributed (hereafter, i.i.d.) from an elliptically symmetric distribution, then the eigenvectors of $\mathbb{E}(X_1 - \tilde{\mu})(X_1 - \tilde{\mu})^T$ and of $\mathbb{E}S_1 S_1^T$ are the same for suitable centering parameters $\tilde{\mu}$ and μ_x, that is, the population principal components from the original data and from its sign transformations are the same [31]. However, 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 S are asymptotically inadmissible [21] and Tyler's M-estimate of scatter [32] has uniformly lower asymptotic risk.

In this paper, we propose to alleviate this low efficiency problem, by associating a data-driven weight W_i with the generalized sign S_i, that can be used to adaptively trade-off between efficiency and robustness considerations in any given application. We demonstrate the utility of using the proposed *weighted generalized sign* functions in a number of problems of current interest, including robust estimation of location and scatter.

Specifically, we propose using product of the generalized sign function and a weight function derived as a transformation of a data-depth function [28, 34]. Like data-depth functions, the weight functions used in this paper are non-negative reals defined over $\mathcal{X} \times \mathcal{F}$, where $\mathcal{F}$ is a fixed family of probability measures. For every choice of parameters $\mu \in \mathcal{X}$ and $\mathbb{F} \in \mathcal{F}$, in this paper

$$R(X_i; \mu, \mathbb{F}) = \mu + S(X_i, \mu) W(X_i, \mathbb{F}). \tag{2}$$

is used as a robust surrogate for observation X_i. Notice that for the trivial choice $W(x, \mathbb{F}) = |x - \mu_{\mathbb{F}}|$, $\mu = \mu_{\mathbb{F}}$, we get $R(X_i; \mu, \mathbb{F}) = X_i$, the original observations. With the other trivial choice of $W(x, \mathbb{F}) \equiv 1$, we get the generalized sign $R(X_i; \mu, \mathbb{F}) = S(X_i, \mu) = S_i$. However, in this paper we illustrate how using other weight functions can lead to interesting robustness and efficiency trade-offs in a variety of situations. The various technical conditions and assumptions that we impose on the weight function $W(x, \mathbb{F})$ are valid for weights derived from three well-known data depth functions: the *half-space depth*, the *Mahalanobis depth*, and the *projection depth*.

We primarily focus on the task of *robust dispersion/scatter estimation* and *robust principal component analysis* in this paper. Figure 1 presents an illustrative example of bivariate data with outliers in the top left panel, where the outliers are marked with red points. In the other panels, the generalized sign values of the same data are presented as black points on the unit circle, with the outliers again marked with red points. Notice that the black points from either the top right or bottom panels have very similar eigenvector structure as the original data without the outliers. The green and blue triangles are examples of the proposed *weighted sign* values: the top right (respectively, bottom left, bottom right) panels depict these values where the weights have been generated using Mahalnobis depth (respectively Tukey's half-space depth and the projection depth). The blue triangles are the weighted sign values of the outliers. Notice that the eigenvectors from the weighted signs also capture the pattern from the original data without the outliers.

There are two unknown quantities in the generalized sign function as defined in (1): μ and $\mathbb{F}$, To estimate dispersion and its eigen-structure robustly, we must start with a robust estimator for μ. In Section 2 we present the case for *weighted spatial quantiles*, which can be defined and studied in very general spaces $\mathcal{X}$. One special case of this is the *weighted spatial median*. As a location estimator, it has several interesting robustness properties and can be shown to be more efficient that some existing robust location estimators, thus making it a perfect candidate to estimate μ. On the other hand, setting $\mathbb{F}$ as $\mathbb{F}_X$, i.e. the distribution of X has the clear interpretation of differentially weighting observations based on their inlyingness to the overall data distribution. This mirrors the context of how data depth has been used in the literature [19, 28]. Due to this reason, we assume $\mathbb{F} \equiv \mathbb{F}_X$ for the rest of the paper. As and when required, we shall use sample versions of these quantities, stating the theoretical conditions for the corresponding approximation results to hold.

Following that, we restrict to the $\mathbb{R}^p$ for fixed p, and present detailed discussions on our primary proposal for a robust measure of dispersion in Section 3, followed by a proposed *affine equivariant version* of it in Section 4, robust estimation of eigenvalues and a third robust estimator for dispersion in Section 5, and a thorough study of robustness and efficiency using influence functions in Section 6. We then report multiple simulation-based numeric studies in Section 7, present several real data examples in Section 8, and concluding remarks in Section 9.

In the rest of this paper, all finite-dimensional vectors are column vectors, and

for a vector or matrix a, the notation a^T stands for its transpose. The Gaussian distribution with mean μ and variance Σ is denoted by $N(\mu, \Sigma)$. The identity matrix is denoted by $\mathbb{I}$, with or without a subscript to denote its dimension. The notations $A^{-1}, det(A), \lambda_{\min}(A), \lambda_{\max}(A)$ respectively stand for the inverse, determinant, minimum and maximum eigenvalues of a matrix A, whenever these are well-defined. For a scalar or vector valued random variable Y, $\mathbb{E}Y$ denotes its expected value, while $\mathbb{V}Y$ denotes its variance or covariance matrix.

2. The Weighted Spatial Median

Suppose the open unit sphere in $\mathcal{X}$ is given by $\text{int}\mathcal{X}_{0;1} = \{x \in \mathcal{X} : |x| < 1\}$, and let $u \in \text{int}\mathcal{X}_{0;1}$. We also fix the set of probability measures $\mathcal{M}$, and select $\mathbb{F} \in \mathcal{M}$. Consider a random element $X \in \mathcal{X}$. We define the $(u, \mathbb{F})$-th *weighted spatial quantile* of $\mathcal{X}$ as the minimizer $q(u, \mathbb{F}) \in \mathcal{X}$ of

$$\Psi(q; u, \mathbb{F}) = \mathbb{E}\Big[W(X, \mathbb{F})\big\{|X - q| + \langle u, X - q\rangle\big\}\Big]. \tag{3}$$

This is a natural generalization of the spatial median [7, 14, 18] ($W(X, \mathbb{F}) \equiv 1$ and $u = \mathbf{0}_p$), or more general spatial quantiles [5, 6, 24] ($W(X, \mathbb{F}) \equiv 1$).

In what follows, for brevity we elaborate only the case of the *weighted spatial median* (thus $\Psi(q; 0, \mathbb{F}) = \mathbb{E}\big[W(X, \mathbb{F})|X - q|\big]$) for the case of finite dimensional $\mathcal{X}$. Specifically, we demonstrate the utility of using the *weighted* spatial median as opposed to using the traditional, unweighted versions found in literature. The analysis and computation of the general weighted spatial quantiles will largely follow by extending the results of the above cited references, and we postpone details to a future study.

Assume that we have independent and identically distributed observations $X_1, \ldots, X_n \in \mathcal{X}$, and the sample weighted spatial median is computed by minimizing $\Psi_n(q; 0, \mathbb{F}) = \sum_{i=1}^n W(X_i, \mathbb{F})|X_i - q|$, and is denoted by $\hat{q}_{nW}$, the second subscript is in acknowledgement that the weight function is used. We denote the unweighted version of this estimator, ie, the case where $W(X, \mathbb{F}) \equiv 1$ as $\hat{q}_n$. Assume the following technical conditions:

(A1) $\Psi(q; 0, \mathbb{F})$ is finite for all $q \in \mathcal{X} \subseteq \mathbb{R}^p$ and has a unique minimizer q_0.
(A2) $\Psi(q; 0, \mathbb{F})$ is twice differentiable at q_0 and the second derivative is positive definite.
(A3) $\mathbb{E}W^2(X, \mathbb{F})S(X; q)S^T(X; q)$ exists for all q in a neighborhood of q_0.

These assumptions are very broad-based and general. The first one essentially requires the existence of a population parameter, the second one requires that the minimization approach is meaningful in the population, and the third one essentially requires that the weight function has a finite variance. No further restrictions are placed on the tuning parameter $\mathbb{F}$ or the choice of the weight function.

Theorem 2.1. *Under assumptions [A1]-[A3], we have*

$$n^{1/2}(\hat{q}_{nW} - q_0) \stackrel{\mathcal{D}}{\Rightarrow} N(0, \Psi_{2W}^{-1}\Psi_{1W}\Psi_{2W}^{-1}), \text{ where}$$
$$\Psi_{2W} = \mathbb{E}W(X,\mathbb{F})\Big[|X-q_0|^{-1}(\mathbb{I}_p - S(X;q_0)S^T(X;q_0))\Big]$$
$$\Psi_{1W} = \mathbb{E}W^2(X,\mathbb{F})S(X;q_0)S^T(X;q_0).$$

Thus, under very standard regularity conditions, the sample weighted spatial median is consistent and is asymptotically normal. Theorem 2.1 can be proved in several different ways. Here we use techniques following [13, 26]. Specifically, following Theorem 4 in [26], which traces back to [13] with slightly relaxed conditions, we get

$$n^{1/2}(\hat{q}_{nW} - q_0) = -\frac{\Psi_{2W}}{\sqrt{n}}\sum_{i=1}^{n} W(X_i,\mathbb{F})S(X_i;q_0) + o_P(1).$$

Theorem 2.1 follows by applying the central limit theorem.

Remark. Note that for the result in Theorem 2.1 to go through, it is not necessary for $\mathbb{F}$ to be the distribution of X. Other choices of $\mathbb{F}$, e.g. [22], may lead to interesting interpretations of $W(.,\mathbb{F})$ and the resulting location estimator and can be explored further.

