Journal of Multivariate Analysis 191 (2022) 105013

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On weighted multivariate sign functions

Subhabrata Majumdar *, Snigdhansu Chatterjee

University of Minnesota, Minneapolis, MN, USA

ARTICLE INFO

Article history:
Received 16 July 2021
Received in revised form 11 April 2022
Accepted 11 April 2022
Available online 21 May 2022

AMS 2020 subject classifications:
primary 62H10
secondary 62H25

Keywords:
Data depth
Multivariate sign
Outlier detection
Principal component analysis
Sufficient dimension reduction

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.

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

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

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 [36] 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 $\{X_i \in \mathbb{R}^p : i \in \{1, \dots, n\}\}$, that we collect together in the $n \times p$ matrix $\mathbf{X} = (X_1; \dots; X_n)^\top$, an approach for robust estimation and inference in multivariate data starts by evaluating (1) on each observation—defining $S_i = S(X_i; \mu_i)$ with respect to some suitable center $\mu_i \in \mathbb{R}^p$—then using these for robust location and scale estimation and inference, including inference for μ_i [27,42,51]. Suppose $\mathbf{S} = (S_1; \dots; S_n)^\top \in \mathbb{R}^{n\times p}$. If the data $\{X_i \in \mathbb{R}^p : i \in \{1, \dots, n\}\}$ are independent, identically distributed (hereafter, i.i.d.) from an elliptically symmetric distribution, then the eigenvectors of $\mathbb{E}(X_1 - \tilde{\mu})(X_1 - \tilde{\mu})^\top$ and of $\mathbb{E}S_1S_1^\top$ are the same for suitable centering parameter $\tilde{\mu} \equiv \mu_i$, that is, the population principal components from the original data and from its sign transformations are the same [48]. 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

* Correspondence to: Splunk, USA.
E-mail address: smajumdar@splunk.com (S. Majumdar).

https://doi.org/10.1016/j.jmva.2022.105013

example, eigenvector estimates obtained from the covariance matrix of S are asymptotically inadmissible [29] and Tyler's M-estimate of scatter [49] 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 the product of the generalized sign function and a weight function derived as a transformation of a data depth function [46,55]. 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

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

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 $R(X_i; \mu, \mathbb{F}) = X_i - \mu_{\mathbb{F}}$, the original centered 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.

We primarily focus on the task of robust dispersion/scatter estimation and robust principal component analysis in this paper. Fig. 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 Mahalanobis 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.

We assume that $\mathcal{X} = \mathbb{R}^p$, that is, the support of the random variable under study is the p-dimensional Euclidean plane for fixed p. However, several of the results of this paper 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. Similarly, we will consider in future the case where $p \to \infty$ as $n \to \infty$, and we anticipate most of the analysis ad results of this paper to be applicable in the situation $p^2/n \to 0$. We assume that the data $X_1, \ldots, X_n$ are independent and identically distributed from an elliptical distribution $\mathbb{F}$ with parameters μ and Σ. A formal definition follows, from [17]:

Definition 1. 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 $\exp\{it^\top \mu\}\phi(t^T \Sigma t)$ for $t \in \mathbb{R}^p$.

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.

There are two unknown quantities in the generalized sign function 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 2.1 we briefly 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. Following that, we present detailed discussions on our primary proposal for a robust measure of dispersion in Section 2.2, followed by a proposed affine equivariant version of it in Section 3, robust estimation of eigenvalues and a third robust estimator for dispersion in Section 4, and a thorough study of robustness and efficiency using influence functions in Section 5. We then report multiple simulation-based numeric studies in Section 6, present several real data examples in Section 7, and concluding remarks in Section 8. All proofs and other technical details are presented in the appendices. All computations are performed on R, with the help of the packages `ddalpha` [44], `fda.usc` [18] and `robustbase` [28] for computing depth-based estimators. Code and data are available in the supplementary material.

In the rest of this paper, all finite-dimensional vectors are column vectors, and for a vector or matrix a, the notation $a^\top$ 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.

The various technical conditions and assumptions that we impose later on the weight function $W(x, \mathbb{F})$ are valid for weights derived from three well-known data depth functions: the halfspace depth, the Mahalanobis depth, and the projection depth. Note that for i.i.d. data, there are three parameters involved here: the location parameter μ in (1), the distribution $\mathbb{F}$ used for the weight function, and the distribution $\mathbb{F}_X$ of X_1. For clarity and to mirror the contexts of how data depth has been used in the literature [26,46,55], we fix $\mathbb{F} = \mathbb{F}_X$ for this paper, although in Section 2.1 we briefly remark

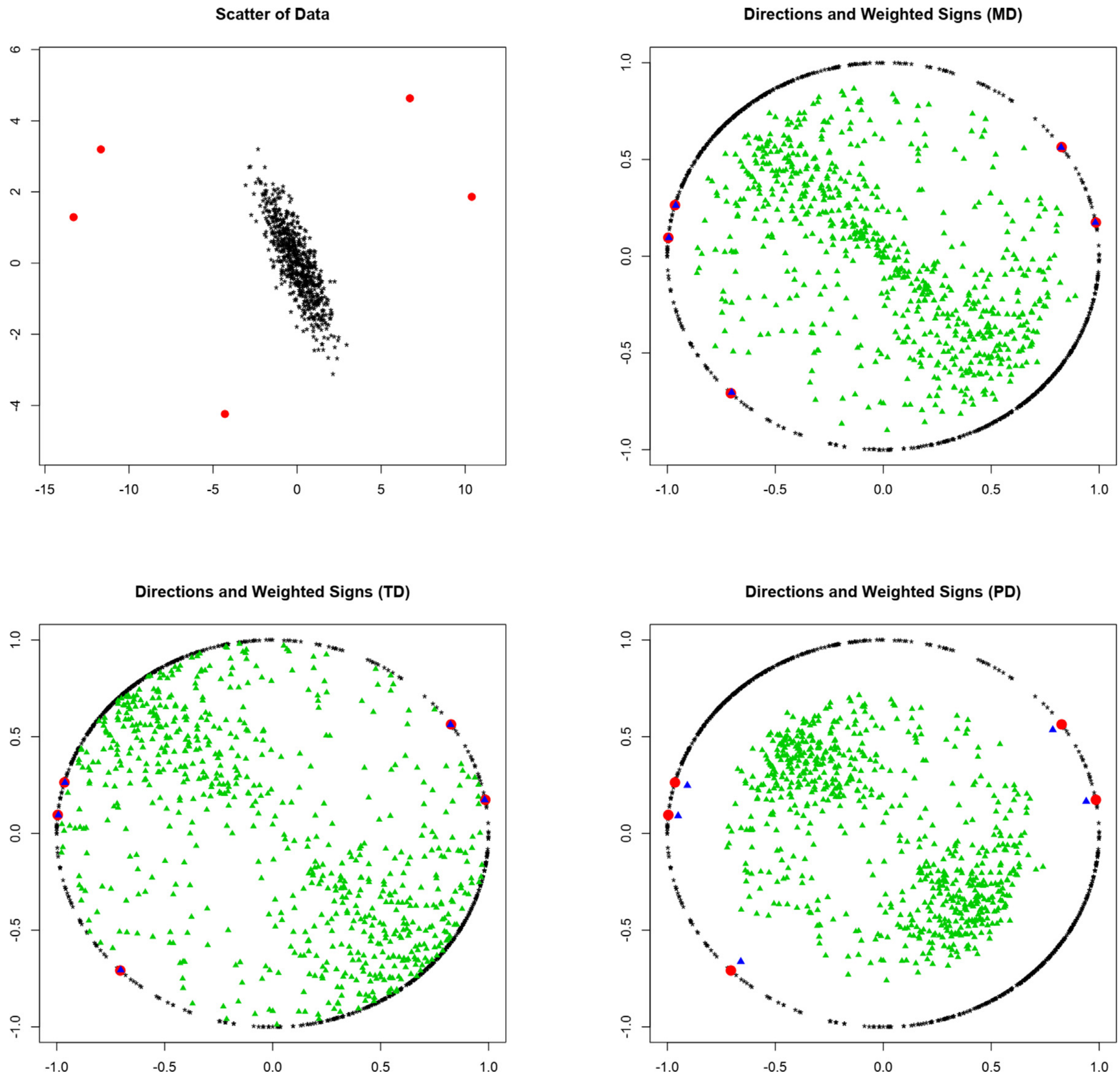

Fig. 1. An illustrative bivariate scatter plot in the top left panel where the outliers are identified with red circles, and generalized signs from the same data (black points on the unit radius circle, outliers are red points) in the other panels. In the top right (bottom left, bottom right) panel, 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). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

on the case when these distributions are different. Also note that in many problems of interest, $\mathbb{F}$ is unknown and data depths are computed using $\mathbb{F}_n$, however, under very standard regularity conditions (for example, see assumptions B1–B3 below), the properties of $W(x, \mathbb{F})$ and $W(x, \mathbb{F}_n)$ are close enough for both asymptotic theory and practical applicability. Additionally, we assume that the weight function is affine invariant, i.e., $W(Ax+b, A\mathbb{F}+b) = W(x, \mathbb{F})$ for $A \in \mathbb{R}^{p\times p}$, $b \in \mathbb{R}^p$.

