1 Introduction

???? I will write this section once the rest of the work is complete ????

We consider data $X_1, \ldots, X_n$ from some set $\mathcal{X} \subseteq \mathbb{R}^p$. We consider two functions defined below. First, we consider the *sign function* $S : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^p$ for any p-dimensional vector, defined as

$$S(x; \mu_x) = ||x - \mu_x||^{-1}(x - \mu_x)\mathcal{I}_{\{x \neq 0\}}.$$

This sign function is defined with respect to the *location parameter* $\mu_x \in \mathbb{R}^p$. This is a direct multivariate generalization of the univariate $p = 1$ case of the indicator of whether the point x is to the right, left or at μ_x. This function has been used many times in statistics, see **???? insert several references. ????**

Suppose $\mathcal{F}_p$ is the set of all probability measures on $\mathbb{R}^p$. The second function we consider is the *peripherality function* $P : \mathbb{R}^p \times \mathcal{F}_p \to \mathbb{R}$, which, for every $x \in \mathbb{R}^p$ and every probability measure $F \in \mathcal{F}_p$, satisfies the condition

$$\text{There exists a constant } \mu_F \in \mathbb{R}^p \text{ such that for every } t \in [0,1] \text{ and every } x \in \mathbb{R}^p \text{ we have } P\Big(\mu_F; F\Big) \leq P\Big(\mu_F + t(x - \mu_F); F\Big).$$

That is, for every fixed F, the peripherality function achieves a minimum at μ_F, and is non-decreasing in every direction away from μ_F. If we impose the practical restriction that $\inf_x P(x; F)$ is bounded below, then we may as well impose without loss of generality $P(\mu_F; F) = 0$ and consequently $P(x; F) \geq 0$ for all $x \in \mathbb{R}^p$ and $F \in \mathcal{F}_p$. The peripherality function quantifies whether the point x is near or far from μ_F. We will impose additional conditions on this function as we proceed, but it can be seem immediately that any distance measure between x and μ_F satisfies the bare minimum requirement mentioned above.

In this paper, we demonstrate certain interesting applications arising from composing the sign function and the peripherality function together, to form the *signed-peripherality function* $\kappa(\cdot)$. We define this function with three parameters $\mu_x \in \mathbb{R}^p$, $F \in \mathcal{F}_p$ and $\mu_y \in \mathbb{R}^p$, argument $x \in \mathbb{R}^p$ and range $\mathbb{R}^p$.

More precisely, $\kappa : \mathbb{R}^p \times \mathbb{R}^p \times \mathcal{F}_p \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^p$ is defined as

$$\kappa(x; \mu_x, F, \mu_y) = S(x; \mu_x)P(x; F) + \mu_y.$$

Notice that if we consider $\mu_y = \mu_F = \mu_x$ and take the very simple peripherality function $P(x; F) = ||x - \mu_F||$, we have $\kappa(x; \mu_x, F, \mu_y) \equiv x$ for all choices of parameters μ_x, F, μ_y. Consequently, under this choice of parameters for the κ-transformation, analyzing a dataset $\{X_1, \ldots, X_n\}$ and its κ-transformed version $\{Y_i = \kappa(X_i; \ldots),\ i = 1, \ldots, n\}$ are equivalent. However, in this paper we illustrate how other choices of the peripherality function lead to interesting robustness results. We have deliberately set the location parameters μ_x, μ_F, μ_y to be potentially non-identical, this additional flexibility has some advantage for robust data analysis. In many applications, the value of these three parameters may be identical, which leads to no conflict in our framework.

A whole class of peripherality functions can be defined from *data - depth*, which are center-outward ranking of multivariate data. Data-depths have been extensively used in statistics also, see **???? multiple references. ????**Peripherality functions can be defined as some inverse ranking based on data depth, and the concept of *outlyingness* associated with data depth is essentially same as what we use in this paper. We use the term *peripherality* to keep track of the difference in application contexts and technical assumptions.