2.1. Asymptotic efficiency of weighted spatial median

Let $V_W = \Psi_{2W}^{-1}\Psi_{1W}\Psi_{2W}^{-1}$ be the asymptotic variance of $\hat{q}_{n,W}$ from Theorem 2.1, where we use the subscript "$_W$" to denote that this depends on the weight function. We use the notation V_1 for the case where $W(x,\mathbb{F}) \equiv 1$, that is, all weights are one. The asymptotic relative efficiency of two statistics is the p-th root of the reciprocals of their determinants. That is,

$$ARE(\hat{q}_{nW}, \hat{q}_n) = \Big\{\frac{det(V_1)}{det(V_W)}\Big\}^{1/p}.$$

One easy result from Theorem 2.1 is that under reasonable conditions the asymptotic relative efficiency of the weighted spatial median over the unweighted version is always greater than one. We document this in the following corollary:

Corollary 2.2. *Assume that the weight function $W(X,\mathbb{F})$ is bounded above by k for some $k > 0$, and the matrices $\Psi_1 = \mathbb{E}S(X;q_0)S^T(X;q_0)$ and Ψ_{1W} are positive definite. Then*

$$ARE(\hat{q}_{nW}, \hat{q}_n) \geq \frac{\lambda_{\min}(\Psi_1)\lambda^2_{\min}(\Psi_{2W})}{k\lambda_{\max}(\Psi_{1W})\lambda^2_{\max}(\Psi_2)}$$

Consequently, if $k/\lambda^2_{\min}(\Psi_{2W}) < \lambda_{\min}(\Psi_1)/(\lambda_{\max}(\Psi_{1W})\lambda^2_{\max}(\Psi_2))$ then this asymptotic relative efficiency is larger than 1.

Proof of Corollary 2.2. Using the facts that $det(AB) = det(A)det(B)$ for square matrices A, B and $det(A^{-1}) = 1/det(A)$ for non-singular A, we write

$$\begin{aligned}\frac{det(V_1)}{det(V_W)} &= det(\Psi_2^{-1}\Psi_1\Psi_2^{-1})det(\Psi_{2W}\Psi_{1W}^{-1}\Psi_{2W}) \\ &= det(\Psi_1)det(\Psi_{1W}^{-1})[det(\Psi_2^{-1})det(\Psi_{2W})]^2\end{aligned}$$

The result follows, using the facts that $\det(A) \geq \lambda_{\min}(A)$ and $\det(A^{-1}) \geq 1/\lambda_{\max}(A)$, and the upper bound on W. □

It is possible to make the above asymptotic relative efficiency to tend to infinity with p, for example by choosing $k = \exp\{-p^2\}$. Such tending to infinity efficiencies are also achievable for fixed k, for example, if $|X - q_0|^2$ is a Gamma random variable with shape parameter $= 2 + \exp\{-p^2\}$. We leave it to future work to study such efficiencies more generally.

We may wish to further explore the conditions of Corollary 2.2. Let us concentrate on the case of where the distribution of X is spherically symmetric. Following [12]:

Definition 2.3. A p-dimensional random vector X is said to elliptically distributed if there exist a vector $\mu \in \mathbb{R}^p$, a positive semi-definite matrix $\Sigma \in \mathbb{R}^{p\times p}$ and a function $\phi : (0, \infty) \to \mathbb{R}$ such that the characteristic function of X is $\exp\{it^T\mu\}\phi(\mathbf{t}^T\Sigma\mathbf{t})$ for $\mathbf{t} \in \mathbb{R}^p$.

There are several alternative formulations, see the above reference and citations within it for details.

Corollary 2.4. *Assume that $\mathbb{F} \equiv \mathbb{F}_X$ is an elliptically symmetric distribution, with location parameter $\mu = q_0$ and the covariance matrix Σ satisfies the following conditions:*

1. $\exp\left\{-\frac{Tr^2(\Sigma)}{256\lambda^2_{\max}(\Sigma)}\right\} = o\left(\min\left(\frac{\lambda_{\max}(\Sigma)}{Tr(\Sigma)}, \frac{\lambda_{\min}(\Sigma)}{\lambda_{\max}(\Sigma)}\right)\right)$,
2. $\lambda_{\max}(\Sigma) = o(Tr(\Sigma))$.

Also assume that the weight function is (a) bounded above by some $W_{\max} > 0$, (b) affine invariant, i.e. $W(Ax + b, A\mathbb{F} + b) = W(x, \mathbb{F})$ for $A \in \mathbb{R}^{p\times p}, b \in \mathbb{R}^p$, and (c) satisfies the following:

$$\frac{W_{\max}}{[\mathbb{E}|X - q_0|^{-1}W(X, \mathbb{F})]^2} < \frac{\lambda_{\min}(\Psi_1)}{\lambda_{\max}(\Psi_1)[\mathbb{E}|X - q_0|^{-1}]^2]}. \tag{4}$$

Then we have $ARE(\hat{q}_{n,W}, \hat{q}_n) > 1$.

The conditions 1 and 2 in Corollary 2.4 are due to [33], and ensure that the eigenvalues of Σ are bounded away from 0 and ∞. Corollary 2.4 can be proved using Corollary 2.2, the fact that for elliptical distributions Ψ_1 is non-singular [31], then using and expanding upon Lemma A.5 in [33]. We give the technical details in the supplementary material.

Obtaining affine invariant weight functions is not challenging. Most functions arising in the context of data depths are affine invariant. Corollary 2.4 implies

that there is a wide frame of distributions and choices of weight functions where there is a benefit to considering weighted spatial medians. In fact, the choice of weight functions is even broader: only location invariance is required for Corollary 2.4 to hold. We only choose to restrict ourselves here because the subsequent analysis uses weights based on affine invariant data depth functions. Also note that in case (4) does not hold for the original form of a weight function, one can scale all weights by an appropriate constant to reduce the value in the left hand side of (4) to satisfy the condition.

In actual computations, in place of $W(x, \mathbb{F})$ we propose using $W(x, \mathbb{F}_n)$ where $\mathbb{F}_n(z) = (np)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{p} \mathcal{I}(X_{ij} \leq z_j), z \in \mathbb{R}^p$ is the empirical distribution function. Up to first order asymptotics, the analysis remain unchanged from the above for this modified weight function.

2.2. Examples of affine invariant weights

We now illustrate some specific choices of weight functions that are compatible with the conditions of Corollary 2.4 and the results in the rest of this paper. These arise as easy transformations of *data-depth functions*. A data depth function is defined on $\mathcal{X} \times \mathcal{F}$, where $\mathcal{F}$ is a fixed set of probability measures. The main property of a data-depth function is that for every probability measure $\mathbb{F} \in \mathcal{F}$, there exists a constant $\mu_{\mathbb{F}} \in \mathcal{X}$ such that for any $t \in [0, 1]$ and $x \in \mathcal{X}$,

$$D(\mu_{\mathbb{F}}; \mathbb{F}) \geq D(\mu_{\mathbb{F}} + t(\mathbf{x} - \mu_{\mathbb{F}}); \mathbb{F}). \tag{5}$$

That is, for every fixed $\mathbb{F}$, the data-depth function achieves a supremum at μ_F, and is non-decreasing in every direction away from μ_F, thus providing a center-outward partial ordering of points in $\mathcal{X}$. There are generally several algebraic and analytic properties assumed for data-depth functions to elicit interesting mathematical properties, see for example [28, 34] for details.

The spherically symmetric case of an elliptical distribution is realized with $\Sigma = \sigma^2 \mathbb{I}_p$ for some $\sigma^2 > 0$. We fix the notation $Z = \Sigma^{-1/2}(X - \mu)$, and let $Z \sim \mathbb{F}_Z$. Note that $\mathbb{F}_Z$ is a spherically symmetric distribution and hence depends only on $|z|$, and $\mathbb{E}Z = \mathbf{0}_p \in \mathbb{R}^p$ and $\mathbb{V}Z = \mathbb{I}_p$. Taking affine invariant data depth functions as weights ensures that $W(X, \mathbb{F}) = W(Z, \mathbb{F}_Z)$. It is now easy to show that results in this paper are valid for the weight functions (with appropriate scaling to satisfy (4)) (*i*) $W_{HSD}(X) = \mathbb{F}_Z(|Z|)$ derived from the *half-space depth*, (*ii*) $W_{MhD}(X) = |Z|^2/(1 + |Z|^2)$ derived from the *Mahalanobis depth*, and (*iii*) $W_{PD}(X) = |Z|/(1 + |Z|/MAD(\mathbb{F}_Z))$, where MAD stands for median absolute deviation, derived from the *projection depth*. We omit the technical details. These three weight functions give a center-inward partial ordering, thus essentially quantifying *peripherality* instead of *depth*. Note however, that our results below are of much more general form, and these three special choices of weights only serve as important illustrative examples to achieve desirable robustness and efficiency balance in data analysis.

3. A robust measure of dispersion

From this section onwards, we assume that $\mathcal{X} = \mathbb{R}^p$, that is, the support of the random variable under study is the p-dimensional Euclidean plane, and the data $X_1, \ldots, X_n$ are independent and identically distributed from an elliptical distribution $\mathbb{F}$ with parameters μ and Σ. We also assume that X_1 is absolutely continuous, with $\mathbb{P}[|X_1| = 0] = 0$, and that Σ is positive definite. This eliminates technicalities arising from rank deficient cases, and makes the weight functions affine invariant. Thus, we essentially restrict the rest of this paper to the same framework as in Corollary 2.4. Most of the results below generalize to the case where $\mathcal{X}$ is a separable Hilbert space, however additional technicalities are involved, as in [3], and will be considered in a future project.

3.1. The Weighted Sign Covariance Matrix

Consider the spectral decomposition of Σ given by $\Sigma = \Gamma\Lambda\Gamma^T$, where Γ is an orthogonal matrix and Λ is diagonal with positive diagonal elements $\lambda_1 \geq \ldots \geq \lambda_p$. Also denote the i-th eigenvector of Σ by $\boldsymbol{\gamma}_i = (\gamma_{i,1}, \ldots, \gamma_{i,p})^T$ for $1 \leq i \leq p$. Thus, the i-th column of Γ is $\boldsymbol{\gamma}_i$. In the rest of this paper we use the notation $\Sigma^{-1/2} = \Lambda^{-1/2}\Gamma^T$, and hence $Z = \Lambda^{-1/2}\Gamma^T(X - \mu)$. Recall from Section 2 that we use the notation $\mathbb{F}_Z$ for the distribution of Z, and that $\mathbb{F}_Z$ is a spherically symmetric distribution and hence depends only on $|z|$. Additionally, to simplify notations, for any random variable $X \sim \mathbb{F}$, we occasionally use the abbreviated notation $W(X) \equiv W(X, \mathbb{F})$. Note that $W(X)$ is a *random weight*, and takes the same value as $W(Z, \mathbb{F}_Z)$. Also, as noted in Corollary 2.4, $W(Z)$ is a function of $|Z|$ only. We additionally assume that $\mathbb{E}W^2(X) < \infty$.