Some specific choices of weight functions that are compatible with our conditions 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 $x \in \mathcal{X}$, we have $D(\mu_{\mathbb{F}}; \mathbb{F}) \geq D(\mu_{\mathbb{F}} + t(\mathbf{x} - \mu_{\mathbb{F}}); \mathbb{F})$. 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 [46,55] 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 $\mathbb{E}Z = 0_p \in \mathbb{R}^p$ 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 easy to show that results in this paper are valid for the weight functions (*i*) $W_{HSD}(X) \propto \mathbb{F}_{Z_1}(|Z|)$ derived from the half-space depth, (*ii*) $W_{MhD}(X) \propto |Z|^2/(1+|Z|^2)$ derived from the Mahalanobis depth, and (*iii*) $W_{PD}(X) \propto |Z|/(1+|Z|/MAD(Z_1))$, 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. Note that in all these cases, $W(Z)$ is a function of $|Z|$ only. We additionally assume that $\mathbb{E}W^2(Z) < \infty$.

2. Robust measures of multivariate location and scale

2.1. The weighted spatial median

We first propose a robust measure of location below, and also use this section to set up some notations and conditions that we will assume for the rest of the paper. 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 $X \in \mathcal{X}$, and define the function $\Phi(q; X, u, \mathbb{F}) = W(X, \mathbb{F})\big\{|X - q| + \langle u, X - q\rangle\big\}$. We define the $(u, \mathbb{F})$th weighted spatial quantile of $\mathcal{X}$ as the minimizer $q(u, \mathbb{F}) \in \mathcal{X}$ of the expectation of $\Phi(q; X, u, \mathbb{F})$, that is

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

This is a natural generalization of the spatial median [9,20,25,35] ($W(X, \mathbb{F}) \equiv 1$ and $u = \mathbf{0}_p$), or more general spatial quantiles [6,7,34] ($W(X, \mathbb{F}) \equiv 1$). We assume that $\Phi(q; X, u, \mathbb{F})$ is convex in q for $\mathbb{F}$-almost all values of $x \in \mathcal{X}$. For brevity we elaborate only the case of the weighted spatial median (thus $\Psi(q; 0, \mathbb{F}) = \mathbb{E}\big[W(X, \mathbb{F})|X - q|\big]$). 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 acknowledgment that the weight function is used. We denote the unweighted version of this estimator, i.e., 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 $\frac{\partial^2}{\partial q^2}\Psi(q; 0, \mathbb{F})$ exists for all q in a neighborhood of q_0, and we use the notations

$$\Psi_{1W} = \Big(\frac{\partial}{\partial q}\Psi(q_0; 0, \mathbb{F})\Big)\Big(\frac{\partial}{\partial q}\Psi(q_0; 0, \mathbb{F})\Big)^\top, \quad \Psi_{2W} = \frac{\partial^2}{\partial q^2}\Psi(q_0; 0, \mathbb{F}).$$

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 1. *Under assumptions [A1]–[A3], we have*

$$n^{1/2}(\hat{q}_{nW} - q_0) \overset{\mathcal{D}}{\Rightarrow} N(0, \Psi_{2W}^{-1}\Psi_{1W}\Psi_{2W}^{-1}).$$

The proof can be found in Appendix A. Thus, under very standard regularity conditions, the sample weighted spatial median is consistent and is asymptotically normal.

Remark 1. Note that the technical conditions for the result presented in Theorem 1 is one of several alternatives that can conceived, and the scope of this result is broader than what is presented above. First, note that if $\mathbb{F}$ and $\mathbb{F}_X$ are different and the weights are not a function of q, a situation that may arise in hypothesis testing problems where the weights are based on the null distribution, the convexity of $\Phi(q; X_i, u, \mathbb{F})$ follows automatically and is not an assumption. Second, even if $\Phi(q; X_i, u, \mathbb{F})$ is not convex but sufficiently smooth, we can have a central limit theorem, for example, by using techniques similar to [8]. Choices of $\mathbb{F}$ other than $\mathbb{F}_X$, e.g., [30], may lead to interesting interpretations of $W(., \mathbb{F})$ and the resulting location estimator and will be explored further in future.

Let $V_W = \Psi_{2W}^{-1}\Psi_{1W}\Psi_{2W}^{-1}$ be the asymptotic variance of $\hat{q}_{n,W}$ from Theorem 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 pth root of the reciprocals of their determinants. That is,

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

We obtain the following result (see Appendix A for proof):

Table 1
Table of $ARE(\hat{q}_{nW}, \hat{q}_n)$ for spherical distributions: p-variate normal and t_ν-distributions with degree of freedom $\nu \in \{3, 5, 10, 20\}$.

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

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

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

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

Table 1 summarizes the AREs for several families of elliptic distributions, numerically calculated using 10,000 random samples, and taking W as projection depth [52]. Weighted spatial median outperforms its unweighted counterpart for all data dimensions and distribution families considered, with higher ARE for smaller values of p. We shall explore in a principled manner the choices of W and $\mathbb{F}$ that lead to when the weighted spatial median is more efficient than the usual spatial median in future work.

2.2. The weighted sign covariance matrix

We now initiate discussion on the main topic of this paper, on robust dispersion estimation and associated quantities. Consider the spectral decomposition of Σ given by $\Sigma = \Gamma\Lambda\Gamma^\top$, where Γ is an orthogonal matrix and Λ is diagonal with positive diagonal elements $\lambda_1 \geq \cdots \geq \lambda_p$. Also denote the ith eigenvector of Σ by $\gamma_i = (\gamma_{i,1}, \ldots, \gamma_{i,p})^\top$ for $1 \leq i \leq p$. Thus, the ith column of Γ is γ_i. In the rest of this paper we use the notation $\Sigma^{-1/2} = \Lambda^{-1/2}\Gamma^\top$, and hence $Z = \Lambda^{-1/2}\Gamma^\top(X-\mu)$. Recall 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 $W(Z, \mathbb{F}_Z) \equiv W(Z)$.

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

$$S(X; \mu) = \frac{X-\mu}{|X-\mu|} = \frac{R\Gamma\Lambda^{1/2}U}{|\Lambda^{1/2}RU|} = \frac{\Gamma\Lambda^{1/2}U}{|\Lambda^{1/2}U|} = \frac{\Gamma\Lambda^{1/2}Z}{|\Lambda^{1/2}Z|}.$$

As a robust surrogate for $X - \mu$, we consider the following random variable

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

In samples, the equivalent for $\tilde{X}$ is $\hat{\tilde{X}} = W(X, \mathbb{F}_n)S(X; \hat{\mu})$ for a suitable location estimator $\hat{\mu}$, for example, the weighted spatial median. We fix the notation $\mathbb{S}(X; \mu) = S(X; \mu)S(X; \mu)^\top$, and define the following dispersion parameter:

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

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

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

Note that the eigenvalues of $\tilde{\Sigma}$ and Σ are not necessarily the same: $\tilde{\lambda}_i = \lambda_i\mathbb{E}W^2(X)\mathbb{E}[U_i^2/(U^\top\Lambda U)]$. However, since the coordinates of U are iid, the ordering of eigenvalues is preserved post-transformation.

2.3. 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*:

$$\sup_{x\in\mathcal{X}} |W(x,\mathbb{F}_n) - W(x,\mathbb{F})| \to 0 \text{ almost surely as } n\to\infty.$$

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 $\epsilon\in(0,\delta)$

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

In the above, δ_x denotes point mass at x. These properties are easily satisfied 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 —similar to the unweighted case in [15]. 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 2.1. The proof is given in Appendix A.

Lemma 1. *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 for* $|c|=1$, *we have* $\mathbb{E}c^\top R_n c = o(n^{-1})$.

If $\mathbb{F}$ is the p-dimensional standard normal distribution, the moment condition $\mathbb{E}|X-\mu|^{-4}<\infty$ is easily seen to hold when $p>4$. Similar verification of this condition can be done for several other distributions. Also, the moment condition $\mathbb{E}|X-\mu|^{-4}<\infty$ can be circumvented by slightly redefining the generalized sign function $S(x;\mu)$ as zero whenever $|x-\mu|<\epsilon_n$ for an appropriately decreasing sequence $\{\epsilon_n\}$.

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 $\tilde{\Sigma}$ (see Appendix A for proof).

Theorem 3. *Assume the conditions of Lemma 1. Then*

$$n^{1/2}\left[\frac{1}{n}\sum_{i=1}^n W_n^2(X_i,\mathbb{F}_n)vec(\mathbb{S}(X_i;\hat{\mu}_n)) - \mathbb{E}W^2(X)vec(\mathbb{S}(X;\mu))\right] \stackrel{\mathcal{D}}{\Rightarrow} N_{p^2}\big(0,V_W\big),$$

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

The asymptotic normality follows from our assumptions and as a direct consequence of Lemma 1. 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 Appendix B.

We now use Theorem 3 to obtain consistency results for the eigenvectors obtained from

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

Suppose that $\tilde{\Lambda}_1>\cdots>\tilde{\Lambda}_p$ are the eigenvalues of $\tilde{\Sigma}$, which we assume are all distinct values.

