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

There exists a constant $\mu_F \in \mathbb{R}^p$ such that for every $t \in [0, 1]$ and every $x \in \mathbb{R}^p$ 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 Hawkins-regression

Consider the $p+1$ dimensional data $\{(\mathbf{X}_i\ Y_i)^T \in \mathbb{R}^{p+1},\ i = 1, \ldots, n\}$. Do the robust PCA on this, and get all the eigenvectors. The typical eigenvector looks like $(\gamma\ \gamma_0) \in \mathbb{R}^{p+1}$. Under the condition that the joint distribution of $(\mathbf{X}\ Y)^T$ is absolutely continuous, we have $\gamma_0 \neq 0$ almost surely. We implement the following for any estimated eigenvector where $|\gamma_0| > \epsilon$ for some fixed $\epsilon > 0$: multiply the eigenvector by $-\gamma_0^{-1}$, thus consider $(-\gamma_0^{-1}\gamma\ -1) \in \mathbb{R}^{p+1}$, which is a "Hawkins-regression". The variance of this linear combination of variables is given by the corresponding eigenvalue times γ_0^{-2}, which may be small or large in the $p+1$ eigenvectors, so the order in which eigenvalues/vectors appear in PCA may not be maintained here.

Also look up what Hawkins had originally proposed, and implement those.

Do a simple simulation, building on the simulation you have for the previous section.

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.

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 [3]. Specifically, we first implicitly define the Affine-equivariant Depth Covariance Matrix (ADCM) as

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

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 circular 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 1, let us consider the class of scatter estimators Σ_M that are obtained as solutions of the following equation:

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

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 [3]:

(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:

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

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 $(\boldsymbol{\mu}_{Dw}, \Sigma_{Dw})$:

$$E\left[\tilde{D}_{\mathbf{X}}(\mathbf{x})\frac{\Sigma_{Dw}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}_{Dw})}{\|\Sigma_{Dw}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}_{Dw})\|}\right] = \mathbf{0}_p \quad (3)$$

$$E\left[\frac{(\tilde{D}_{\mathbf{X}}(\mathbf{x}))^2\Sigma_{Dw}^{-1/2}(\mathbf{x}-\boldsymbol{\mu}_{Dw})(\mathbf{x}-\boldsymbol{\mu}_{Dw})^T\Sigma_{Dw}^{-1/2}}{(\mathbf{x}-\boldsymbol{\mu}_{Dw})^T\Sigma_{Dw}^{-1}(\mathbf{x}-\boldsymbol{\mu}_{Dw})}\right] = Var(\tilde{Z}_1)I_p \quad (4)$$

In the framework of (1), 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 3 and 4 is guaranteed when $\mathbf{X}$ has a symmetric distribution [3].

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 $(\boldsymbol{\mu}^{(0)}_{Dw}, \Sigma_{Dw,(0)})$. Set $t = 0$;
2. Calculate the standardized observations $\mathbf{z}^{(t)}_i = \Sigma^{-1/2}_{Dw,(t)}(\mathbf{x}_i - \boldsymbol{\mu}^{(t)}_{Dw})$;

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 [2]. 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 (1), then [3] 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 [1] this equals

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

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

$$ARE(\hat{\boldsymbol{\gamma}}_{\Sigma_M,i}, \hat{\boldsymbol{\gamma}}_{Cov,i}; F) = \frac{E_{F_0}\alpha_{Cov}(\|\mathbf{z}\|)^2}{E_{F_0}\alpha_C(\|\mathbf{z}\|)^2} = \frac{E_{F_0}\|\mathbf{z}\|^4 . \left[E_{F_0}(pu\|\mathbf{z}\| + u'(\|\mathbf{z}\|)\|\mathbf{z}\|)\right]^2}{E_{F_0}(u(\|\mathbf{z}\|))^2} \tag{5}$$

The table below below 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 the three choices of depths we consider, and compares them with those of from its unweighted counterpart, i.e. Tyler's scatter matrix. Due to difficulty of analytically obtain the AREs, we calculate them using Monte-Carlo simulation of 10^6 samples and subsequent numerical integration.

(Table of ARE to be added)

References

[1] 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.

[2] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Staehl.

Robust Statistics: The Approach Based on Influence Functions. Wiley, 1986.

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