It is convenient to write

$$X = \mu + R\Gamma\Lambda^{1/2}U$$

where U is a random variable uniformly distributed on the unit sphere $\mathcal{S}_{0;1} = \{x \in \mathcal{X} : |x| = 1\}$ and R is another random variable independent of U satisfying $\mathbb{E}R^2 = p$. Note that $Z = RU$, and $|Z| = R$, $Z/|Z| = U$. Then we have

$$\begin{aligned} S(X;\mu) = |X - \mu|^{-1}(X - \mu) &= |\Lambda^{1/2}RU|^{-1}R\Gamma\Lambda^{1/2}U \\ &= |\Lambda^{1/2}U|^{-1}\Gamma\Lambda^{1/2}U = |\Lambda^{1/2}Z|^{-1}\Gamma\Lambda^{1/2}Z \end{aligned}$$

The rest of this paper is built on the following transformed random variable:

$$\tilde{X} = W(X, \mathbb{F})S(X;\mu) \equiv W(Z, \mathbb{F}_Z)|\Lambda^{1/2}Z|^{-1}\Gamma\Lambda^{1/2}Z. \tag{6}$$

We fix the notation $\mathbb{S}(X;\mu) = S(X;\mu)S(X;\mu)^T$, and define the following dispersion parameter:

$$\tilde{\Sigma} = \mathbb{E}\tilde{X}\tilde{X}^T = \mathbb{E}W^2(X, \mathbb{F})\mathbb{S}(X;\mu). \tag{7}$$

In the following Theorem, we establish that the eigenvectors of Σ and $\tilde{\Sigma}$ are identical, although their eigenvalues may be different.

Theorem 3.1. *Under the conditions listed above, we have* $\tilde{\Sigma} = \Gamma\tilde{\Lambda}\Gamma^T$*, where* $\tilde{\Lambda} = \Lambda^{1/2}\mathbb{E}W^2(X)\mathbb{E}UU^T/(U^T\Lambda U)\Lambda^{1/2}$ *is a diagonal matrix. Thus, the eigenvectors of* Σ *and* $\tilde{\Sigma}$ *are identical.*

Proof of Theorem 3.1. Fix any index $i \in \{1, \ldots, p\}$. Consider the vector $\tilde{U}$ such that

$$\tilde{U}_j = \begin{cases} U_j & \text{if } j \neq i, \\ -U_i & \text{if } j = i. \end{cases}$$

Then $\tilde{U}$ and U have the same distribution, and note that $U^T\Lambda U = \tilde{U}^T\Lambda\tilde{U}$ almost surely. Consequently, for any $j \neq i$ we have

$$\mathbb{E}\frac{U_iU_j}{U^T\Lambda U} = \mathbb{E}\frac{\tilde{U}_i\tilde{U}_j}{\tilde{U}^T\Lambda\tilde{U}} = -\mathbb{E}\frac{U_iU_j}{U^T\Lambda U}.$$

Consequently, $\mathbb{E}S(X;\mu)S(X;\mu)^T = \Gamma\Lambda_S\Gamma^T$, as established in Theorem 1 of Taskinen et al. [31].

Also, since the weight $W(X)$ is a function of $|Z| = R$, we have that $W(X)$ is independent of $S(X;\mu)$. Consequently, we have

$$\begin{aligned}\tilde{\Sigma} &= \mathbb{E}\tilde{X}\tilde{X}^T = \mathbb{E}W^2(X)S(X;\mu)S(X;\mu)^T \\ &= \mathbb{E}W^2(X)\mathbb{E}S(X;\mu)S(X;\mu)^T = \Gamma\Lambda_W\Gamma^T,\end{aligned}$$

where Λ_W is a diagonal matrix. □

3.2. Sample version of $\tilde{\Sigma}$

We now discuss the properties of the sample version of $\tilde{\Sigma}$, say $\hat{\tilde{\Sigma}}$ computed from $\mathbf{X}$. In practice, we cannot obtain $W(x) \equiv W(x, \mathbb{F})$, and consequently use $W(x, \mathbb{F}_n)$ instead. We assume the following conditions:

(B1) *Bounded weights*: The weights $W(\cdot, \cdot)$ are bounded functions.

(B2) *Uniform convergence*:

$$\sup_{x \in \mathcal{X}} |W(x, \mathbb{F}_n) - W(x, \mathbb{F})| \to 0$$

almost surely as $n \to \infty$.

(B3) *Smoothness under perturbation:* For all $\mathbb{F} \in \mathcal{F}$, there exists a $\delta > 0$, possibly depending on $\mathbb{F}$, such that for any $\epsilon \in (0, \delta)$

$$\sup_{x \in \mathcal{X}} \Big| W\big(x, \mathbb{F}\big) - W\big(x, (1-\epsilon)\mathbb{F} + \epsilon\delta_x\big) \Big| \leq \epsilon.$$

In the above, δ_x denotes point mass at x. These properties are easily satisfied under for weight functions derived from standard depth functions, for example, $W_{HSD}(\cdot)$, $W_{MhD}(\cdot)$ and $W_{PD}(\cdot)$ discussed earlier.

The following result allows us to use the empirical, plug-in weights and an estimated location parameter in the weighted dispersion estimator. A natural choice for the location parameter estimator is the solution to $\sum_{i=1}^{n} \tilde{X}_i = 0$, which is the same as the sample version of the weighted spatial median discussed in Section 2.

Lemma 3.2. *Assume that* $\mathbb{E}\|X - \mu\|^{-4} < \infty$. *Also assume that we have a location estimator* $\hat{\mu}_n$ *satisfying* $\mathbb{E}\|\hat{\mu}_n - \mu\|^4 = O(n^{-2})$. *Then*

$$\frac{1}{n}\sum_{i=1}^{n} W_n^2(X_i, \mathbb{F}_n)\mathbb{S}(X_i; \hat{\mu}_n) = \frac{1}{n}\sum_{i=1}^{n} W^2(X_i, \mathbb{F})\mathbb{S}(X_i; \mu) + R_n,$$

where for any $c \in \mathbb{R}^p$ *such that* $|c| =$, *we have* $\mathbb{E}c^T R_n c = o(n^{-1})$.

Proof of Lemma 3.2. This proof is mostly algebra, and we provide a sketch of the main arguments. We have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} W_n^2(X_i, \mathbb{F}_n)\mathbb{S}(X_i; \hat{\mu}_n) \\
=&\frac{1}{n}\sum_{i=1}^{n} W^2(X_i, \mathbb{F})\mathbb{S}(X_i; \mu) + \frac{1}{n}\sum_{i=1}^{n}\big\{W_n^2(X_i, \mathbb{F}_n) - W^2(X_i, \mathbb{F})\big\}\mathbb{S}(X_i; \mu) \\
&+ \frac{1}{n}\sum_{i=1}^{n} W^2(X_i, \mathbb{F})\big\{\mathbb{S}(X_i; \hat{\mu}_n) - \mathbb{S}(X_i; \mu)\big\} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\big\{W_n^2(X_i, \mathbb{F}_n) - W^2(X_i, \mathbb{F})\big\}\big\{\mathbb{S}(X_i; \hat{\mu}_n) - \mathbb{S}(X_i; \mu)\big\} \\
&= \frac{1}{n}\sum_{i=1}^{n} W^2(X_i, \mathbb{F})\mathbb{S}(X_i; \mu) + T_2 + T_3 + T_4.
\end{aligned}$$

Using the stated technical conditions, we can now show that $\mathbb{E}c^T T_j c = o(n^{-1})$ for $j = 2, 3, 4$. For illutration, we present the case for T_2 below.

Notice that the (j, k)-th element of T_2 is given by

$$n^{-1}\sum_{i=1}^{n} |X_i - \mu|^{-2}\big\{W_n^2(X_i, \mathbb{F}_n) - W^2(X_i, \mathbb{F})\big\}(X_{i,j} - \mu_j)(X_{i,k} - \mu_k),$$

and hence

$$
\begin{aligned}
c^T T_2 c &= \sum_{j,k} c_j c_k T_{2,j,k} \\
&= n^{-1} \sum_{i=1}^{n} |X_i - \mu|^{-2} \big\{ W_n^2(X_i, \mathbb{F}_n) - W^2(X_i, \mathbb{F}) \big\} \big(\sum_j c_j (X_{i,j} - \mu_j) \big)^2 \\
&\leq M n^{-1} \sum_{i=1}^{n} |X_i - \mu|^{-2} \big\{ |W_n(X_i, \mathbb{F}_n) - W(X_i, \mathbb{F})| \big\} \big(c^T (X_i - \mu) \big)^2 \\
&\leq M n^{-1} \sum_{i=1}^{n} |X_i - \mu|^{-2} \big\{ |W_n(X_i, \mathbb{F}_n) - W_n(X_i, \mathbb{F}_{n,-i})| \big\} \big(c^T (X_i - \mu) \big)^2 \\
&\qquad + M n^{-1} \sum_{i=1}^{n} |X_i - \mu|^{-2} \big\{ |W_n(X_i, \mathbb{F}_{n,-i}) - W(X_i, \mathbb{F})| \big\} \big(c^T (X_i - \mu) \big)^2 \\
&= M n^{-1} \sum_{i=1}^{n} T_{21i} + M n^{-1} \sum_{i=1}^{n} T_{22i} \\
&= T_{21} + T_{22}.
\end{aligned}
$$

Let $H(X_i) = |X_i - \mu|^{-2} \big(c^T (X_i - \mu) \big)^2$, and notice that $H(X_i) \leq 1$ almost surely for $|c| = 1$. Now notice that conditional on X_i except for a null set A_i (possibly depending on X_i) we have $T_{21i} \leq n^{-1} H(X_i)$. Thus, except for a null set $A_1 \cap \ldots \cap A_n$, $T_{21} \leq M n^{-2} H(X_i)$ and the conclusion follows for this part.