Theorem 4. *Suppose the spectral decomposition of* $\hat{\tilde{\Sigma}}$ *is given by* $\hat{\tilde{\Sigma}} = \widehat{\Gamma}\hat{\tilde{\Lambda}}\widehat{\Gamma}^\top$. *Then the matrix of centered and scaled eigenvectors* $G_n = n^{1/2}(\widehat{\Gamma}-\Gamma)$ *and the vector of centered and scaled eigenvalues* $L_n = n^{1/2}(\hat{\tilde{\Lambda}}-\tilde{\Lambda})$ *have asymptotically 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*

$$\sum_{k=1,k\neq i}^p \left[\tilde{\Lambda}_i-\tilde{\Lambda}_k\right]^{-2}\mathbb{E}\Big[W^4(Z,\mathbb{F}_Z)\big(\mathbb{S}_{i,k}(\Lambda^{1/2}Z;\mathbf{0})\big)^2\Big]\gamma_k\gamma_k^\top, \quad i=j,$$

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

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*

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

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

The proof of this result follows from using Theorem 3 and using techniques similar to a corresponding result in [48]. We omit the algebraic details here and present it in Appendix A.

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

$$A\mathbb{V}(\sqrt{n}\hat{\gamma}_i) = \sum_{k=1;k\neq i}^{p} \frac{\lambda_i\lambda_k}{(\lambda_i-\lambda_k)^2}\gamma_k\gamma_k^\top; \quad 1 \le i \le p. \tag{2}$$

Suppose $\widehat{\tilde{\gamma}}_i$ is the ith eigenvector of $\widehat{\tilde{\Sigma}}$, whose asymptotic behavior is presented above in Theorem 4.

This leads to the following useful result:

Corollary 2. *The asymptotic relative efficiency of $\widehat{\tilde{\gamma}}_i$, relative to $\hat{\gamma}_i$, is given by*

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

The proof of this Corollary is immediate, by plugging in the asymptotic variances of $\hat{\gamma}_i$ and $\widehat{\tilde{\gamma}}_i$ from (2) and Theorem 4, respectively.

3. 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 $T_Y = AT_XA^\top$. 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 [23] to construct an affine equivariant counterpart of the $\tilde{\Sigma}$. Specifically, we implicitly define

$$\Sigma_* = \frac{p}{\mathbb{V}W(X)}\mathbb{E}\left[\frac{W^2(X)(X-\mu)(X-\mu)^\top}{(X-\mu)^\top\Sigma_*^{-1}(X-\mu)}\right]. \tag{3}$$

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

$$\mathbb{E}\left[u(|Z_M|)\frac{Z_MZ_M^\top}{|Z_M|^2} - v(|Z_M|)\mathbb{I}_p\right] = 0, \tag{4}$$

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 [23]:

- 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) $P(H) \le 1/p$.

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 (3) 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

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

The asymptotic properties of $\hat{\Sigma}_*$ can be obtained using methods similar to those of Section 2, and techniques presented in [13] and elsewhere. We summarize these properties in the following result. The proof is provided in Appendix A.

Theorem 5. *The asymptotic covariance matrix of an eigenvector of the sample affine equivariant scatter functional $\hat{\Sigma}_*$ is given by*

$$V_{12}\sum_{k=1,k\neq i}^{p}\frac{\lambda_i\lambda_k}{\lambda_i-\lambda_k}\gamma_i\gamma_k^\top,$$

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 $\hat{\gamma}_{*,i}$ is the ith eigenvector of $\hat{\Sigma}_*$,

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

4. Robust estimation of eigenvalues and Σ

As seen in Theorem 2, 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 2.1 suffices. We start by computing the sample version $\widehat{\tilde{\Sigma}}$ and its spectral decomposition: $\widehat{\tilde{\Sigma}} = \widehat{\Gamma}\widehat{\tilde{\Lambda}}\widehat{\Gamma}^\top$. We then use the following steps:

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

$$\lambda_{i,j}^\dagger = \frac{1}{|G_j|}\sum_{l \in G_j}\left(S_{li} - \bar{S}_{G_j,i}\right)^2; \quad i \in \{1, \dots, p\};\ j \in \{1, \dots, k\}, \quad \text{where}$$

$\bar{\mathbf{S}}_{G_j} = (\bar{S}_{G_j,1}, \dots, \bar{S}_{G_j,p})^\top$ 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:

$$\lambda_i^\dagger = \text{median}(\lambda_{i,1}^\dagger, \dots, \lambda_{i,k}^\dagger); \quad i \in \{1, \dots, p\}.$$

We collect $\lambda_i^\dagger, i \in \{1, \dots, p\}$ in the diagonal matrix $\Lambda^\dagger = \text{diag}(\lambda_1^\dagger, \dots, \lambda_p^\dagger)$. The number of subgroups used to calculate this median-of-small-variances estimator can be determined following [34]. 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 $\Sigma^\dagger = \widehat{\Gamma}\Lambda^\dagger\widehat{\Gamma}^\top$. Let $|A|_F$ denote the Frobenius norm of a matrix A, in other words, $|A|_F = (\text{trace}(A^\top A))^{1/2}$. The following result establishes that this is a consistent estimator of Σ (see Appendix A for proof):

Theorem 6. *Suppose that as $n \to \infty$, $k \to \infty$ and $n/k \to \infty$. Then we have*

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

5. Influence functions of dispersion measures

We retain the framework adopted in Section 2, 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 $\lambda_1 > \cdots > \lambda_p$, 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 present some results relating to influence functions in the current context, with proofs given in Appendix A. Influence functions quantify how much influence a sample point, especially an infinitesimal contamination, has on any functional of a probability distribution [22]. 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 $T(\mathbb{H}) = \mathbb{E}_{\mathbb{H}}f$ 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 $\tilde{\lambda}_1 > \cdots > \tilde{\lambda}_p$ are the eigenvalues of $\tilde{\Sigma}$, which we assume are all distinct values.

Proposition 1. *The influence function of $\tilde{\gamma}_i$, the ith eigenvector of $\tilde{\Sigma}$, is as follows:*

$$IF(x_0; \tilde{\gamma}_i, \mathbb{F}) = W^2(x_0)\sum_{k=1; k \neq i}^{p}\frac{\mathbb{S}_{ik}(x_0; \mu)}{\tilde{\lambda}_i - \tilde{\lambda}_k}\gamma_k.$$

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 1 is bounded, indicating good robustness properties of the principal component analysis.

We now derive the influence function for Σ_*.

Proposition 2. *The influence function of Σ_* is given by $IF(x_0, \Sigma_*, \mathbb{F}) = \alpha_{\Sigma_*}(|x_0|; \mathbb{F}_Z)\mathbb{S}(x_0; \mu) - \beta_{\Sigma_*}(|x_0|; \mathbb{F}_Z)\Sigma_*$, for constants $\alpha_{\Sigma_*}(|x_0|; \mathbb{F}_Z)$ and $\beta_{\Sigma_*}(|x_0|; \mathbb{F}_Z)$ that depend on Σ_* and $\mathbb{F}_Z$.*

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

Proposition 3. *The influence function of γ_{*i} may be obtained as*

$$IF(x_0; \gamma_{*i}, \mathbb{F}) = \alpha_{\Sigma_*}(|x_0|; \mathbb{F}_Z) \sum_{k=1; k\neq i}^{p} \frac{\mathbb{S}_{ik}(x_0; \mu)}{\lambda_i - \lambda_k}\gamma_k; \quad \alpha_{\Sigma_*}(|x_0|; \mathbb{F}_Z) = \frac{p(p+2)u(|z_0|)}{\mathbb{E}(pu(|Z|) + u'(|Z|))},$$

where $z_0 = \Sigma^{-1/2}(x_0 - \mu)$.

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 [23], which in turn implies that the influence function for a principal component based on Σ_* is also bounded.

6. Simulation studies

We report extensive simulation studies on several properties relating to $\tilde{\Sigma}$, Σ_*, and their eigenvalues and eigenvectors—on datasets with or without influential outlying points—to illustrate the finite sample efficiency and robustness properties of the proposed weighted estimators. We compare these proposed estimators with techniques that exist in literature, specifically, the Sign Covariance Matrix (SCM) and Tyler's M estimate of scatter [49].

6.1. 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 $p = 4$. To generate the data matrices $\mathbf{X}$, we consider four elliptical distributions, as well as three non-elliptical settings: (i) normal distribution with the first coordinates of 10% randomly selected rows of $\mathbf{X}$ shifted by $1000 \max_{i,j} x_{ij}$, (ii) normal distribution with the first coordinates of 30% random rows shifted, and (iii) a gaussian copula with correlation 0.8, paired with Beta(3, 3) marginals. From every distribution we draw 10 000 samples each for sample sizes $n \in \{50, 100, \ldots, 500\}$. 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 [33] to find out error estimates for the first eigenvector of a scatter matrix. In our case, the first eigenvector is

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

We measure the prediction error for an eigenvector estimate $\tilde{\gamma}_1$ using the smallest angle between the true and predicted vectors, i.e., $\cos^{-1}|\tilde{\gamma}_1^T\hat{\gamma}_1|$. 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):

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

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

We present the results from this simulation exercise in Fig. 2. It can be seen that $\tilde{\Sigma}$-based estimators outperform SCM and Tyler's M-estimator of scatter. Among the depth functions considered, Mahalanobis depth has highest efficiency without the presence of any outliers. However its performance degrades in the contaminated settings—evidently because of the use of non-robust location and scatter estimates in its calculation. Projection and halfspace depth-weighted estimators outperform sign-based methods in presence of a small amount of (10%) outliers, but the SCM is more robust for small sample sizes and high (30%) contamination. In terms of computational load, SCM and Tyler take the lowest time, but $\tilde{\Sigma}$-H and $\tilde{\Sigma}$-P have competitive orders of magnitude. The affine equivariant Σ_* estimators are more efficient than their $\tilde{\Sigma}$ counterparts. However this efficiency comes at the cost of high computation cost, and loss of robustness. Overall, projection depth based estimators ($\tilde{\Sigma}$-P) strike a good balance between efficiency and computation time in our simulation settings.