In this paper, we consider a few illustrative cases of the use of the κ-transformation. Suppose the data at hand is $X_1, \ldots, X_n$, and we define $Y_i = \kappa(X_i; \mu_X, F, \mu_Y)$ for some choice of parameters μ_X, F, μ_Y. For interpretability and convenience, we assume that $\mathbb{E}S(X_i; \mu_X)P(X_i; F) = 0$, thus $\mathbb{E}Y_i = \mu_Y$. We thus have

$$\begin{aligned}\mathbb{V}Y_i &= \mathbb{E}P(X_i; F)^2 S(X_i; \mu_X)S(X_i; \mu_X)^T \\ &= \mathbb{E}P(X_i; F)^2 ||X_i - \mu_X||^{-2}(X_i - \mu_X)(X_i - \mu_X)^T.\end{aligned}$$

???? Need to include (a) Biman-PC idea for affine equivariance for $p \ll n$, (b) kernel versions as an example of generalization. (c) anything else? ????

2 Develop a robust estimator of variance

Simply do sample variance of the transformed variables $Y_i = \kappa(X_i)$.

Show simulation results like above.

3 Outlier detection

Expand and generalize what you have in the paper already, where I think this is a small example. Refer to a standard method for multivariate outlier detection, and show that such a method used on $Y_i = \kappa(X_i)$ works. I will send you some papers for referencing in this part.

4 Robust principal component analysis

This will be one of the bigger and major sections of the paper, essentially copied and pasted from the previous version. Don't do anything here as of now.

5 Robust PCA and supervised models

In the presence of a vector of univariate responses, say $\mathbf{Y} = (Y_1, Y_2, ..., Y_n)^T$, there is substantial literature devoted to utilizing the subspace generated by the basis of $Cov(\mathbf{X})$ in modelling $E(Y|\mathbf{X})$. This ranges from the simple Principal Components Regression (PCR) to Partial Least Squares (PLS) and Envelope methods [6]. Here we concentrate on robust inference using Sufficient Dimension Reduction (SDR) [1], mainly because it provides a general framework for reducing dimensionality of data directly using top eigenvectors of the covariance matrix of X (albeit in a different manner than PCR) or an appropriate affine transformation of it.

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

$$\mathbf{X}_y = \bar{\boldsymbol{\mu}} + \Gamma \mathbf{v}_y + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{X}_y = \mathbf{X}|Y = y, \bar{\boldsymbol{\mu}} = E\mathbf{X}$, Γ is a $p \times d$ semi-orthogonal basis for $\mathcal{S}_\Gamma$, the spanning subspace of $\{E\mathbf{X}_y - \bar{\boldsymbol{\mu}} | y \in S_Y\}$ (S_y is sample space of Y) and $\mathbf{v}_y = (\Gamma^T\Gamma)^{-1}\Gamma^T(E\mathbf{X}_y - \bar{\boldsymbol{\mu}}) \in \mathbb{R}^d$. The random error term $\boldsymbol{\epsilon}$ follows a multivariate normal distribution with mean $\mathbf{0}_p$ and covariance matrix Δ. This formulation is straightforward to implement when Y is categorical, while for continuous responses, the vector $\mathbf{y}$ is divided into a number of slices.

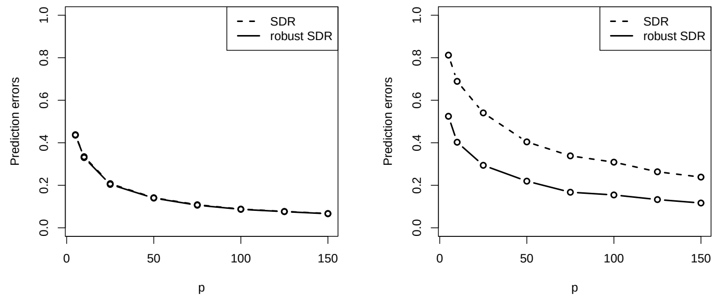
Under this model the minimal sufficient transformation is $R(\mathbf{X}) = \Gamma^T\Delta^{-1}\mathbf{X}$. The simplest case of this model is when $\Delta = \sigma^2 I_p$, for which the maximum likelihood estimator of $R(\mathbf{X})$ turns out to be the first d PCs of $Cov(\mathbf{X})$. Taking $\hat{E}\mathbf{X}_y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

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

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

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

The following simulation study using the same setup as in [1] compares the performance of our robust SDR with the original method with or without the presence of bad leverage points in the covariate matrix X. For a fixed dimension p, we take $n = 200, d = 1$, generate the responses Y as independent standard normal, and the predictors as $\mathbf{X}_Y = \boldsymbol{\gamma}^* v_Y^* + \boldsymbol{\epsilon}$, with $\boldsymbol{\gamma}^*_{p\times 1} = (1, ..., 1)^T, v_Y = Y + Y^2 + Y^3$ and $Var(\boldsymbol{\epsilon}) = 25I_p$. We measure performance of both SDR models by their mean squared prediction error on another set of 200 observations $(Y^*, \mathbf{X}^*)$ generated similarly, and taking the average



(a) (b)

Figure 1: Average prediction errors for two methods of SDR (a) in absence and (b) in presence of outliers

of these errors on 100 such training-test pair of datasets. Finally we repeat the whole setup for different choices of $p = 5, 10, 25, 50, 75, 100, 125, 150$.