The argument for T_{22} follows a similar argument. □

Let $vec(\mathbb{S}(X; \mu))$ be the vectorized version of $\mathbb{S}(X; \mu)$. We are now in a position to state the result for consistency of the sample version of $\tilde{\Sigma}$,

Theorem 3.3. *Assume the conditions of Lemma 3.2. Then*

$$
\begin{aligned}
n^{1/2} \sum_{i=1}^{n} \Big(& W_n^2(X_i, \mathbb{F}_n) vec(\mathbb{S}(X_i; \hat{\mu}_n)) - \mathbb{E} W^2(X_i) vec(\mathbb{S}(X_i; \mu)) \Big) \\
& \stackrel{\mathcal{D}}{\Rightarrow} N_{p^2} \big(0, V_W \big),
\end{aligned}
$$

where $V_W = \mathbb{V}[W^2(X, \mathbb{F}) vec(\mathbb{S}(X; \mu))]$.

The asymptotic normality follows from our assumptions and as a direct consequence of Lemma 3.2. Incidentally, an expression for V_W can be explicitly obtained in terms of Γ, Λ and $\mathbb{F}$, but is algebraic in nature. We present it in the supplementary material.

We now use Theorem 3.3 to obtain consistency results for the eigenvectors obtained from $\hat{\tilde{\Sigma}} = n^{-1} \sum_{i=1}^{n} W^2(X_i, \mathbb{F}) \mathbb{S}(X_i; \hat{\mu}_n)$. Suppose that $\tilde{\Lambda}_1 > \tilde{\Lambda}_2 > \ldots > \tilde{\Lambda}_p$ are the eigenvalues of $\tilde{\Sigma}$, which we assume are all distinct values.

Theorem 3.4. *Suppose the spectral decomposition of* $\hat{\tilde{\Sigma}}$ *is given by* $\hat{\tilde{\Sigma}} = \hat{\Gamma} \hat{\tilde{\Lambda}} \hat{\Gamma}^T$. *Then the matrix of centered and scaled eigenvectors* $G_n = n^{1/2} (\hat{\Gamma} - \Gamma)$ *and the*

vector of centered and scaled eigenvalues $L_n = n^{1/2}(\widehat{\tilde{\Lambda}} - \tilde{\Lambda})$ *have independent distributions. The distribution of the random variable vec* (G_n) *converges weakly to a* p^2*-variate normal distribution with mean* $\mathbf{0}_{p^2}$ *and the variance matrix whose* (i,j)*-th block of* $p \times p$ *entries are given by*

$$\sum_{k=1,k\neq i}^{p} \Big[\tilde{\Lambda}_i - \tilde{\Lambda}_k\Big]^{-2} \mathbb{E}\Big[W^4(Z,\mathbb{F}_Z)\big(\mathbb{S}_{i,k}(\Lambda^{1/2}Z;\mathbf{0})\big)^2\Big]\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T, \text{ if } i=j, \tag{8}$$

$$-\Big[\tilde{\Lambda}_i - \tilde{\Lambda}_j\Big]^{-2} \mathbb{E}\Big[W^4(Z,\mathbb{F}_Z)\big(\mathbb{S}_{i,j}(\Lambda^{1/2}Z;\mathbf{0})\big)^2\Big]\boldsymbol{\gamma}_i\boldsymbol{\gamma}_j^T; \text{ if } i\neq j. \tag{9}$$

The distribution of L_n *converges weakly to a* p*-dimensional normal distribution with mean* $\mathbf{0}_p$ *and the variance-covariance matrix whose* (i,j)*-the element is*

$$\mathbb{E}\Big[W^4(Z,\mathbb{F}_Z)\big(\mathbb{S}_{i,i}(\Lambda^{1/2}Z;\mathbf{0})\big)^2\Big] - \tilde{\Lambda}_i^2, \text{ if } i=j,$$
$$\mathbb{E}\Big[W^4(Z,\mathbb{F}_Z)\big(\mathbb{S}_{i,j}(\Lambda^{1/2}Z;\mathbf{0})\big)^2\Big] - \tilde{\Lambda}_i\tilde{\Lambda}_j, \text{ if } i\neq j.$$

The proof of this result follows from using Theorem 3.3 and using techniques similar to a corresponding result in Taskinen et al. [31]. We omit the algebraic details here and put it in the supplementary material.

Recall that the asymptotic variance of the i-th eigenvector of the *sample covariance matrix*, say $\hat{\boldsymbol{\gamma}}_i$ is [2]:

$$A\mathbb{V}(\sqrt{n}\hat{\boldsymbol{\gamma}}_i) = \sum_{k=1;k\neq i}^{p} \frac{\lambda_i\lambda_k}{(\lambda_i-\lambda_k)^2}\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T; \quad 1\leq i\leq p. \tag{10}$$

Suppose $\widehat{\tilde{\boldsymbol{\gamma}}}_i$ is the i-th eigenvector of $\widehat{\tilde{\Sigma}}$, whose asymptotic behavior is presented above in Theorem 3.4.

This leads to the following useful result:

Corollary 3.5. *The asymptotic relative efficiency of* $\widehat{\tilde{\boldsymbol{\gamma}}}_i$*, relative to* $\hat{\boldsymbol{\gamma}}_i$*, is given by*

$$ARE(\widehat{\tilde{\boldsymbol{\gamma}}}_i, \hat{\boldsymbol{\gamma}}_i; \mathbb{F})$$
$$= \Big[\sum_{k=1;k\neq i}^{p} \frac{\lambda_i\lambda_k}{(\lambda_i-\lambda_k)^2}\Big]\Big[\sum_{k=1,k\neq i}^{p}\Big[\tilde{\Lambda}_i - \tilde{\Lambda}_k\Big]^{-2}\mathbb{E}\Big[W^4(Z,\mathbb{F}_Z)\big(\mathbb{S}_{i,k}(\Lambda^{1/2}Z;\mathbf{0})\big)^2\Big]\Big]^{-1}.$$

The proof of this Corollary is immediate.

4. An affine equivariant robust measure of dispersion

A desirable invariance property of any dispersion parameter T_X corresponding to a random variable X is that under affine transformation $Y = AX + b$ the dispersion parameter scales to $T_Y = AT_XA^T$. It is clear that $\tilde{\Sigma}$ does not possess

this property, since it remains unchanged for X and $Y = cX$ for any scalar $c > 0$.

We follow the general framework of M-estimation with data-dependent weights [16] to construct an affine equivariant counterpart of the $\tilde{\Sigma}$. Specifically, we implicitly define

$$\Sigma_* = \frac{p}{\mathbb{V}W(X)}\mathbb{E}\left[\frac{W^2(X)(X-\mu)(X-\mu)^T}{(X-\mu)^T\Sigma_*^{-1}(X-\mu)}\right]. \tag{11}$$

To ensure existence and uniqueness of Σ_*, consider the class of dispersion parameters Σ_M that are obtained as solutions of the following equation:

$$\mathbb{E}\left[u(|Z_M|)\frac{Z_M Z_M^T}{|Z_M|^2} - v(|Z_M|)\mathbb{I}_p\right] = 0 \tag{12}$$

with $Z_M = \Sigma_M^{-1/2}(X-\mu)$. Under the following assumptions on the scalar valued functions u and v, the above equation produces a unique solution [16]:

(C1) The function $u(r)/r^2$ is monotone decreasing, and $u(r) > 0$ for $r > 0$;
(C2) The function $v(r)$ is monotone decreasing, and $v(r) > 0$ for $r > 0$;
(C3) Both $u(r)$ and $v(r)$ are bounded and continuous,
(C4) $u(0)/v(0) < p$,
(C5) For any hyperplane in the sample space $\mathcal{X}$, (i) $P(H) = \mathbb{E}\{\mathcal{I}_{\{X\in H\}}\} < 1 - pv(\infty)/u(\infty)$ and (ii) $P(H) \leq 1/p$.

Putting things into context, in our case we have $v(\cdot) = p^{-1}\mathbb{V}W(X)$, $u(\cdot) = W^2(X)$. We proceed to verify the other conditions for the weight functions $W_{HSD}(\cdot)$, $W_{MhD}(\cdot)$ and $W_{PD}(\cdot)$ discussed earlier.

It is easy to verify that the resulting $u(\cdot)$ from the above choices satisfy (C1) and (C3). Note that $v(\cdot)$ is a finite positive constant, and (C2) and (C3) are also easily satisfied. Since $u(0) = 0$ in all the above cases, (C4) is also easy to check. Since X is absolutely continuous, (C5) holds trivially.

To compute the sample version of Σ_*, we solve (11) iteratively by obtaining a sequence of positive definite matrices $\hat{\Sigma}_*^{(k)}$ until convergence. Thus, using the location estimator $\hat{\mu}_n$, we may iterate

$$\hat{\Sigma}_*^{(k+1)} = \frac{p}{\mathbb{V}W(X)}\mathbb{E}\left[\frac{W^2(X)(X-\hat{\mu}_n)(X-\hat{\mu}_n)^T}{(X-\hat{\mu}_n)^T(\hat{\Sigma}_*^{(k)})^{-1}(X-\hat{\mu}_n)}\right].$$

The asymptotic properties of $\hat{\Sigma}_*$ can be obtained using methods similar to those of Section 3, and techniques presented in [10] and elsewhere. We state the following result and omit its proof.