6.2. Influence function comparison

In Fig. 3 we consider first eigenvectors of $\tilde{\Sigma}$, the Sign Covariance Matrix (SCM) and Tyler's shape matrix [49]. We generate data from and set $\mathbb{F} \equiv \mathcal{N}_2(0, \text{diag}(2, 1))$ and plot norms of the eigenvector influence functions for different

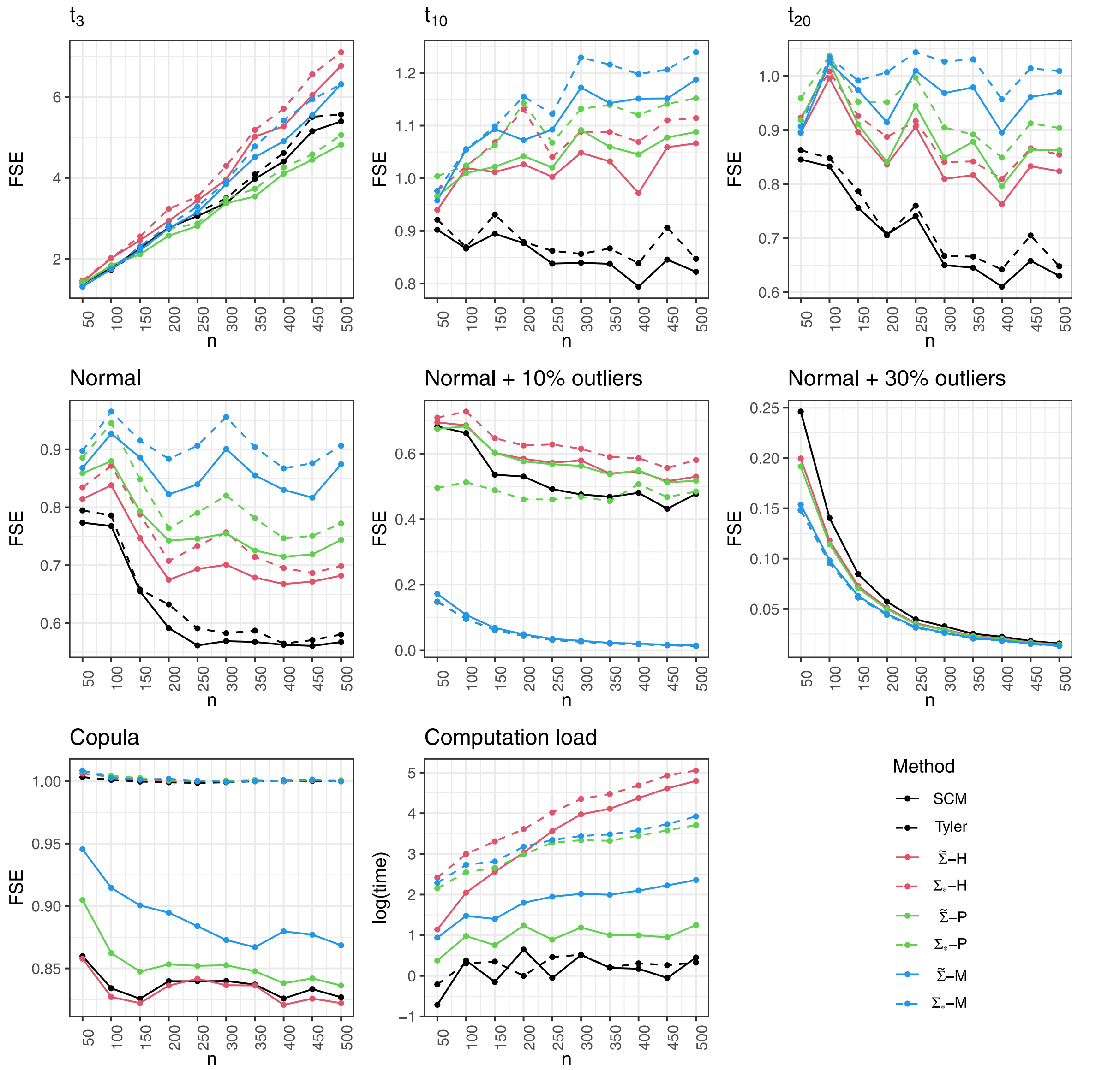

Fig. 2. Finite sample efficiencies of the first eigenvector estimates based on scatter matrices of several distributions in dimension $p = 4$. The notation H, M or P after $\tilde{\Sigma}$ or Σ_* indicates the depth function used for the weights: H = halfspace depth, M = Mahalanobis depth, P = projection depth. Computation load is calculated for the normal distribution case.

values of x_0. Let us denote the ith 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:

$$IF(x_0; \gamma_{S,i}, \mathbb{F}) = \sum_{k=1;k\neq i}^{p} \frac{\mathbb{S}_{ik}(x_0; \mu)}{\lambda_{S,i} - \lambda_{S,k}} \gamma_k; \quad \text{where } \lambda_{S,i} = \mathbb{E}_Z \left(\frac{\lambda_i z_i^2}{\sum_{j=1}^{p} \lambda_j z_j^2} \right),$$

$$IF(x_0; \gamma_{T,i}, \mathbb{F}) = (p+2) \sum_{k=1;k\neq i}^{p} \frac{\mathbb{S}_{ik}(x_0; \mu)}{\lambda_i - \lambda_k} \gamma_k.$$

Panels (b) and (c) in Fig. 3, 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

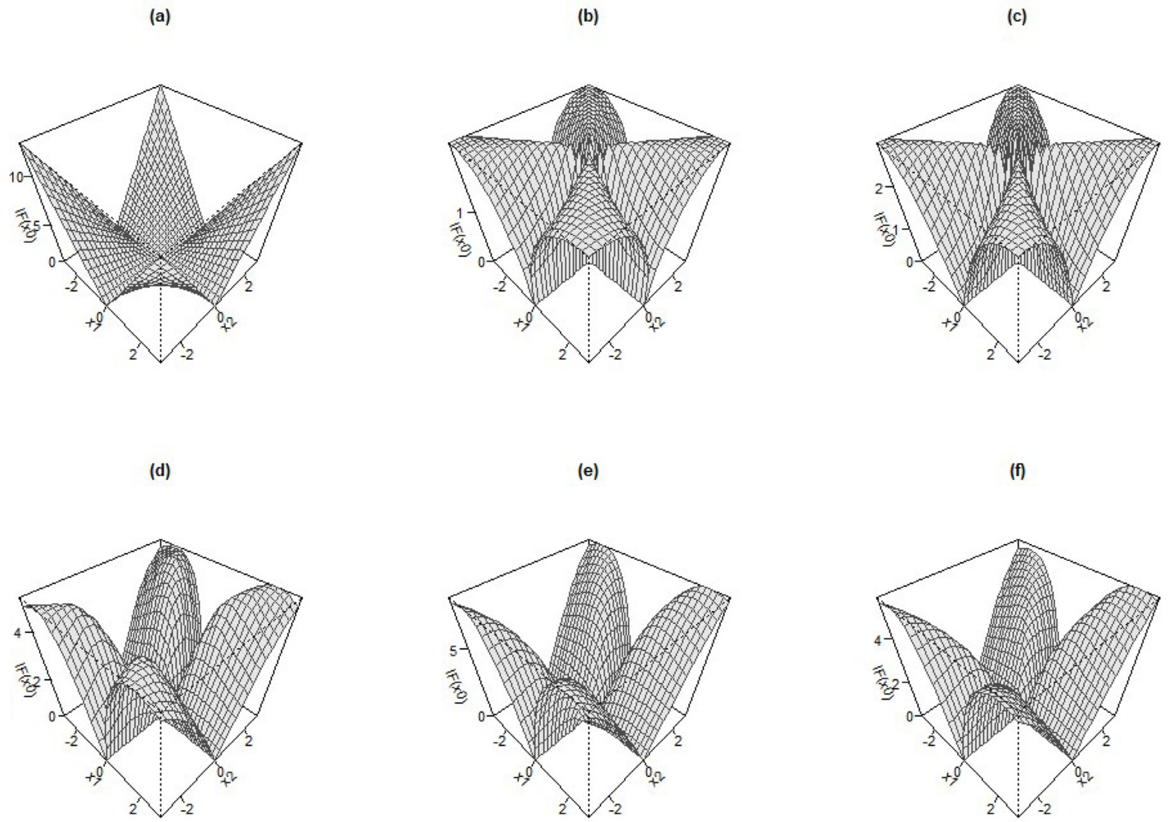

Fig. 3. Plot of the norm of influence function for first eigenvectors of (a) sample covariance matrix, (b) SCM, (c) Tyler's scatter matrix, and $\tilde{\Sigma}$ for weights obtained from (d) halfspace depth, (e) Mahalanobis depth, (f) projection depth for a bivariate normal distribution with $\mu = \mathbf{0}$, $\Sigma = \text{diag}(2, 1)$.