Panel (a) of figure 1 compares prediction errors using robust and maximum likelihood SDR estimates when X contains no outliers, and the two methods are virtually indistinguishable. We now introduce outliers in each of the 100 datasets by adding 100 to first $p/5$ coordinates of the first 10 observations in X, and repeat the analysis. Panel (b) of the figure shows that although our robust method performs slightly worse than the case when there were no outliers, it remains more accurate in predicting our of sample observations for all values of p.

6 Robust inference with functional data

This section is to show something beyond the $p \ll n$ setting. Both robust PCA and location testing are important problems for functional data. Abhirup can take charge of this part once we have everything else settled.

The main idea here is to use any decent location estimator to start with (some version of median is fine). Then we may test if the functional location for resting and active state are identical or not.

Also do functional PCA robust version, and then maybe project the data on the first few principal components and then do the 2 sample (or paired sampel) testing again.

(Some technical notations)

We use the approach of [2] for performing robust PCA on functional data. Given data on n functions, say $f_1, f_2, ..., f_n \in L^2[0,1]$, each observed at a set of common design points $\{t_1, ..., t_m\}$, we model each function as a linear combination of p mutually orthogonal B-spline basis functions $\delta_1, ..., \delta_p$. Following this, we map data for each of the functions onto the coordinate system formed by the spline basis:

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

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

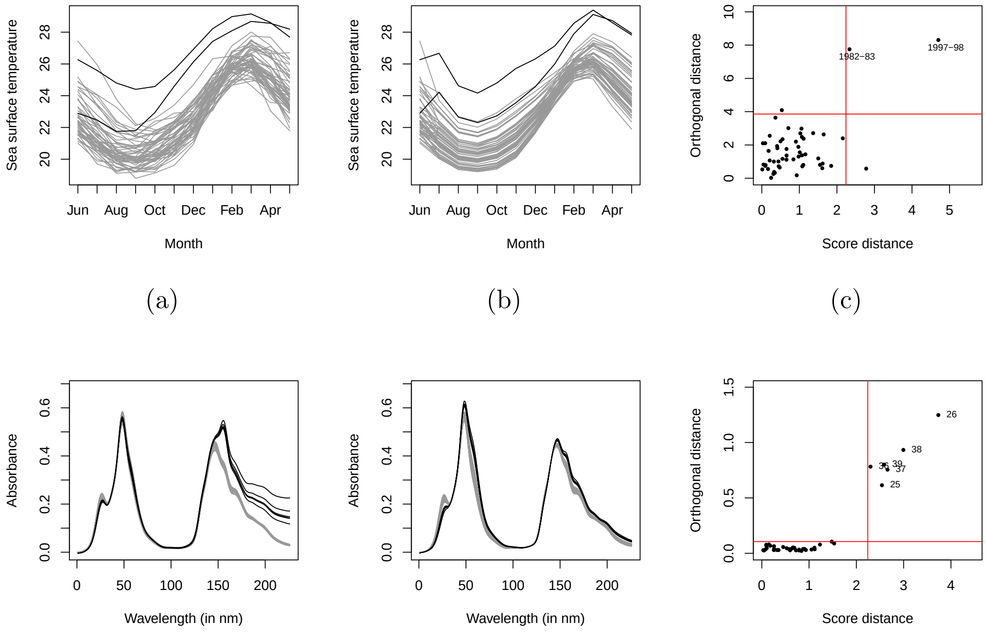
$$\widehat{\tilde{\mathbf{x}}}_i = \tilde{\boldsymbol{\mu}} + \tilde{P}\tilde{\mathbf{s}}_i$$