Theorem 4.1. *The asymptotic covariance matrix of an eigenvector of the sample affine equivariant scatter functional $\hat{\Sigma}_*$ is given by*

$$V_{12}\sum_{k=1,k\neq i}^{p}\frac{\lambda_i\lambda_k}{\lambda_i-\lambda_k}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_k^T,$$

where V_{12} is the asymptotic variance of an off-diagonal element of $\hat{\Sigma}_*$ when the underlying distribution is $\mathbb{F}_Z$. It follows that if $\hat{\boldsymbol{\gamma}}_{*,i}$ is the i-th eigenvector of $\hat{\Sigma}_*$,

$$ARE(\hat{\boldsymbol{\gamma}}_{*,i}, \hat{\boldsymbol{\gamma}}_i; \mathbb{F}) = V_{12}^{-1} = \frac{[\mathbb{E}(pu(|Z|) + u'(|Z|)|Z|)]^2}{p^2(p+2)^2\mathbb{E}(u(|Z|)^2\mathbb{E}(\mathbb{S}_{12}(Z;\mathbf{0}))^2}. \tag{13}$$

5. Robust estimation of eigenvalues, and a plug-in estimator of $\boldsymbol{\Sigma}$

As seen in Theorem 3.1, eigenvalues of the $\tilde{\Sigma}$ are not same as the population eigenvalues. In this section, we discuss on robust estimation of λ_i's using $\tilde{\Sigma}$. Assume the data is centered, the robust estimator from Section 2 suffices. We start by computing the sample version $\widehat{\tilde{\Sigma}}$ and its spectral decomposition:

$$\widehat{\tilde{\Sigma}} = \widehat{\Gamma}\widehat{\tilde{\Lambda}}\widehat{\Gamma}^T$$

We then use the following steps:

1. Randomly divide the sample indices $\{1, 2, \ldots, n\}$ into k disjoint groups $\{G_1, \ldots, G_k\}$ of size $\lfloor n/k \rfloor$ each.
2. Transform the data matrix: $\mathbf{S} = \widehat{\Gamma}^T\mathbf{X}$.
3. Calculate coordinate-wise variances for each group of indices G_j:

$$\lambda^\dagger_{i,j} = \frac{1}{|G_j|}\sum_{l\in G_j}\left(S_{li} - \bar{S}_{G_j,i}\right)^2; \quad i = 1, \ldots, p;\ j = 1, \ldots, k.$$

   where $\bar{\mathbf{S}}_{G_j} = (\bar{S}_{G_j,1}, \ldots, \bar{S}_{G_j,p})^T$ is the vector of column-wise means of $\mathbf{S}_{G_j}$, the submatrix of $\mathbf{S}$ with row indices in G_j.
4. Obtain estimates of eigenvalues by taking coordinate-wise medians of these variances:

$$\lambda^\dagger_i = \text{median}(\lambda^\dagger_{i,1}, \ldots, \lambda^\dagger_{i,k}); \quad i = 1, \ldots, p.$$

We collect $\lambda^\dagger_i$, $i = 1, \ldots, p$ in the diagonal matrix $\Lambda^\dagger = \text{diag}(\lambda^\dagger_1, \ldots, \lambda^\dagger_p)$. The number of subgroups used to calculate this median-of-small-variances estimator can be determined following [24]. There can be other ways of estimating the eigenvalues of Σ using $\mathbf{S}$ also, we will pursue such methods elsewhere. We construct a consistent plug-in estimator of the population covariance matrix Σ as

$$\Sigma^\dagger = \widehat{\Gamma}\Lambda^\dagger\widehat{\Gamma}^T.$$

Let $|A|_F$ denote the Frobenius norm of a matrix A, in other words, $|A|_F = (\text{trace}A^TA)^{1/2}$. The following result establishes that this is a consistent estimator of Σ:

Theorem 5.1. *Suppose that as $n \to \infty$, $k \to \infty$ and $n/k \to \infty$. Then we have*

$$\|\Sigma^\dagger - \Sigma\|_F \xrightarrow{P} 0.$$

Proof of Theorem 5.1. This proof has many algebraic steps, and we sketch the main arguments below.

Suppose $\hat{A} = \widehat{\Gamma}^T \Sigma \widehat{\Gamma}$. Owing to the fact that the Frobenius norm is invariant under rotations and that p is finite and fixed, it suffices to show that the off-diagonal elements of $\hat{A}$ converge in probability to zero, and that the difference between the i-th diagonal element of $\hat{A}$ and $\lambda_i^\dagger$ converges to zero for any $i = 1, \ldots, p$.

Now notice that from Theorem 3.4 we have that $\widehat{\Gamma} = \Gamma + R_{n1}$, where the (i,j)-th element of the remainder $R_{n1,i,j}$ satisfies $\mathbb{E}R_{n1,i,j}^2 = O(n^{-1})$. We can show, using standard algebra, that

$$\hat{A} = \Lambda + R_{n2},$$

where the (i,j)-th element of the remainder $R_{n2,i,j}$ satisfies $\mathbb{E}R_{n2,i,j}^2 = O(n^{-1})$. This follows immediately from above, the fact that p is finite and fixed, and all elements of Λ are constants. This immediately establishes the case for the off-diagonal elements.

For the diagonal elements, notice that since $k \to \infty$, each coordinate-wise variance $\lambda_{i,j}^\dagger$ for each group of indices G_j is a consistent estimator of λ_i. The result follows.

□

6. Influence Functions of Dispersion Measures

We retain the framework adopted in Section 3, and discuss in this section the robustness and efficiency properties associated with $\tilde{\Sigma}$ and Σ_*, and principal components derived therefrom. We do not discuss $\Sigma^\dagger$ here, since the properties of that approach follow from those of $\tilde{\Sigma}$. We additionally assume that the eigenvalues of Σ are distinct, and given by $\lambda_1 > \lambda_2 > \ldots > \lambda_p$, to avoid several additional technical conditions for the theoretical results to follow. The case where the eigenvalues of Σ can have multiplicity greater than one requires no additional conceptual development, but does require considerable algebraic manipulations.

For studying the robustness aspect, we first present some results relating to influence functions in the current context. Influence functions quantify how much influence a sample point, especially an infinitesimal contamination, has on any functional of a probability distribution [15]. Given any probability distribution $\mathbb{H} \in \mathcal{M}$, the influence function of any point $x_0 \in \mathcal{X}$ for some functional $T(\mathbb{H})$ on the distribution is defined as

$$IF(x_0; T, \mathbb{H}) = \lim_{\epsilon \to 0} \frac{1}{\epsilon}(T(\mathbb{H}_\epsilon) - T(\mathbb{H})),$$

where $\mathbb{H}_\epsilon = (1-\epsilon)\mathbb{H} + \epsilon\delta_{x_0}$; δ_{x_0} being the distribution with point mass at x_0. When $T(\mathbb{H}) = E_{\mathbb{H}} f$ for some $\mathbb{H}$-integrable function f, $IF(x_0; T, \mathbb{H}) = f(x_0) - T(\mathbb{H})$.

It now follows that

$$IF(x_0; \tilde{\Sigma}, \mathbb{F}) = W^2(x_0)\mathbb{S}(x_0; \mu) - \tilde{\Sigma}.$$

Recall that $\tilde{\lambda}_1 > \tilde{\lambda}_2 > \ldots > \tilde{\lambda}_p$ are the eigenvalues of $\tilde{\Sigma}$, which we assume are all distinct values.

Proposition 6.1. *The influence function of $\boldsymbol{\gamma}_i$ as follows:*

$$\begin{aligned} IF(x_0; \boldsymbol{\gamma}_i, \mathbb{F}) &= \sum_{k=1;k\neq i}^{p} \frac{1}{\tilde{\lambda}_i - \tilde{\lambda}_k} \left\{ \boldsymbol{\gamma}_k^T IF(x_0; \tilde{\Sigma}, \mathbb{F}) \boldsymbol{\gamma}_i \right\} \boldsymbol{\gamma}_k \\ &= \sum_{k=1;k\neq i}^{p} \frac{1}{\tilde{\lambda}_i - \tilde{\lambda}_k} W^2(x_0) \left\{ \boldsymbol{\gamma}_k^T \mathbb{S}(x_0; \mu) \boldsymbol{\gamma}_i \right\} \boldsymbol{\gamma}_k. \end{aligned}$$

The proof of Proposition 6.1 follows from [29] and [10], we omit the details.

If the weight function $W(\cdot)$ is a bounded function, as is the case of W_{HSD}, W_{MhD}, and W_{PD}, the influence function given in Proposition 6.1 is bounded, indicating good robustness properties of the principal component analysis.

We now derive the influence function for Σ_*.

Proposition 6.2. *The influence function of Σ_* is given by*

$$IF(x_0, \Sigma_*, \mathbb{F}) = \alpha_{\Sigma_*}(|x_0|; \mathbb{F}_Z)\mathbb{S}(x_0; \mu) - \beta_{\Sigma_*}(|x_0|; \mathbb{F}_Z)\Sigma_*.$$

Proof of Proposition 6.2. Let $z_0 = \Lambda^{-1/2}\Gamma^T(x_0 - \mu) = (z_{01}, \ldots, z_{0p})^T$. As a first step, since Σ_* is affine equivariant, we obtain from [10] that

$$IF(x_0, \Sigma_*, \mathbb{F}) = \Sigma_*^{1/2} IF(z_0, \Sigma_*, \mathbb{F}_Z)\Sigma_*^{1/2}.$$

From Lemma 1 of [15], page 276, we obtain that there exist scalar valued functions $\alpha_{\Sigma_*}(|x_0|; \mathbb{F}_Z)$ and $\beta_{\Sigma_*}(|x_0|; \mathbb{F}_Z)$ such that

$$IF(z_0, \Sigma_*, \mathbb{F}_Z) = \alpha_{\Sigma_*}(|x_0|; \mathbb{F}_Z)\mathbb{S}(z_0; \mathbf{0}) - \beta_{\Sigma_*}(|x_0|; \mathbb{F}_Z)\mathbb{I}_p,$$

consequently we obtain

$$IF(x_0, \Sigma_*, \mathbb{F}) = \alpha_{\Sigma_*}(|x_0|; \mathbb{F}_Z)\mathbb{S}(x_0; \mu) - \beta_{\Sigma_*}(|x_0|; \mathbb{F}_Z)\Sigma_*.$$

□

Suppose $\lambda_{*1} > \lambda_{*2} > \ldots > \lambda_{*p}$ are the eigenvalues of Σ_*, which we assume are all distinct values. Also denote the i-th eigenvector of Σ_* by $\boldsymbol{\gamma}_{*i} = (\gamma_{*i1}, \ldots, \gamma_{*ip})^T$ for $1 \leq i \leq p$.