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.

6.3. Efficiency of affine equivariant robust estimator

To study the finite sample efficiency properties of Σ_*, 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 5. 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 covariance matrix even when the data comes from multivariate normal distribution.

6.4. 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 [1,10,11] and references and citations of these articles. In the context of multivariate-response ($Y_i \in \mathbb{R}^q$) linear regression, the envelope method

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).

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

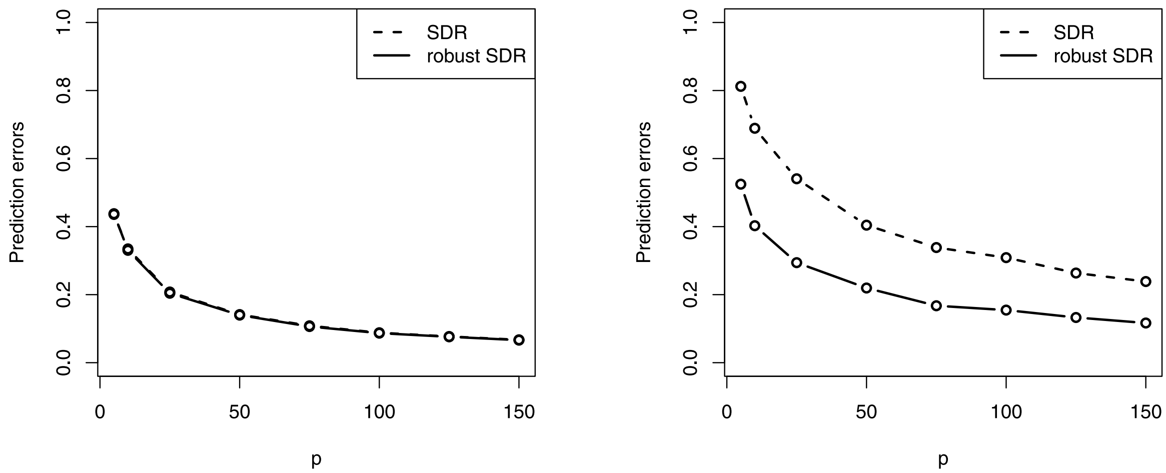

Fig. 4. Average SDR prediction errors (left) in absence and (right) in presence of outliers. For each value of p, prediction errors are calculated over 100 replications of the data setting.

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

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

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 [1], and may be used for supervised learning. Nonlinear regression models may also be handled similarly.

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

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

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]. In a 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 $Y_1, \dots, Y_n$ 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 $p \in \{5, 10, 25, 50, 75, 100, 125, 150\}$.

The left panel of Fig. 4 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. The right panel 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.

7. 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 outlier detection functional (and hence infinite-dimensional) data.

We follow the approach of [5] for performing robust principal component analysis on functional data using the estimated eigenvectors from $\widehat{\tilde{\Sigma}}$. Suppose the data consistent 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 $\{t_1, \ldots, t_m\}$. We model each of these functions as a linear combination of p mutually orthogonal B-spline basis functions $\mathcal{D} = \{\delta_1, \ldots, \delta_p\}$. We map data for each of the functions onto the coordinate system formed by the spline basis:

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

We then model the ith row of the $n \times p$ matrix $T(\mathcal{F}, \mathcal{D}) \equiv T$, denoted by $\mathbf{T}_i = \mu + Ps_i + e_i$, 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 $\widehat{\tilde{\Sigma}}$. Define $\widehat{\mathbf{T}}_i = \hat{\mu} + \hat{P}\hat{s}_i$. The orthogonal distance (OD) and score distance (SD) corresponding to this projection are defined as

$$OD_i = |\mathbf{T}_i - \widehat{\mathbf{T}}_i|, \quad SD_i = \sqrt{\sum_{j=1}^{q} \frac{\hat{s}_{ij}^2}{\hat{\lambda}_j}};$$

where $\hat{\lambda}_1, \ldots, \hat{\lambda}_q$ are the top eigenvalues from $\widehat{\tilde{\Sigma}}$. For outlier detection, following [24] 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 datasets. 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) (Fig. 5 top-left panel). Second, we consider the octane data, which consists of 226 variables and 39 observations [16]. 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. (Fig. 5 top-right panel).

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 middle-top panel of Fig. 5. Using our proposed methodology with $q = 1$ results in two points having their SD and OD larger than cutoff, depicted in top-right panel of Fig. 5. 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 the bottom panels of Fig. 5.

8. Conclusion

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 on 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.

Although our weight functions are broad-based, we focus on transformations of data depth functions as weights in this paper. It may be possible to link existing work on multivariate rank-based methods [31,50] to potentially interesting choices of weighted signs, such as with other application areas. Examples of such directions include the use of ℓ_1-norm based methods [14,41], the use of signed ranks [43], parametric tests for eigenvalue/vector estimates [43], and non-parametric tests for location [21,54]. As pointed by one of the reviewers, clustering is another interesting domain of application where existing notions of multivariate rank [37–39] can inform the use of weighted signs. For example, the weights themselves may be optimized to increase efficiency relative to cluster sizes.

Along with Tyler's M estimate, other variations of SCM and robust estimation of scatter matrices are worth exploring for comparison and generalization. For example, the k-step SCM—a finite-iteration intermediary between SCM and Tyler's M

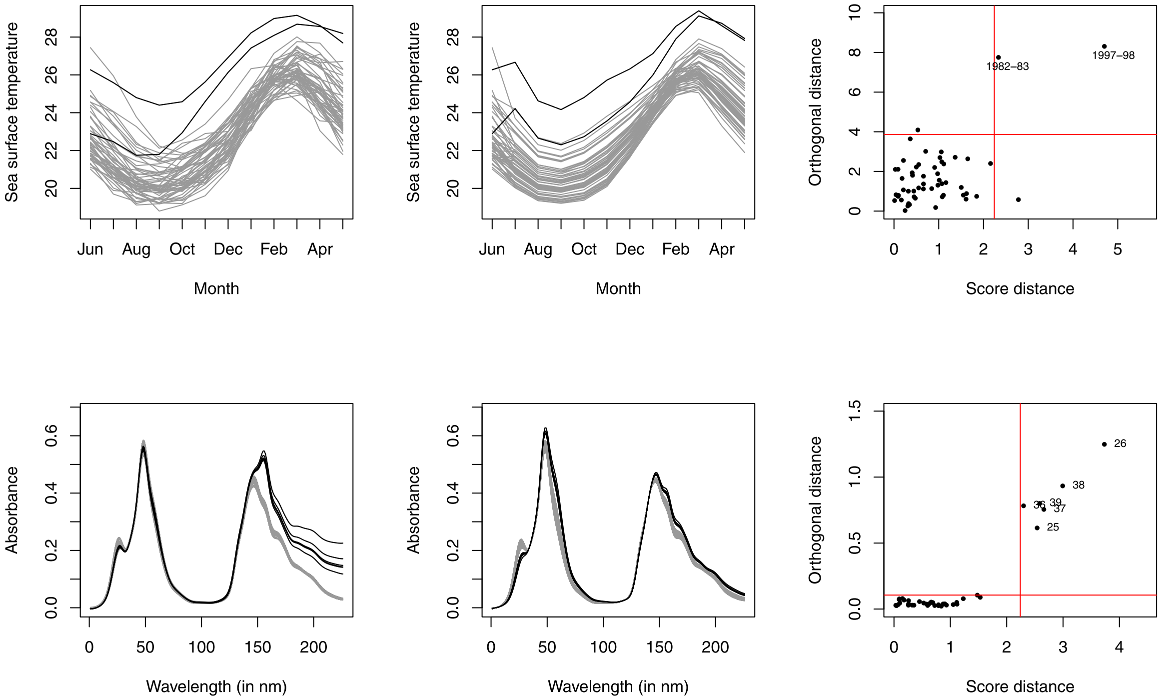

Fig. 5. Actual sample curves, their spline approximations and diagnostic plots, respectively, for (top) El-Niño and (bottom) Octane datasets.

estimate—aims to balance robustness and efficiency [12], and the generalized SCM [45] in essence uses an orthogonally equivariant weight function. Finally, the depth-weighted Stahel–Donoho estimates of location and scatter [53] may be incorporated in a slightly relaxed version of our weighted signs framework by considering (transformations of) depth functions multiplied by the ℓ_2-norm of a vector as the weight function.

Several of our results stated in this paper 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 only few conceptual challenges to such extensions. However, there are several technical and algebraic challenges (e.g., see [4]), which may be tackled using recent developments in high-dimensional M-estimation [32].

CRediT authorship contribution statement

Subhabrata Majumdar: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Snigdhansu Chatterjee:** Conceptualization, Data curation, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Acknowledgments

We thank the Editor, Associate Editor, and two anonymous referees for their insightful comments that led to improvements in the paper. The research of SC is partially supported by the US National Science Foundation grants 1737918, 1939916 and 1939956 and a grant from Cisco Systems Inc.