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

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

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

We first look into the El-Niño data, which is part of a larger dataset on potential factors behind El-Niño oscillations in the tropical pacific available in `http://www.cpc.ncep.noaa.gov/data/indices/`. This records monthly average Sea Surface Temperatures from June 1970 to May 2004, and the yearly oscillations follow more or less the same pattern (see panel a of figure 2). Using a cubic spline basis with knots at alternate months starting in June gives a close approximation of the yearly time series data (panel b), and performing depth-based PCA with $q = 1$ results in two points having their SD and OD larger than cutoff (panel c). These points correspond to



(a) (b) (c)

(d) (e) (f)

Figure 2: Actual sample curves, their spline approximations and diagnostic plots respectively for El-Niño (a-c) and Octane (d-f) datasets

the time periods June 1982 to May 1983 and June 1997 to May 1998 are marked by black curves in panels a and b), and pinpoint the two seasons with strongest El-Niño events.

Our second application is on the Octane data, which consists of 226 variables and 39 observations [9]. Each sample is a gasoline compound with a certain octane number, and has its NIR absorbance spectra measured in 2 nm intervals between 1100 - 1550 nm. There are 6 outliers here: compounds 25, 26 and 36-39, which contain alcohol. We use the same basis structure as the one in El-Niño data here, and again the top robust PC turns out to be sufficient in identifying all 6 outliers (panels d, e and f of figure 2).

7 An example with images

8 A depth-based M estimate of scatter

8.1 Formulation

The DCM is orthogonally equivariant and remains constant only under rotations of the original variables. To construct its affine equivariant counterpart, we need to follow the general framework of M-estimation with data-dependent weights [11]. Specifically, we first implicitly define the Affine-equivariant Depth Covariance Matrix (ADCM) as

$$\Sigma_{Dw} = \frac{1}{Var(\tilde{Z}_1)} E \left[\frac{(\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2 (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T}{(\mathbf{x} - \boldsymbol{\mu})^T \Sigma_{Dw}^{-1} (\mathbf{x} - \boldsymbol{\mu})} \right] \tag{3}$$

Its affine equivariance follows from the fact that the weights $(\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2$ depend only on the standardized quantities $\mathbf{z}$ that depend only on the underlying spherical distribution G. We solve this iteratively by obtaining a sequence of positive definite matrices $\Sigma_{Dw}^{(k)}$ until convergence:

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

To ensure existence and uniqueness of the estimator in 3, let us consider the class of scatter estimators Σ_M that are obtained as solutions of the following equation:

$$E_{\mathbf{Z}_M} \left[u(\|\mathbf{z}_M\|) \frac{\mathbf{z}_M \mathbf{z}_M^T}{\|\mathbf{z}_M\|^2} - v(\|\mathbf{z}_M\|) I_p \right] = 0 \tag{4}$$

with $\mathbf{z}_M = \Sigma_M^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$. Under the following assumptions on the scalar valued functions u and v, the above equation produces a unique solution [11]:

(M1) The function $u(r)/r^2$ is monotone decreasing, and $u(r) > 0$ for $r > 0$;

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

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

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

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

$$Var(\tilde{Z}_1) = Var\left(\tilde{D}_{\mathbf{Z}}(\mathbf{Z}) \frac{Z_1}{\|\mathbf{Z}\|}\right) = Var(\tilde{D}_{\mathbf{Z}}(\mathbf{Z}))Var(S_1(\mathbf{Z})) = \frac{1}{p} Var(\tilde{D}_{\mathbf{Z}}(\mathbf{Z}))$$

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

8.2 Calculation

(Comes right after the calculations discussion of DCM)

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

$$E\left[\tilde{D}_{\mathbf{X}}(\mathbf{x})\frac{\Sigma_{Dw}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}_{Dw})}{\|\Sigma_{Dw}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}_{Dw})\|}\right] = \mathbf{0}_p \tag{5}$$

$$E\left[\frac{(\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2\Sigma_{Dw}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}_{Dw})(\mathbf{x}-\boldsymbol{\mu}_{Dw})^T\Sigma_{Dw}^{-1/2}}{(\mathbf{x}-\boldsymbol{\mu}_{Dw})^T\Sigma_{Dw}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{Dw})}\right] = Var(\tilde{Z}_1)I_p \tag{6}$$

In the framework of (3), for any fixed Σ_M there exists a unique and fixed solution of the location problem $E_{\mathbf{Z}_M}(w(\|\mathbf{z}_M\|\mathbf{z}_M) = \mathbf{0}_p$ under the following condition:

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

This condition is easy to verify for our choice of the weights: $w(\|\mathbf{z}_M\|) = \tilde{D}_{\mathbf{Z}_M}(\mathbf{z}_M)/\|\mathbf{z}_M\|$. Uniqueness of simultaneous fixed point solutions of 5 and 6 is guaranteed when $\mathbf{X}$ has a symmetric distribution [11].