Proposition 6.3. *The influence function of $\boldsymbol{\gamma}_{*i}$ may be obtained as*

$$\begin{aligned} IF(x_0; \boldsymbol{\gamma}_{*i}, \mathbb{F}) &= \sum_{k=1;k\neq i}^{p} \frac{1}{\lambda_{*i} - \lambda_{*k}} \left\{ \boldsymbol{\gamma}_{*k}^T IF(x_0; \tilde{\Sigma}, \mathbb{F}) \boldsymbol{\gamma}_{*i} \right\} \boldsymbol{\gamma}_{*k} \\ &= \alpha_{\Sigma_*}(|x_0|; \mathbb{F}_Z) \sum_{k=1;k\neq i}^{p} \frac{1}{\lambda_{*i} - \lambda_{*k}} \left\{ \boldsymbol{\gamma}_{*k}^T \mathbb{S}(x_0; \mu) \boldsymbol{\gamma}_{*i} \right\} \boldsymbol{\gamma}_{*k}. \end{aligned}$$

We omit the proof of Proposition 6.3, which follows along similar lines to to rest of the computations of this section. It can be shown that when $W(\cdot)$ is a bounded function, $\alpha_{\Sigma_*}(|x_0|;\mathbb{F}_Z)$ is also bounded, along the lines of [16], which in turn implies that the influence function for a principal component based on Σ_* is also bounded.

7. Simulation Studies

We report a number of numerical simulation studies on several properties relating to $\tilde{\Sigma}$ and Σ_*, and their eigenvalues and eigenvectors, on datasets with or without influential points, to illustrate the finite sample efficiency and robustness properties of the proposed weighted estimators. We compare these proposed estimators with techniques that exists in literature, specifically, the Sign Covariance Matrix (SCM) and Tyler's shape matrix [32].

7.1. Efficiency of different robust estimators

We compare the performance of $\tilde{\Sigma}$ and Σ_* with that of the SCM and Tyler's scatter matrix. For this study, we fix the dimension $p = 4$. We consider six elliptical distributions, and from every distribution draw 10000 samples each for sample sizes $n = 20, 50, 100, 300$ and 500. All distributions are centered at $\mathbf{0}_p$, and have covariance matrix $\Sigma = \text{diag}(4, 3, 2, 1)$.

We use the concept of principal angles [23] to find out error estimates for the first eigenvector of a scatter matrix. In our case, the first eigenvector is

$$\boldsymbol{\gamma}_1 = (1, \overbrace{0, \ldots, 0}^{p-1})^T.$$

We measure the prediction error for an eigenvector estimator (say, $\tilde{\boldsymbol{\gamma}}_1$), using the smallest angle between the true and predicted vectors, i.e. $\cos^{-1}|\tilde{\boldsymbol{\gamma}}_1^T\hat{\boldsymbol{\gamma}}_1|$. A small absolute value of this angle means to better prediction. We repeat this 10,000 times and calculate the **Mean Squared Prediction Angle**:

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

where $\tilde{\boldsymbol{\gamma}}_1^{(m)}$ is the value of $\tilde{\boldsymbol{\gamma}}_1$ in the m-th replication, $m = 1, \ldots, 10,000$. The finite sample efficiency of $\tilde{\boldsymbol{\gamma}}_1$ relative to that from the sample covariance matrix, i.e. $\hat{\boldsymbol{\gamma}}_1$ is obtained as:

$$FSE(\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_1) = \frac{MSPA(\hat{\boldsymbol{\gamma}}_{0,1})}{MSPA(\hat{\boldsymbol{\gamma}}_1)}.$$

The results from this simulation exercise are presented in Table 1. It can be seem that $\tilde{\Sigma}$-based estimators (columns 3-5) outperform SCM and Tyler's M-estimator of scatter. Among the 3 depth functions considered, projection depth

4-variate t_5	SCM	Tyler	$\tilde{\Sigma}$-H	$\tilde{\Sigma}$-M	$\tilde{\Sigma}$-P	Σ_*-H	Σ_*-M	Σ_*-P
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	$\tilde{\Sigma}$-H	$\tilde{\Sigma}$-M	$\tilde{\Sigma}$-P	Σ_*-H	Σ_*-M	Σ_*-P
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	$\tilde{\Sigma}$-H	$\tilde{\Sigma}$-M	$\tilde{\Sigma}$-P	Σ_*-H	Σ_*-M	Σ_*-P
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	$\tilde{\Sigma}$-H	$\tilde{\Sigma}$-M	$\tilde{\Sigma}$-P	Σ_*-H	Σ_*-M	Σ_*-P
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	$\tilde{\Sigma}$-H	$\tilde{\Sigma}$-M	$\tilde{\Sigma}$-P	Σ_*-H	Σ_*-M	Σ_*-P
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	$\tilde{\Sigma}$-H	$\tilde{\Sigma}$-M	$\tilde{\Sigma}$-P	Σ_*-H	Σ_*-M	Σ_*-P
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 1

Finite sample efficiencies of estimators of the first eigenvector based on several scatter matrices in dimension $p = 4$. The notation H, M or P after $\tilde{\Sigma}$ or Σ_* indicates the depth function used for the weights: H = halfspace depth, M = Mahalanobis depth, P = projection depth.

gives 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 that are much more computationally intensive (see Table 4 in [30]). The affine equivariant Σ_*-based estimators (columns 6-8) are even more efficient.

7.2. Influence function comparison

In Figure 2 we consider first eigenvectors of $\tilde{\Sigma}$, the Sign Covariance Matrix (SCM) and Tyler's shape matrix [32]. We generate data from and set $\mathbb{F} \equiv \mathcal{N}_2(0, \mathrm{diag}(2, 1))$ and plot norms of the eigenvector influence functions for different values of x_0. Let us denote the i-th eigenvector of the Sign Covariance Matrix and Tyler's shape matrix by $\gamma_{S,i}$ and $\gamma_{T,i}$, respectively. Their influence

functions are given as follows:

$$IF(x_0;\gamma_{S,i},\mathbb{F}) = \sum_{k=1;k\neq i}^{p} \frac{\mathbb{S}_{ik}(\Lambda^{1/2}z_0,0)}{\lambda_{S,i}-\lambda_{S,k}}\gamma_k;$$
$$\text{where } \lambda_{S,i} = \mathbb{E}_Z\left(\frac{\lambda_i z_i^2}{\sum_{j=1}^{p}\lambda_j z_j^2}\right),$$
$$IF(x_0;\gamma_{T,i},\mathbb{F}) = (p+2)\sum_{k=1;k\neq i}^{p} \frac{\sqrt{\lambda_i\lambda_k}}{\lambda_i-\lambda_k}\mathbb{S}_{ik}(z_0;0)\gamma_k.$$

Panels (b) and (c) in Figure 2, corresponding to Sign Covariance Matrix and Tyler's shape matrix respectively, exhibit an 'inlier effect', that is, points close to the center having high influence, which results in loss of efficiency. On the other hand, the influence function for eigenvector estimates of the sample covariance matrix (panel (a)) is unbounded and makes the corresponding estimates non-robust. In comparison, the $\tilde{\Sigma}$ corresponding to weights derived from projection depth, half-space depth and Mahalanobis depth have bounded influence functions *and* small values of the influence function at 'deep' points.

7.3. Efficiency of affine equivariant robust estimator

We study the finite sample efficiency properties of Σ_* using a simulation exercise. We consider 6 different elliptic distributions, namely, the p-variate multivariate Normal distribution and the multivariate t distributions corresponding to degrees of freedom 5, 6, 10, 15 and 25. We compute the ARE of the estimator for the first eigenvector using Σ_*, using weights based on the projection depth (PD) and the halfspace depth (HSD), thus this simulation is an illustration of how different choices of weights affect the results in the context of Theorem 4.1. We consider using the sample covariance matrix as the baseline method for this study. The ARE values are computed by using Monte-Carlo simulation of 10^6 samples and subsequent numerical integration. We report the results of this exercise in Table 2. Based on these results, we notice that Σ_* is 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, note that Σ_* is almost as efficient as the sample covarnace matrix even when the data comes from multivariate normal distribution.

7.4. Robust sufficient dimension reduction and supervised learning

One of the main usages of obtaining dispersion estimators and their eigenvalues and eigenvectors is in *dimension reduction* techniques. Examples of such uses are in *principal component regression, partial least squares* and *envelope methods*. We illustrate below the latter technique, in the context of *sufficient dimension reduction* (SDR). For details on envelope methods and other uses of robust estimators of dispersion and eigen-structures, see Cook et al. [8], Adragni and Cook

Distribution	PD				HSD			
	$p=2$	$p=5$	$p=10$	$p=20$	$p=2$	$p=5$	$p=10$	$p=20$
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 2

Table of AREs of the estimator for the first eigenvector estimation using Σ_*, relative to using the sample covariance matrix, for different choices of dimension p. The data-generating distributions are the multivariate Normal (MVN), and multivariate t-distributions with degrees of freedom 5, 6, 10, 15 and 25. Weights for Σ_* are based on either the projection depth (PD) or the half-space depth (HSD).