Appendix A. Proofs

Proof of Theorem 1. This result can be proved in several different ways. We use the techniques of [19,40] for our approach. Specifically, following Theorem 4 in [40]—which traces back to [19] with slightly relaxed conditions—we get

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

where $\nabla \Phi(q; X_i, u, \mathbb{F})$ is any measurable subgradient of $\Phi(q; X_i, u, \mathbb{F})$. Theorem 1 now follows using the techniques of [40], specifically Theorem 4 and the first paragraph of page 1517 therein. □

Proof of Corollary 1. 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

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

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. □

Proof of Theorem 2. Fix any index $i \in \{1, \ldots, p\}$. Consider the vector $\tilde{U}$ such that

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

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

$$\mathbb{E}\frac{U_i U_j}{U^\top \Lambda U} = \mathbb{E}\frac{\tilde{U}_i\tilde{U}_j}{\tilde{U}^\top \Lambda \tilde{U}} = -\mathbb{E}\frac{U_i U_j}{U^\top \Lambda U}.$$

so that $\mathbb{E}S(X;\mu)S(X;\mu)^\top = \Gamma\Lambda_S\Gamma^\top$, as established in Theorem 1 of [48]. 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

$$\tilde{\Sigma} = \mathbb{E}\tilde{X}\tilde{X}^\top = \mathbb{E}W^2(X)S(X;\mu)S(X;\mu)^\top = \mathbb{E}W^2(X)\mathbb{E}S(X;\mu)S(X;\mu)^\top = \Gamma\Lambda_W\Gamma^\top,$$

where Λ_W is a diagonal matrix. □

Proof of Lemma 1. The proof is mostly algebra, and we provide a sketch of the main arguments. We have

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

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

Notice that the (j,k)th element of T_2 is given by $n^{-1}\sum_{i=1}^n |X_i - \mu|^{-2}\left\{W_n^2(X_i,\mathbb{F}_n) - W^2(X_i,\mathbb{F})\right\}(X_{i,j} - \mu_j)(X_{i,k} - \mu_k)$, and hence

$$\begin{aligned}c^\top T_2 c = \sum_{j,k} c_j c_k T_{2,j,k} &\leq Mn^{-1}\sum_{i=1}^n |X_i - \mu|^{-2}\left\{|W_n(X_i,\mathbb{F}_n) - W_n(X_i,\mathbb{F}_{n,-i})|\right\}\left(c^\top(X_i - \mu)\right)^2\\ &+ Mn^{-1}\sum_{i=1}^n |X_i - \mu|^{-2}\left\{|W_n(X_i,\mathbb{F}_{n,-i}) - W(X_i,\mathbb{F})|\right\}\left(c^\top(X_i - \mu)\right)^2\\ &= Mn^{-1}\sum_{i=1}^n T_{21i} + Mn^{-1}\sum_{i=1}^n T_{22i} = T_{21} + T_{22}.\end{aligned}$$

Let $H(X_i) = |X_i - \mu|^{-2}\left(c^\top(X_i - \mu)\right)^2$, 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 $A_1 \cap \ldots \cap A_n$, $T_{21} \leq Mn^{-2}H(X_i)$ and the conclusion follows for this part.

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

Proof of Theorem 3. The quantity in the statement of the theorem can be broken down as:

$$n^{1/2}\left[\frac{1}{n}\sum_{i=1}^{n} W_n^2(X_i,\mathbb{F}_n)vec(\mathbb{S}(X_i;\hat{\mu}_n)) - \frac{1}{n}\sum_{i=1}^{n} W^2(X_i,\mathbb{F})vec(\mathbb{S}(X_i;\mu))\right]$$
$$+n^{1/2}\left[\frac{1}{n}\sum_{i=1}^{n} W^2(X_i,\mathbb{F})vec(\mathbb{S}(X_i;\mu)) - \mathbb{E}W^2(X)vec(\mathbb{S}(X;\mu))\right]$$

The first part goes to 0 in probability by Lemma 1. Applying Slutsky's theorem along with Central Limit Theorem, we get the required convergence. □

Proof of Theorem 4. We suppose $G_n = (g_1,\dots,g_p), L_n = \text{diag}(l_1,\dots,l_p)$. In spirit, this corollary is similar to Theorem 13.5.1 in [2]. We start with the following result, due to [48], which allows us to obtain asymptotic joint distributions of eigenvectors and eigenvalues of $\hat{\hat{\Sigma}}$, provided we know the limiting distribution of $\hat{\hat{\Sigma}}$ itself:

Theorem 7. *Let $\mathbb{F}_\Lambda$ be defined as before, and $\hat{C}$ be any positive definite symmetric $p\times p$ matrix such that at F_Λ the limiting distribution of $\sqrt{n}\,\text{vec}(\hat{C}-\Lambda)$ is a p^2-variate (singular) normal distribution with mean zero. Write the spectral decomposition of $\hat{C}$ as $\hat{C}=\hat{P}\hat{\Lambda}\hat{P}^\top$. Then the limiting distributions of $\sqrt{n}\,\text{vec}(\hat{P}-\mathbb{I}_p)$ and $\sqrt{n}\,\text{vec}(\hat{\Lambda}-\Lambda)$ are multivariate (singular) normal and*

$$\sqrt{n}\,\text{vec}(\hat{C}-\Lambda) = \left[(\Lambda\otimes\mathbb{I}_p)-(\mathbb{I}_p\otimes\Lambda)\right]\sqrt{n}\,\text{vec}(\hat{P}-\mathbb{I}_p) + \sqrt{n}\,\text{vec}(\hat{\Lambda}-\Lambda) + o_P(1) \tag{A.1}$$

The first matrix picks only off-diagonal elements of the left-hand side and the second one only diagonal elements. We shall now use this as well as the form of the asymptotic covariance matrix of the vectorized $\hat{\hat{\Sigma}}$, i.e., V_W to obtain limiting variance and covariances of eigenvalues and eigenvectors.

Due to the decomposition (A.1) we have, for $\mathbb{F}_\Lambda$, the following relation between any off-diagonal element of $\hat{\hat{\Lambda}}$ and the corresponding element in the estimate of eigenvectors, say $\hat{\hat{\Gamma}}_\Lambda$ as $\sqrt{n}\hat{\hat{\gamma}}_{\Lambda,ij} = \sqrt{n}\frac{\hat{\hat{\Lambda}}(i,j)}{\tilde{\lambda}_i-\tilde{\lambda}_j}$; $\quad i\neq j$. So that for eigenvector estimates of the original $\mathbb{F}$ we have

$$\sqrt{n}(\hat{\hat{\gamma}}_i-\gamma_i) = \sqrt{n}\Gamma(\hat{\hat{\gamma}}_{\Lambda,i}-e_i) = \sqrt{n}\left[\sum_{k=1;k\neq i}^{p}\hat{\hat{\gamma}}_{\Lambda,i,k}\gamma_k + (\hat{\hat{\gamma}}_{\Lambda,i,i}-1)\gamma_i\right] \tag{A.2}$$

Now $\sqrt{n}(\hat{\hat{\gamma}}_{\Lambda,i,i}-1)=o_P(1)$ and $A\mathbb{V}(\sqrt{n}\hat{\hat{\Lambda}}(i,k),\sqrt{n}\hat{\hat{\Lambda}}(i,l))=0$ for $k\neq l$, so the above equation implies

$$A\mathbb{V}(g_i) = A\mathbb{V}(\sqrt{n}(\hat{\hat{\gamma}}_i-\gamma_i)) = \sum_{k=1;k\neq i}^{p}\frac{A\mathbb{V}(\sqrt{n}\hat{\hat{\Lambda}}(i,k))}{(\tilde{\lambda}_i-\tilde{\lambda}_k)^2}\gamma_k\gamma_k^\top$$

For the covariance terms, from (A.2) we get, for $i\neq j$,

$$A\mathbb{V}(g_i,g_j) = A\mathbb{V}(\sqrt{n}(\hat{\hat{\gamma}}_i-\gamma_i),\sqrt{n}(\hat{\hat{\gamma}}_j-\gamma_j)) = -\frac{A\mathbb{V}(\sqrt{n}\hat{\hat{\Lambda}}(i,j))}{(\tilde{\lambda}_i-\tilde{\lambda}_j)^2}\gamma_j\gamma_i^\top.$$

The exact forms given in the statement of the corollary now follow from the Form of V_W in Appendix B.