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

1. Start from some initial estimates $(\boldsymbol{\mu}_{Dw}^{(0)}, \Sigma_{Dw,(0)})$. Set $t = 0$;

2. Calculate the standardized observations $\mathbf{z}_i^{(t)} = \Sigma_{Dw,(t)}^{-1/2}(\mathbf{x}_i - \boldsymbol{\mu}_{Dw}^{(t)})$;

3. Update the location estimate:

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

4. Update the scatter estimate:

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

5. Continue until convergence.

8.3 Influence functions

The influence function of any affine equivariant estimate of scatter can be expressed as

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

for scalar valued functions α_C, β_C [10]. Following this, the influence function of an eigenvector $\boldsymbol{\gamma}_{C,i}$ of C is derived:

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

For Tyler's estimate of scatter, we have $\alpha_C(\|\mathbf{z}_0\|) = p + 2$. Considering a more general case, when $C = \Sigma_M$, i.e. the solution to (3), then [11] shows that

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

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

8.4 ARE calculations

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

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

where $ASV(C_{12}, F_0)$ is the asymptotic variance of an off-diagonal element of C when the underlying distribution is F_0. Following [7] this equals

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

	PD-ACM				HSD-ACM			
Distribution	$p=2$	$p=5$	$p=10$	$p=20$	$p=2$	$p=5$	$p=10$	$p=20$
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 1: Table of AREs of the ADCM for different choices of p and data-generating distributions, and two choices of depth functions

again using the fact that $\|\mathbf{Z}\|$ and $\mathbf{S}(\mathbf{Z})$ are independent with $\mathbf{Z} \sim F_0$. When $C = Cov$, i.e. the sample covariance matrix, we have $\alpha_{Cov}(\|\mathbf{z}\|) = \|\mathbf{z}\|^2$. It now follows that

$$ARE(\hat{\boldsymbol{\gamma}}_{\Sigma_M,i}, \hat{\boldsymbol{\gamma}}_{Cov,i}) = \frac{E_{F_0}\alpha_{Cov}(\|\mathbf{z}\|)^2}{E_{F_0}\alpha_C(\|\mathbf{z}\|)^2} = \frac{E_{F_0}\|\mathbf{z}\|^4 . \, [E_{F_0}(pu(\|\mathbf{z}\|) + u'(\|\mathbf{z}\|)\|\mathbf{z}\|)]^2}{p^2(p+2)^2 E_{F_0}(u(\|\mathbf{z}\|))^2} \tag{7}$$

Table 1 considers 6 different elliptic distributions (namely, bivariate t with df = 5, 6, 10, 15, 25 and bivariate normal) and summarizes ARE for first eigenvectors for ADCMs corresponding to projection depth (PD-ACM) and halfspace depth (HSD-ACM). Due to difficulty of analytically obtain the AREs, we calculate them using Monte-Carlo simulation of 10^6 samples and subsequent numerical integration. The ADCM seems to be particularly efficient in lower dimensions for distributions with heavier tails (t_5 and t_6), while for distributions with lighter tails, the AREs increase with data dimension. At higher values of p the ADCM is almost as efficient as the sample covarnace matrix when the data comes from multivariate normal distribution.

9 The robust location problem

Consider the general situation of estimation or testing for the location parameter of an elliptical distribution using weighted sign vectors. For now the only condition we impose on these weights, say $w(.)$, is that they need to be scalar-valued affine equivariant and square-integrable functions of the data, or in other words functions of the norm of the standardized random variable $\mathbf{Z}$. In that sense $w(\mathbf{X})$ can be equivalently written as $f(r)$, with $r = \|\mathbf{Z}\|$. The simplest use of weighted signs here would be to construct a robust alternative to the Hotelling's T^2 test using their sample mean vector and covariance matrix. Formally, this means testing for $H_0 : \boldsymbol{\mu} = \mathbf{0}_p$ vs. $H_1 : \boldsymbol{\mu} \neq \mathbf{0}_p$ based on the test statistic:

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

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

Proposition 9.1. *Consider n random variables $Z = (\mathbf{Z}_1, ..., \mathbf{Z}_n)^T$ distributed independently and identically as $\mathcal{E}(\boldsymbol{\mu}, kI_p, G); k \in \mathbb{R}$, and the class of hypothesis tests defined above. Then, given any $\alpha \in (0, 1)$, local power at $\boldsymbol{\mu} \neq \mathbf{0}_p$ for the level-α test based on $T_{n,w}$ is maximum when $w(\mathbf{Z}_1) = c$, a constant independent of $\mathbf{Z}_1$.*

This essentially means that power-wise the (unweighted) spatial sign test [13] is optimal in the given class of hypothesis tests when the data comes from a spherically symmetric distribution. Our simulations show that this empirically holds for non-spherical but elliptic distributions as well.

9.1 The weighted spatial median

Our weight functions are affine equivariant functions of the data, i.e. they are not affected by the population location parameter $\boldsymbol{\mu}$. This ensures that there exists a unique solution of the following of optimization problem:

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

This can be seen as a generalization of the Fermat-Weber location problem (which has the spatial median [3, 4] as the solution) using data-dependent weights. We call its solution, $\boldsymbol{\mu}_w$, the *weighted spatial median* of F. In a sample setup it is estimated by iteratively solving the equation $\sum_{i=1}^n w(\mathbf{X}_i)\mathbf{S}(\mathbf{X}_i - \hat{\boldsymbol{\mu}}_w)/n = \mathbf{0}_p$.

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

Theorem 9.2. *Let A_w, B_w be two matrices, dependent on the weight function w such that*

$$A_w = E\left[\frac{w(\boldsymbol{\epsilon})}{\|\boldsymbol{\epsilon}\|}\left(1 - \frac{\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T}{\|\boldsymbol{\epsilon}\|^2}\right)\right]; \quad B_w = E\left[\frac{(w(\boldsymbol{\epsilon}))^2\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T}{\|\boldsymbol{\epsilon}\|^2}\right]$$

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

$$\sqrt{n}(\hat{\boldsymbol{\mu}}_w - \boldsymbol{\mu}_w) \rightsquigarrow N_p(\mathbf{0}_p, A_w^{-1}B_wA_w^{-1}) \tag{8}$$

This theorem generalizes equivalent results for the spatial median, and can be proved using the same steps to obtain those results [13]. Setting $w(\boldsymbol{\epsilon}) = 1$ above yields the asymptotic covariance matrix for the spatial median. Following this, the asymptotic relative efficiency (ARE) of $\boldsymbol{\mu}_w$ corresponding to some non-uniform weight function with respect to the spatial median, say $\boldsymbol{\mu}_s$ will be:

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

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

Corollary 9.3. *For the spherical distribution $\mathcal{E}(\boldsymbol{\mu}, kI_p, G); k \in \mathbb{R}, \boldsymbol{\mu} \in \mathbb{R}^p$, we have*

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

9.2 A high-dimensional test of location

It is possible to take an alternative approach to the location testing problem by using the covariance-type U-statistic $C_{n,w} = \sum_{i=1}^n \sum_{j=1}^{i-1} \mathbf{X}_{w,i}^T \mathbf{X}_{w,j}$. This class of test statistics are especially attractive since they are readily generalized to cover high-dimensional situations, i.e. when $p > n$. The Chen and

Qin (CQ) high-dimensional test of location for multivariate normal $\mathbf{X}_i$ [5] is a special case of this test that uses the statistic $C_n = \sum_{i=1}^n \sum_{j=1}^{i-1} \mathbf{X}_i^T \mathbf{X}_j$, and a recent paper ([14], from here on referred to as WPL test) shows that one can improve upon the power of the CQ test for non-gaussian elliptical distributions by using spatial signs $\mathbf{S}(\mathbf{X}_i)$ in place of the actual variables.

Given these, and some mild regularity conditions, the following holds for our generalized test statistic $C_{n,w}$ under H_0 as $n, p \to \infty$:

$$\frac{C_{n,w}}{\sqrt{\frac{n(n-1)}{2}\text{Tr}(B_w^2)}} \rightsquigarrow N(0,1) \tag{9}$$

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

$$\frac{C_{n,w} - \frac{n(n-1)}{2}\boldsymbol{\mu}_0^T A_w^2 \boldsymbol{\mu}_0 (1 + o(1))}{\sqrt{\frac{n(n-1)}{2}\text{Tr}(B_w^2)}} \rightsquigarrow N(0,1) \tag{10}$$

we provide the details behind deriving these two results in the supplementary material, which involve modified regularity conditions and sketches of proofs along the lines of [14].