[1], Cook and Zhang [9] and references and citations of these articles. In the context of multivariate-response ($Y_i \in \mathbb{R}^q$) linear regression, the envelope method proposes the model $Y_i = \alpha + \Gamma_1 \eta x_i + e_i$, where e_i are independent mean zero Gaussian noise terms with covariance matrix Σ whose spectral representation can be written as

$$\begin{aligned}\Sigma = \Gamma\Lambda\Gamma^T &= \begin{pmatrix} \Gamma_0 & \Gamma_1 \end{pmatrix} \begin{pmatrix} \Lambda_0 & 0 \\ 0 & \Lambda_1 \end{pmatrix} \begin{pmatrix} \Gamma_0 \\ \Gamma_1 \end{pmatrix} \\ &= \Gamma_0\Lambda_0\Gamma_0^T + \Gamma_1\Lambda_1\Gamma_1^T.\end{aligned}$$

Thus, the eigenvectors of Σ are partitioned into two blocks: $\Gamma_1 \in \mathbb{R}^q \times \mathbb{R}^d$ and $\Gamma_0 \in \mathbb{R}^q \times \mathbb{R}^{q-d}$, and the regression coefficient of Y_i on x_i is given by $\Gamma_1\eta$ for some $\eta \in \mathbb{R}^d \times \mathbb{R}^p$. Dimension reduction is achieved when $d \ll p$, typically without extraneous assumptions like sparsity. The envelope model for generalized linear models is discussed in Adragni and Cook [1], and may be used for supervised learning. Nonlinear regression models may also be handled similarly.

Given a set of examples $\{(Y_i, X_i), i = 1, \ldots, n\}$, an envelope-based prediction for the response Y for any X may be obtained from

$$\begin{aligned}\hat{Y}(X) &= \Big[\sum_{i=1}^n w_i\Big]^{-1} \sum_{i=1}^n w_i Y_i, \text{ where} \\ w_i &= \exp\left[-\frac{1}{\hat{\sigma}^2}|\hat{\Gamma}_1^T(X - X_i)|^2\right].\end{aligned}$$

The above assumes that the covariates come from the Gaussian distribution $N_p(\mathbf{0}_p, \sigma^2\mathbb{I}_p)$, and appropriate changes may be made for other distributions.

We design a robust version of the above, by using weighted spatial medians for location parameters corresponding to the distributions of X and $X|Y$, and using the first d eigenvectors of $\tilde{\Sigma}$ as $\hat{\Gamma}_1$. A robust location estimator for the distribution of $X|Y$ is required for the estimation of σ^2. Details are available in [1].

Ina non-linear regression model, we compare the performance of the robust version of SDR with the original method of [1] with or without the presence

of bad leverage points in Σ. For any given choice of covariate dimension p, we take $n = 200$ and $d = 1$, and generate the responses $Y_1, \ldots, Y_n$ as independent standard normal, and $X|Y$ as Normal with mean $Y + Y^2 + Y^3$ in each of the p coordinates, and variance $25\mathbb{I}_p$. We measure performances of the SDR models by their mean squared prediction error on another set of 200 observations generated similarly, and taking the average of these errors on 100 such training-test pairs of datasets. The above steps are repeated for the choices of $p = 5, 10, 25, 50, 75, 100, 125, 150$.

Panel (a) of figure 3 compares prediction errors using both robust and maximum likelihood SDR estimates when the covariates contain no outliers: here the two methods are virtually indistinguishable. We then introduce outliers in each of the 100 datasets by adding 100 to first $p/5$ coordinates of the first 10 observed covariate values, and repeat the analysis. Panel (b) of the figure shows that the robust SDR method remains more accurate in predicting out of sample observations for all values of p than the standard SDR.

8. Real Data Applications

We now present an application of our proposed approach to some real data problems.

Robust techniques are useful when in identifying outlying observations, and we illustrate below how to make use of our fixed-dimensional methods presented earlier for functional (and hence infinite-dimensional) data.

We follow the approach of [4] for performing robust principal component analysis on functional data using the estimated eigenvectors from $\widehat{\widehat{\Sigma}}$. Suppose the data consistents of n curves, say $\mathcal{F} = \{f_1, \ldots, f_n\} \in L^2[0, 1]$, each observed at a set of common design points $\{t_1, ..., t_m\}$. We model each of these functions as a linear combination of p mutually orthogonal B-spline basis functions $\mathcal{D} = \{\delta_1, ..., \delta_p\}$. We map data for each of the functions onto the coordinate system formed by the spline basis:

$$T(\mathcal{F}, \mathcal{D})_{ij} = \sum_{l=2}^{m} f_i(t_l)\delta_j(t_l)(t_l - t_{l-1}); \quad 1 \le i \le n, 1 \le j \le p. \tag{14}$$

We then model the i-th row of the $n \times p$ matrix $T(\mathcal{F}, \mathcal{D}) \equiv T$, denoted by $\mathbf{T}_i$ as follows:

$$\mathbf{T}_i = \mu + Ps_i + e_i,$$

where μ is a location parameter, P is a $p \times q$ loading matrix, s_i is a $q \times 1$ score vector, and e_i is the error term. We obtain robust estimators of μ, P and consequently s_i using $\widehat{\widehat{\Sigma}}$. Define $\widehat{\mathbf{T}}_i = \hat{\mu} + \hat{P}\hat{s}_i$. The *orthogonal distance* (OD) corresponding to this projection is defined as

$$OD_i = |\mathbf{T}_i - \widehat{\mathbf{T}}_i|.$$

Analogously, the *score distance* (SD) is defined as

$$SD_i = \sqrt{\sum_{j=1}^{q} \frac{\hat{s}_{ij}^2}{\hat{\lambda}_j}};$$

where $\hat{\lambda}_1, \ldots, \hat{\lambda}_q$ are the top eigenvalues from $\widehat{\widetilde{\Sigma}}$. For outlier detection, following [17] we set the upper cutoff values for score distances at $(\chi^2_{2,.975})^{1/2}$ and orthogonal distances at

$$[\text{median}(OD^{2/3}) + \text{MAD}(OD^{2/3})\Phi^{-1}(0.975)]^{3/2},$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

We apply the above outlier detection method on two data sets. First, we consider the monthly average *sea surface temperature anomaly data* from June 1970 to May 2004, available from http://www.cpc.ncep.noaa.gov/data, depicted in panel (a) of Figure 4)). Second, we consider the *octane data*, which consists of 226 variables and 39 observations [11]. 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. This data is presented in panel (b) of Figure 4)).

In the sea surface temperature data, using a cubic spline basis with knots at alternate months starting in June, we get a close approximation as depicted in panel (c) of Figure 4. Using our proposed methodology with $q = 1$ results in two points having their SD and OD larger than cutoff, depicted in panel (e) of Figure 4. 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 c, and pinpoint the two seasons with strongest El-Niño events.

On the octane data, we use the same methodology, and again the top robust PC turns out to be sufficient in identifying all 6 outliers. Details are available in panels (b), (d) and (f) of Figure 4.

9. Conclusions

We propose the use of a weighted multivariate sign transformation for robust estimation and inference, and as demonstrated by theoretical results and several simulation studies and data examples, in many situations using a data-depth driven weight function leads to considerable efficiency gain without compromising robustness properties. Our methodology seems to suggest new ways of identifying El-Niño or La-Niña events from the sea-surface temperature anomaly data, which will be studied further later.

Several of our results stated above are for data from the Euclidean space $\mathbb{R}^p$, where p is fixed. The cases where p increases with sample size and may be higher than sample size, and where data are from a separable Hilbert space, will be considered in a future work. There are few conceptual difficulties to such extensions, however, there are several technical and algebraic challenges.

Acknowledgements

The research of SC is partially supported by the National Science Foundation (NSF) under grants # DMS-1622483, # DMS-1737918.

References

[1] Adragni, K. P. and Cook, R. D. (2009). Sufficient dimension reduction and prediction in regression. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367(1906):4385 – 4405.

[2] Anderson, T. (2009). *An Introduction to Multivariate Statistical Analysis*. Wiley, Hoboken, NJ, 3 edition.

[3] Bali, J. L., Boente, G., Tyler, D. E., and Wang, J.-L. (2011). Robust functional principal components: A projection-pursuit approach. *The Annals of Statistics*, 39(6):2852 – 2882.

[4] Boente, G. and Salibian-Barrera, M. (2015). *s*-estimators for functional principal component analysis. *Journal of the American Statistical Association*, 110(511):1100 – 1111.

[5] Cardot, H., Cénac, P., and Godichon-Baggioni, A. (2017). Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls. *The Annals of Statistics*, 45(2):591 – 614.

[6] Chakraborty, A. and Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *The Annals of Statistics*, 42(3):1203 – 1231.

[7] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association*, 91(434):862 – 872.

[8] Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*, pages 927 – 960.

[9] Cook, R. D. and Zhang, X. (2015). Foundations for envelope models and methods. *Journal of the American Statistical Association*, 110(510):599 – 611.

[10] 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(3):603–618.

[11] Esbensen, K. H., Schönkopf, S., and Midtgaard, T. (1994). *Multivariate Analysis in Practice*. CAMO As, Trondheim, Germany.

[12] Fang, K.-T., Kotz, S., and Ng, K.-W. (1990). *Symmetric Multivariate and Related Distributions*. CRC Press.

[13] Haberman, S. J. (1989). Concavity and estimation. *The Annals of Statistics*, 17(4):1631 – 1661.

[14] Haldane, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika*, 35(3 - 4):414 – 417.

[15] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust Statistics: The Approach based on Influence Functions*, volume 196. John Wiley & Sons.