For the on-diagonal elements of $\hat{\hat{\Lambda}}$, using Theorem 7 we have for the ith eigenvalue of $\hat{\hat{\Lambda}}$, say $\lambda_{\Lambda,i}$, $\sqrt{n}\hat{\hat{\lambda}}_{\Lambda,i}=\sqrt{n}\hat{\hat{\Lambda}}(i,i)$, for $i=1,\dots,p$. Hence

$$A\mathbb{V}(l_i) = A\mathbb{V}(\sqrt{n}(\hat{\hat{\lambda}}_{\Lambda,i}-\tilde{\lambda}_i)) = A\mathbb{V}(\sqrt{n}(\hat{\hat{\lambda}}_{\Lambda,i}-\tilde{\lambda}_{\Lambda,i})) = A\mathbb{V}(\sqrt{n}\hat{\hat{\Lambda}}(i,i))$$

A similar derivation gives the expression for $A\mathbb{V}(l_i,l_j); i\neq j$. Finally, since the asymptotic covariance between an on-diagonal and an off-diagonal element of $\hat{\hat{\Lambda}}$, it follows that the elements of G_n and diagonal elements of L_n are independent. □

Proof of Theorem 5. Following [22], the influence function for the eigenvector of an affine equivariant dispersion parameter Σ_M of an elliptic distribution—defined as in (4)—is

$$IF(x_0;\gamma_i,\Sigma_M) = \alpha_{\Sigma_M}(|z_0|)\sum_{k=1,k\neq i}^{p}\frac{\sqrt{\lambda_i\lambda_k}}{\lambda_i-\lambda_k}\mathbb{S}_{ik}(z_0;\mathbf{0})\gamma_k,\quad \alpha_{\Sigma_M}(|z_0|) := \frac{p(p+2)u(|z_0|)}{\mathbb{E}(pu(|Z|)+u'(|Z|))}, \tag{A.3}$$

where $z_0=\Sigma^{-1/2}(x_0-\mu)$, and the expression of $\alpha_{\Sigma_M}(|z_0|)$ is due to Huber [23]. Now taking $\Sigma_M\equiv\Sigma_*$, following [13] we have that

$$V_{12} = \mathbb{E}\left[(\alpha_{\Sigma_*}(|Z|)\mathbb{S}_{12}(Z;\mathbf{0}))^2\right] = \mathbb{E}(\alpha_{\Sigma_*}(|Z|))^2.\mathbb{E}(\mathbb{S}_{12}(Z;\mathbf{0}))^2,$$

using the fact that $|Z|$ and $S(Z;\mathbf{0})$ are independent with $Z\sim\mathbb{F}_Z$. Consequently

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

Proof of Theorem 6. This proof has many algebraic steps, and we sketch the main arguments below. Suppose $\hat{A}=\widehat{\Gamma}^\top\Sigma\widehat{\Gamma}$. 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 ith diagonal element of $\hat{A}$ and $\lambda_i^\dagger$ converges to zero for any $i=\{1,\dots,p\}$. Now notice that from Theorem 4 we have that $\widehat{\Gamma}=\Gamma+R_{n1}$, where the (i,j)th element of the remainder $R_{n1,i,j}$ satisfies $\mathbb{E}R^2_{n1,i,j}=O(n^{-1})$. 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 $\mathbb{E}R^2_{n2,i,j}=O(n^{-1})$. 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 $\lambda^\dagger_{i,j}$ for each group of indices G_j is a consistent estimator of λ_i. The result follows. $\square$

Proof of Proposition 1. Recall that $IF(x_0;\tilde{\Sigma},\mathbb{F})=W^2(x_0)\mathbb{S}(x_0;\mu)-\tilde{\Sigma}$. Now we have, following [13,47], that

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

since $\tilde{\gamma}_i=\gamma_i$ for $1\le i\le p$. $\square$

Proof of Proposition 2. Let $z_0=\Lambda^{-1/2}\Gamma^\top(x_0-\mu)=(z_{01},\dots,z_{0p})^\top$. As a first step, since Σ_* is affine equivariant, we obtain from [13] that $IF(x_0,\Sigma_*,\mathbb{F})=\Sigma_*^{1/2}IF(z_0,\Sigma_*,\mathbb{F}_Z)\Sigma_*^{1/2}$. From Lemma 1 of [22], 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 $IF(x_0,\Sigma_*,\mathbb{F})=\alpha_{\Sigma_*}(|x_0|;\mathbb{F}_Z)\mathbb{S}(x_0;\mu)-\beta_{\Sigma_*}(|x_0|;\mathbb{F}_Z)\Sigma_*$. $\square$

Proof of Proposition 3. After (A.3), this proof follows through similar steps as the above proof of Proposition 1. $\square$

Appendix B. Form of V_W

We elaborate on the form of the asymptotic covariance matrix V_W in Theorem 3. First observe that for F having covariance matrix $\Sigma=\Gamma\Lambda\Gamma^\top$,

$$V_W=(\Gamma\otimes\Gamma)V_{W,\Lambda}(\Gamma\otimes\Gamma)^\top,\tag{B.1}$$

where $V_{W,\Lambda}$ is the covariance matrix of $\mathbb{F}_\Lambda$, the elliptic distribution with mean μ and covariance matrix Λ. Now,

$$\begin{aligned}V_{W,\Lambda}&=\mathbb{E}\left[\operatorname{vec}\left\{\frac{W^2(Z,\mathbb{F}_Z)\Lambda^{1/2}ZZ^\top\Lambda^{1/2}}{Z^\top\Lambda Z}-\tilde{\Lambda}\right\}\operatorname{vec}^\top\left\{\frac{W^2(Z,\mathbb{F}_Z)\Lambda^{1/2}ZZ^\top\Lambda^{1/2}}{Z^\top\Lambda Z}-\tilde{\Lambda}\right\}\right]\\&=\mathbb{E}\left[\operatorname{vec}\left\{W^2(Z,\mathbb{F}_Z)\mathbb{S}(\Lambda^{1/2}z;\mathbf{0})\right\}\operatorname{vec}^\top\left\{W^2(Z,\mathbb{F}_Z)\mathbb{S}(\Lambda^{1/2}Z;\mathbf{0})\right\}\right]-\operatorname{vec}(\tilde{\Lambda})\operatorname{vec}^\top(\tilde{\Lambda})\end{aligned}$$

The matrix $\operatorname{vec}(\tilde{\Lambda})\operatorname{vec}^\top(\tilde{\Lambda})$ consists of elements $\lambda_i\lambda_j$ at (i,j)th position of the (i,j)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 $\mathbb{E}[\sqrt{\lambda_a\lambda_b\lambda_c\lambda_d}Z_aZ_bZ_cZ_d.W^4(Z,\mathbb{F}_Z)/(Z^\top\Lambda Z)^2]$, with $1\le a,b,c,d\le p$. Since $W^4(Z,\mathbb{F}_Z)/(Z^\top\Lambda Z)^2$ is even in Z, which has a spherically symmetric distribution, all such expectations will be 0 unless a,b,c,d are all equal or pairwise equal. Following a similar derivation for generalized sign covariance matrices in [29], we collect the non-zero elements and write the matrix of expectations:

$$(\mathbb{I}_{p^2}+\mathbb{K}_{p,p})\left\{\sum_{a=1}^{p}\sum_{b=1}^{p}\tilde{\gamma}_{ab}(e_ae_a^\top\otimes e_be_b^\top)-\sum_{a=1}^{p}\tilde{\gamma}_{aa}(e_ae_a^\top\otimes e_ae_a^\top)\right\}+\sum_{a=1}^{p}\sum_{b=1}^{p}\tilde{\gamma}_{ab}(e_ae_b^\top\otimes e_ae_b^\top)$$

where $\mathbb{I}_k = (e_1, \ldots, e_k)$, $\mathbb{K}_{m,n} = \sum_{i=1}^m \sum_{j=1}^n \mathbb{J}_{ij} \otimes \mathbb{J}_{ij}^\top$ with $\mathbb{J}_{ij} \in \mathbb{R}^{m\times n}$ having 1 as (i,j)th element and 0 elsewhere, and $\tilde{\gamma}_{mn} = \mathbb{E}[\lambda_m \lambda_n Z_m^2 Z_n^2 . W^4(Z, \mathbb{F}_Z)/(Z^\top \Lambda Z)^2]$; $1 \le m, n \le p$.

. Putting everything together, denote by $\hat{\tilde{\Lambda}}$ the sample version of $\tilde{\Lambda}$, the weighted covariance matrix obtained from $\mathbb{F}_\Lambda$, i.e., $\hat{\tilde{\Lambda}} = \sum_{i=1}^n W_n^2(Z_i, \mathbb{F}_{Z,n}) \mathbb{S}(\Lambda^{1/2} Z_i; \hat{\mu}_n)/n$. Then the different types of elements in the matrix $\hat{\tilde{\Lambda}}$ are as given below $(1 \le a, b, c, d \le p)$:

- Variance of on-diagonal elements:

$$A\mathbb{V}(\sqrt{n}\hat{\tilde{\Lambda}}(a,a)) = \mathbb{E}\left[\frac{W^4(Z, \mathbb{F}_Z)\lambda_a^2 Z_a^4}{(Z^\top \Lambda Z)^2}\right] - \tilde{\lambda}_a^2.$$

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

$$A\mathbb{V}(\sqrt{n}\hat{\tilde{\Lambda}}(a,b)) = \mathbb{E}\left[\frac{W^4(Z, \mathbb{F}_Z)\lambda_a \lambda_b Z_a^2 Z_b^2}{(Z^\top \Lambda Z)^2}\right].$$

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

$$A\mathbb{V}(\sqrt{n}\hat{\tilde{\Lambda}}(a,a), \sqrt{n}\hat{\tilde{\Lambda}}(b,b)) = \mathbb{E}\left[\frac{W^4(Z, \mathbb{F}_Z)\lambda_a \lambda_b Z_a^2 Z_b^2}{(Z^\top \Lambda Z)^2}\right] - \tilde{\lambda}_a \tilde{\lambda}_b.$$

- Covariance of two off-diagonal elements $(a \ne b, c \ne d)$: $A\mathbb{V}(\sqrt{n}\hat{\tilde{\Lambda}}(a,b), \sqrt{n}\hat{\tilde{\Lambda}}(c,d)) = 0$.
- Covariance of one off-diagonal and one on-diagonal element $(a \ne b \ne c)$: $A\mathbb{V}(\sqrt{n}\hat{\tilde{\Lambda}}(a,b), \sqrt{n}\hat{\tilde{\Lambda}}(c,c)) = 0$.