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

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

when $\Sigma = kI_p$, this again simplifies to $E^2(f(r)/r)/[Ef^2(r).E^2(1/r)]$.

9.2.1 Robust estimation of eigenvalues, and a plug-in estimator of Σ

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

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

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

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

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

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

The number of subgroups used to calculate this median-of-small-variances estimator can be determined following [12]. After this, we construct a consistent plug-in estimator of the population covariance matrix Σ:

Theorem 9.4. *Consider the estimates $\hat{\lambda}_i$ obtained from the above algorithm, and the matrix of eigenvectors $\hat{\Gamma}_D$ estimated using the sample DCM. Define $\hat{\Sigma} = \hat{\Gamma}_D \hat{\Lambda} \hat{\Gamma}_D^T; \hat{\Lambda} = diag(\hat{\lambda}_1, ..., \hat{\lambda}_p)$. Then as $n \to \infty$,*

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

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

(put in 1 or 2 sentences?)

Proof of Proposition 9.1. Under contiguous alternatives $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, the weighted sign test statistic $T_{n,w}$ has mean $E(w(\mathbf{Z})\mathbf{S}(\mathbf{Z}))$. For spherically symmetric $\mathbf{Z}$, $w(\mathbf{Z})$ depends on $\mathbf{Z}$ only through its norm. Since $\|\mathbf{Z}\|$ and $\mathbf{S}(\mathbf{Z})$ are independent, we get $E(w(\mathbf{Z})\mathbf{S}(\mathbf{Z})) = Ew(\mathbf{Z}).E\mathbf{S}(\mathbf{Z})$. The same kind of decomposition holds for $Cov(w(\mathbf{Z})\mathbf{S}(\mathbf{Z}))$.

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

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

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

Proof of Theorem 9.4. We are going to prove the following:

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

2. $\|\hat{\Lambda} - \Lambda\|_F \xrightarrow{P} 0$

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

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

$$|\tilde{\lambda}_i - \lambda_i| \xrightarrow{P} 0 \qquad (11)$$

using Theorem 3.1 of [12], with $\mu = \lambda_i$. Now $\hat{\lambda}_i = \text{med}_j(Var(X^T_{G_j}\hat{\boldsymbol{\gamma}}_{D,i}))$ and $\tilde{\lambda}_i = \text{med}_j(Var(X^T_{G_j}\boldsymbol{\gamma}_i))$, so that

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

using Cauchy-Schwarz inequality. Combining the facts $\|\hat{\boldsymbol{\gamma}}_{D,i} - \boldsymbol{\gamma}_i\| = O_P(1/\sqrt{n})$ and $\text{med}_j[\text{Tr}(Cov(X_{G_j}))] \xrightarrow{P} \text{Tr}(\Sigma)$ [12] with (11), we get the needed.

□

Sketch of proofs for statements regarding $C_{n,w}$. A first step to obtain asymptotic normality for the high-dimensional location test statistic $C_{n,w}$ is obtaining an equivalent result of Lemma 2.1 in [14]:

Lemma 9.5. *Under the conditions*
(C1) $Tr(\Sigma^4) = o(Tr^2(\Sigma^2))$,
(C2) $Tr^4(\Sigma)/Tr^2(\Sigma^2) \exp[-Tr^2(\Sigma)/128p\lambda^2_{\max}(\Sigma)] = o(1)$

when H_0 is true we have

$$\begin{aligned} E[(\boldsymbol{\epsilon}_{w1}^T \boldsymbol{\epsilon}_{w2})^4] &= O(1)E^2[(\boldsymbol{\epsilon}_{w1}^T \boldsymbol{\epsilon}_{w2})^2] & (12)\\ E[(\boldsymbol{\epsilon}_{w1}^T B_w \boldsymbol{\epsilon}_{w1})^2] &= O(1)E^2[(\boldsymbol{\epsilon}_{w1}^T B_w \boldsymbol{\epsilon}_{w1})^2] & (13)\\ E[(\boldsymbol{\epsilon}_{w1}^T B_w \boldsymbol{\epsilon}_{w2})^2] &= o(1)E^2[(\boldsymbol{\epsilon}_{w1}^T B_w \boldsymbol{\epsilon}_{w1})^2] & (14) \end{aligned}$$