[16] Huber, P. J. (1981). *Robust Statistics*. Wiley.
[17] Hubert, M., Rousseeuw, P. J., and Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics*, 47(1):64 – 79.
[18] Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles. *The Annals of Statistics*, 25(2):435 – 477.
[19] Liu, R., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *The Annals of Statistics*, 27(3):783–858.
[20] Locantore, N., Marron, J. S., Simpson, D. G., et al. (1999). Robust principal component analysis for functional data. *Test*, 8(1):1 – 73.
[21] Magyar, A. F. and Tyler, D. E. (2014). The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions. *Biometrika*, 101(3):673–688.
[22] Majumdar, S. and Chatterjee, S. (2018). Non-convex penalized multitask regression using data depth-based penalties. *Stat*, 7:e174.
[23] Miao, J. and Ben-Israel, A. (1992). On principal angles between subspaces in $\mathbb{R}^n$. *Linear Algebra and its Applications*, 171:81 – 98.
[24] Minsker, S. (2015). Geometric median and robust estimation in Banach spaces. *Bernoulli*, 21(4):2308 – 2335.
[25] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *Journal of Nonparametric Statistics*, 5(2):201–213.
[26] Niemiro, W. (1992). Asymptotics for M-estimators defined by convex minimization. *The Annals of Statistics*, 20(3):1514–1533.
[27] Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. Springer Science & Business Media.
[28] Serfling, R. (2006). Depth functions in nonparametric multivariate inference. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 72:1.
[29] Sibson, R. (1979). Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 217 – 229.
[30] Sirkiä, S., Taskinen, S., and Oja, H. (2007). Symmetrised m-estimators of multivariate scatter. *Journal of Multivariate Analysis*, 98(8):1611 – 1629.
[31] Taskinen, S., Koch, I., and Oja, H. (2012). Robustifying principal component analysis with spatial sign vectors. *Statistics & Probability Letters*, 82(4):765 – 774.
[32] Tyler, D. E. (1987). A distribution-free m-estimator of multivariate scatter. *The Annals of Statistics*, 15(1):234 – 251.
[33] Wang, L., Peng, B., and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*, 110(512):1658 – 1669.
[34] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482.

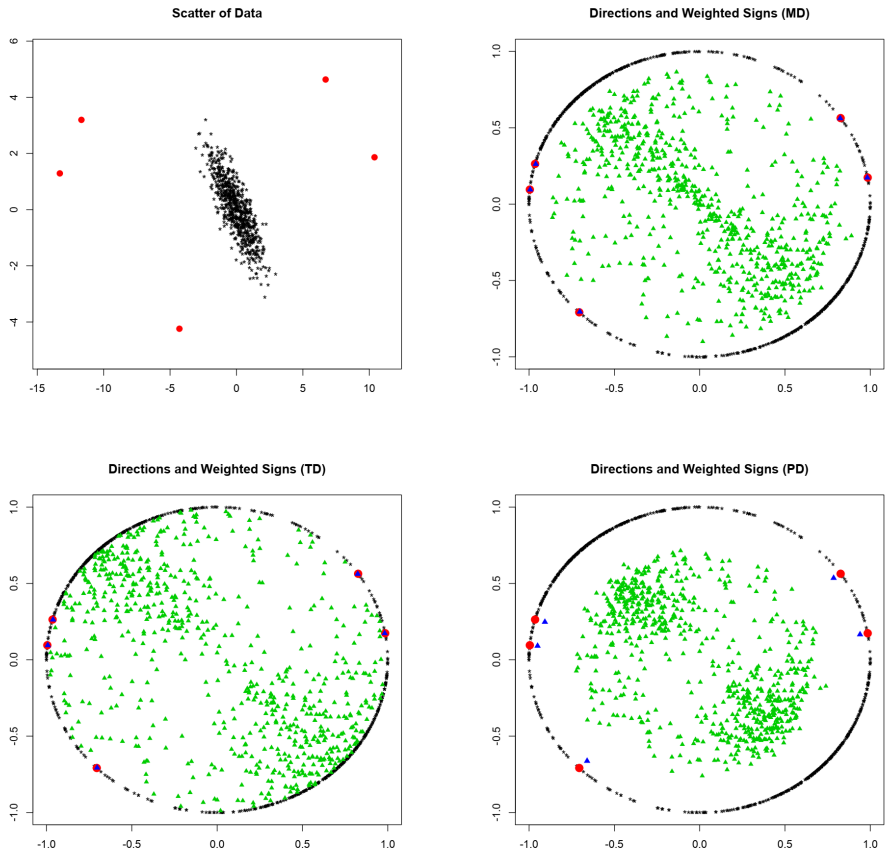

Fig 1: An illustrative bivariate scatter plot in the top left panel where the outliers are identified with red circles, and generalized signs from the same data (black points on the unit radius circle, outliers are red points) in the other panels. In the top right (bottom left, bottom right) panel, weighted signs from the same data with weights obtained using Mahalanobis depth (Tukey depth, projection depth respectively) are presented as green triangles (outliers are identified by blue triangles).

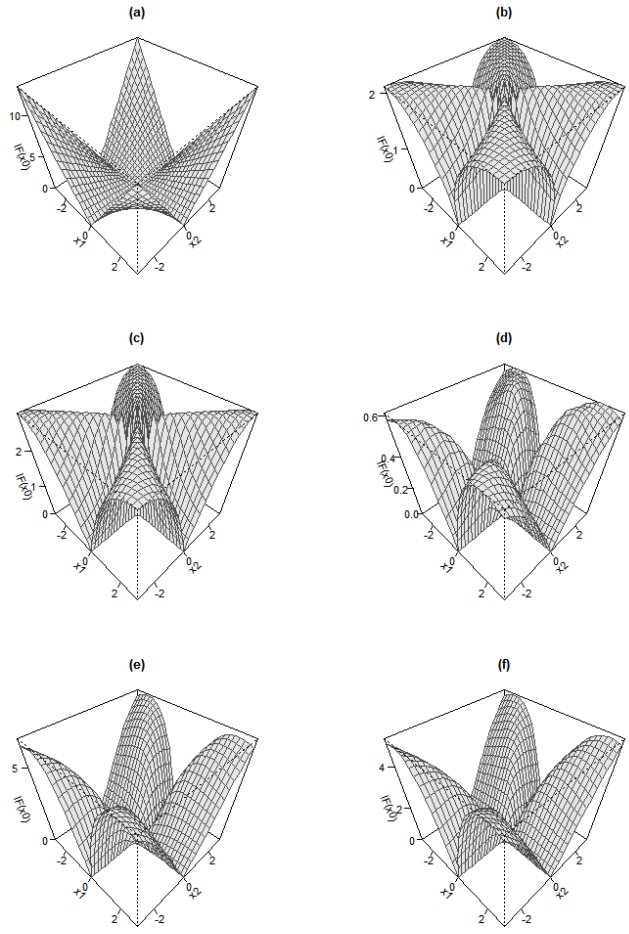

Fig 2: Plot of the norm of influence function for first eigenvector of (a) sample covariance matrix, (b) SCM, (c) Tyler's scatter matrix and $\tilde{\Sigma}$ for weights obtained from (d) Halfspace depth, (e) Mahalanobis depth, (f) Projection depth for a bivariate normal distribution with $\boldsymbol{\mu}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

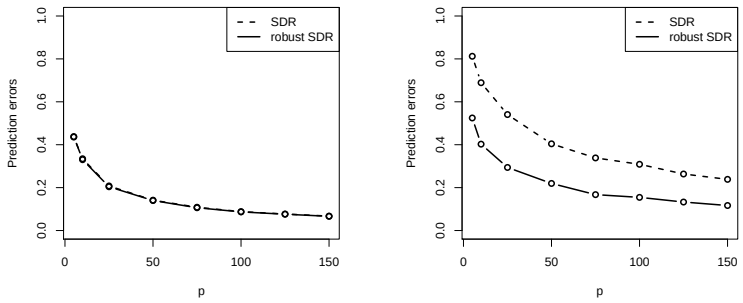

Fig 3: Average prediction errors for two methods of SDR (a) in absence and (b) in presence of outliers

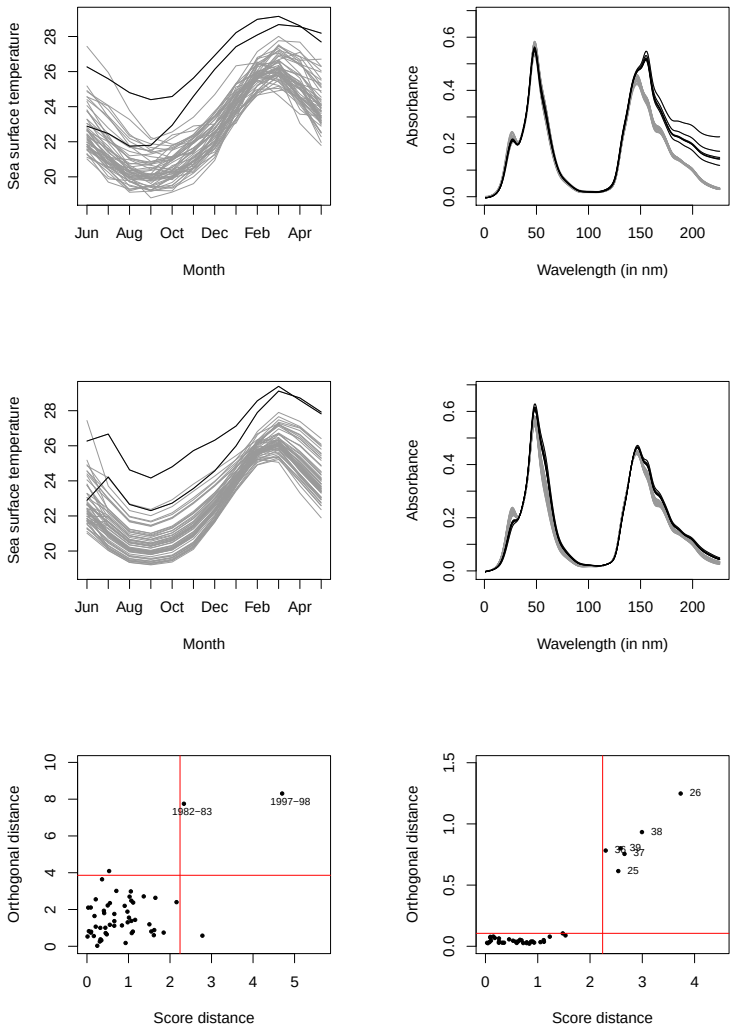

Fig 4: Actual sample curves, their spline approximations and diagnostic plots respectively for El-Niño (a,c,e) and Octane (b,d,f) datasets