The above give all the elements of $V_{W,\Lambda}$. We plug these into (B.1) to recover V_W.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2022.105013. The supplementary material consists of codes to implement numerical methods, and the real datasets analyzed.

References

[1] K.P. Adragni, R.D. Cook, Sufficient dimension reduction and prediction in regression, Phil. Trans. R. Soc. London A 367 (1906) (2009) 4385–4405.
[2] T. Anderson, An Introduction to Multivariate Statistical Analysis, third edn., Wiley, Hoboken, NJ, 2009, p. 744.
[3] J.L. Bali, G. Boente, D.E. Tyler, J.-L. Wang, Robust functional principal components: A projection-pursuit approach, Ann. Statist. 39 (6) (2011) 2852–2882.
[4] G. Boente, D. Rodriguez, M. Sued, The spatial sign covariance operator: Asymptotic results and applications, J. Multivariate Anal. 170 (2019) 115–128.
[5] G. Boente, M. Salibian-Barrera, *S*-Estimators for functional principal component analysis, J. Amer. Statist. Assoc. 110 (511) (2015) 1100–1111.
[6] H. Cardot, P. Cénac, A. Godichon-Baggioni, Online estimation of the geometric median in Hilbert spaces: Nonasymptotic confidence balls, Ann. Statist. 45 (2) (2017) 591–614.
[7] A. Chakraborty, P. Chaudhuri, The spatial distribution in infinite dimensional spaces and related quantiles and depths, Ann. Statist. 42 (3) (2014) 1203–1231.
[8] S. Chatterjee, A. Bose, Generalized bootstrap for estimating equations, Ann. Statist. 33 (1) (2005) 414–436.
[9] P. Chaudhuri, On a geometric notion of quantiles for multivariate data, J. Amer. Statist. Assoc. 91 (434) (1996) 862–872.
[10] R.D. Cook, B. Li, F. Chiaromonte, Envelope models for parsimonious and efficient multivariate linear regression, Statist. Sinica (2010) 927–960.
[11] R.D. Cook, X. Zhang, Foundations for envelope models and methods, J. Amer. Statist. Assoc. 110 (510) (2015) 599–611.
[12] C. Croux, C. Dehon, A. Yadine, The *k*-step spatial sign covariance matrix, Adv. Data Anal. Classif. 4 (2010) 137–150.
[13] C. Croux, G. Haesbroeck, Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies, Biometrika 87 (3) (2000) 603–618.
[14] C. Croux, E. Ollila, H. Oja, Sign and rank covariance matrices: Statistical properties and application to principal components analysis, in: Y. Dodge (Ed.), Statistical Data Analysis Based on the L1-Norm and Related Methods, Birkhäuser Basel, Basel, 2002, pp. 257–269.
[15] A. Dürre, D. Vogel, D. Tyler, The spatial sign covariance matrix with unknown location, J. Mulivariate Anal. 130 (2014) 107–117.
[16] K.H. Esbensen, S. Schönkopf, T. Midtgaard, Multivariate Data Analysis–in Practice, fifth edn., CAMO Software AS, Trondheim, Germany, 1994.
[17] K.-T. Fang, S. Kotz, K.-W. Ng, Symmetric Multivariate and Related Distributions, first edn., Chapman and Hall/CRC, New York, USA, 1990.
[18] M. Febrero-Bande, M.O.d.l. Fuente, Statistical computing in functional data analysis: The R package fda.usc, J. Stat. Softw. 51 (2012) 1–28.
[19] S.J. Haberman, Concavity and estimation, Ann. Statist. 17 (4) (1989) 1631–1661.
[20] J.B.S. Haldane, Note on the median of a multivariate distribution, Biometrika 35 (3 - 4) (1948) 414–417.
[21] M. Hallin, D. Paindaveine, T. Verdebout, Optimal rank-based testing for principal components, Ann. Statist. 38 (6) (2010) 3245–3299.
[22] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, W.A. Stahel, Robust Statistics: The Approach based on Influence Functions, Vol. 196, John Wiley & Sons, New York, USA, 1986.
[23] P.J. Huber, Robust Statistics, Wiley, New York, USA, 1981.
[24] M. Hubert, P.J. Rousseeuw, K. Vanden Branden, ROBPCA: A new approach to robust principal component analysis, Technometrics 47 (1) (2005) 64–79.
[25] V.I. Koltchinskii, *M*-estimation, convexity and quantiles, Ann. Statist. 25 (2) (1997) 435–477.

[26] R. Liu, J.M. Parelius, K. Singh, Multivariate analysis by data depth: descriptive statistics, graphics and inference, Ann. Statist. 27 (3) (1999) 783–858.
[27] N. Locantore, J.S. Marron, D.G. Simpson, et al., Robust principal component analysis for functional data, Test 8 (1) (1999) 1–73.
[28] M. Maechler, P. Rousseeuw, C. Croux, V. Todorov, A. Ruckstuhl, M. Salibian-Barrera, T. Verbeke, M. Koller, E.L.T. Conceicao, M.A. di Palma, Robustbase: Basic robust statistics, 2022.
[29] A.F. Magyar, D.E. Tyler, The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions, Biometrika 101 (3) (2014) 673–688.
[30] S. Majumdar, S. Chatterjee, Non-convex penalized multitask regression using data depth-based penalties, Stat 7 (e174) (2018) 1–15.
[31] J.I. Marden, Some robust estimates of principal components, Statist. Probab. Lett. 43 (4) (1999) 349–359.
[32] T. Mathieu, Concentration study of M-estimators using the influence function, 2021, arXiv:2104.04416.
[33] J. Miao, A. Ben-Israel, On principal angles between subspaces in $\mathbb{R}^n$, Linear Algebra Appl. 171 (1992) 81–98.
[34] S. Minsker, Geometric median and robust estimation in Banach spaces, Bernoulli 21 (4) (2015) 2308–2335.
[35] J. Möttönen, K. Nordhausen, H. Oja, Asymptotic theory of the spatial median, in: Institute of Mathematical Statistics Collections, Vol. 7, Institute of Mathematical Statistics, 2010, pp. 182–193.
[36] J. Möttönen, H. Oja, Multivariate spatial sign and rank methods, J. Nonparametr. Stat. 5 (2) (1995) 201–213.
[37] J. Nevalainen, D. Larocque, H. Oja, A weighted spatial median for clustered data, Stat. Meth. Appl. 15 (2007a) 355–379.
[38] J. Nevalainen, D. Larocque, H. Oja, On the multivariate spatial median for clustered data, Canad. J. Statist. 35 (2) (2007b) 215–231.
[39] J. Nevalainen, D. Larocque, H. Oja, I. Pörsti, Nonparametric analysis of clustered multivariate data, J. Amer. Statist. Assoc. 105 (490) (2010) 864–872.
[40] W. Niemiro, Asymptotics for M-estimators defined by convex minimization, Ann. Statist. 20 (3) (1992) 1514–1533.
[41] K. Nordhausen, H. Oja, Multivariate L1 statistical methods: The package MNM, J. Stat. Softw. 43 (5) (2011) 1–28.
[42] H. Oja, Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks, Springer Science & Business Media, 2010.
[43] D. Peters, R.H. Randles, A multivariate signed-rank test for the one-sample location problem, J. Amer. Statist. Assoc. 85 (410) (1990) 552–557.
[44] O. Pokotylo, P. Mozharovskyi, R. Dyckerhoff, Depth and depth-based classification with r package ddalpha, J. Stat. Softw. 91 (2019) 1–46.
[45] J. Raymaekers, P. Rousseeuw, A generalized spatial sign covariance matrix, J. Multivariate Anal. 171 (2019) 94–111.
[46] R. Serfling, Depth functions in nonparametric multivariate inference, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 72 (2006) 1.
[47] R. Sibson, Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling, J. R. Stat. Soc. Ser. B Stat. Methodol. (1979) 217–229.
[48] S. Taskinen, I. Koch, H. Oja, Robustifying principal component analysis with spatial sign vectors, Statist. Probab. Lett. 82 (4) (2012) 765–774.
[49] D.E. Tyler, A distribution-free m-estimator of multivariate scatter, Ann. Statist. 15 (1) (1987) 234–251.
[50] S. Visuri, V. Koivunen, H. Oja, Sign and rank covariance matrices, J. Statist. Plann. Inference 91 (2) (2000) 557–575.
[51] L. Wang, B. Peng, R. Li, A high-dimensional nonparametric multivariate test for mean vector, J. Amer. Statist. Assoc. 110 (512) (2015) 1658–1669.
[52] Y. Zuo, Projection-based depth functions and associated medians, Ann. Statist. 31 (2003) 1460–1490.
[53] Y. Zuo, H. Cui, Depth weighted scatter estimators, Ann. Statist. 33 (1) (2005) 381–413.
[54] Y. Zuo, X. He, On the limiting distributions of multivariate depth-based rank sum statistics and related tests, Ann. Statist. 34 (6) (2006) 2879–2896.
[55] Y. Zuo, R. Serfling, General notions of statistical depth function, Ann. Statist. 28 (2) (2000) 461–482.