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

A proof of this lemma is derived using results in section 3 of [8], noticing that any-scalar valued 1-Lipschitz function of $\boldsymbol{\epsilon}_w$ is a M_w-Lipschitz function of $\mathbf{S}(\boldsymbol{\epsilon})$, with $M_w = \sup_{\boldsymbol{\epsilon}} w(\boldsymbol{\epsilon})$. Same steps as in the proof of Theorem 2.2 in [14] follow now, using the lemma above in place of Lemma 2.1 therein, to establish asymptotic normality of $C_{n,w}$ under H_0.

To derive the asymptotic distribution under contiguous alternatives we need the conditions (C3)-(C6) in [14], as well as slightly modified versions of Lemmas A.4 and A.5:

Lemma 9.6. *Given that condition (C3) holds, we have $\lambda_{\max}(B_w) \leq 2\frac{\lambda_{\max}}{Tr(\Sigma)}(1+o(1))$.*

Lemma 9.7. *Define $D_w = E\left[\frac{w^2(\boldsymbol{\epsilon})}{\|\boldsymbol{\epsilon}\|^2}(I_p - \mathbf{S}(\boldsymbol{\epsilon})\mathbf{S}(\boldsymbol{\epsilon})^T)\right]$. Then $\lambda_{\max}(A_w) \leq E(w(\boldsymbol{\epsilon})/\|\boldsymbol{\epsilon}\|)$ and $\lambda_{\max}(D_w) \leq E(w(\boldsymbol{\epsilon})/\|\boldsymbol{\epsilon}\|)^2$. Further, if (C3) and (C4) hold then $\lambda_{\min}(A_w) \geq E(w(\boldsymbol{\epsilon})/\|\boldsymbol{\epsilon}\|)(1+o(1))/\sqrt{3}$.*

The proof now exactly follows steps in the proof of theorem 2.3 in [14], replacing vector signs by weighted signs, using the fact that $w(\boldsymbol{\epsilon})$ is bounded above by M_w while applying conditions (C5)-(C6) and lemmas A.1, A.2, A.3, and finally using the above two lemmas in place of lemmas A.4 and A.5 respectively.

□

References

[1] K. P. Adragni and R. D. Cook. Sufficient dimension reduction and prediction in regression. *Phil. Trans. R. Soc. A*, 367:4385–4405, 2009.

[2] G. Boente and M. Salibian-Barrera. S-Estimators for Functional Principal Component Analysis. *J. Amer. Statist. Assoc.*, 110:1100–1111, 2015.

[3] B.M. Brown. Statistical Use of the Spatial Median. *J. Royal Statist. Soc. B*, 45:25–30, 1983.

[4] P. Chaudhuri. On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.*, 91:862–872, 1996.

[5] S. X. Chen and Y. L. Qin. A Two-sample Test for High-dimensional Data with Application to Gene-Set Testing. *Ann. Statist.*, 38:808–835, 2010.

[6] R. D. Cook, B. Li, and F. Chiaromonte. Envelope models for parsimonious and efficient multivariate linear regression. *Biometrika*, 20:927–1010, 2010.

[7] C. Croux and G. Haesbroeck. Principal Component Analysis based on Robust Estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies. *Biometrika*, 87:603–618, 2000.

[8] N. El Karoui. Concentration of Measure and Spectra of Random Matrices: with Applications to Correlation Matrices, Elliptical Distributions and Beyond. *Ann. Applied Probab.*, 19:2362–2405, 2009.

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

[10] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Staehl. *Robust Statistics: The Approach Based on Influence Functions.* Wiley, 1986.

[11] P. J. Huber. *Robust Statistics.* Wiley series in probability and mathematical statistics. Wiley, 1981.

[12] S. Minsker. Geometric median and robust estimation in Banach spaces. *Bernoulli*, 21:2308–2335, 2015.

[13] H. Oja. *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks.* Lecture Notes in Statistics. Springer, 2010.

[14] L. Wang, B. Peng, and R. Li. A High-Dimensional Nonparametric Multivariate Test for Mean Vector. *J. Amer. Statist. Assoc.*, 110:1658–1669, 2015.