Robustness of the Affine Equivariant Scatter Estimator Based on the Spatial Rank Covariance Matrix

Kai Yu [1]

Xin Dang*[2]

Department of Mathematics

and

Yixin Chen[3]

Department of Computer and Information Science

March 31, 2013

[1]Department of Mathematics, University of Mississippi, University, 38677, USA. Email: kyu1@olemiss.edu

[2]Corresponding author, Department of Mathematics, University of Mississippi, 315 Hume Hall, University, MS 38677, USA. Telephone: (662)915-7409. Fax: (662)915-2361. Email: xdang@olemiss.edu. Webpage: http://olemiss.edu/∼xdang

[3]Department of Computer and Information Science, University of Mississippi, 201 Weir Hall, University, MS 38677, USA. Telephone: (662)915-7438. Fax: (662)915-5623. Email: ychen@cs.olemiss.edu. Webpage: http://cs.olemiss.edu/∼ychen

Abstract

Visuri, Oja and Koivunen (2000) proposed a technique for robust covariance matrix estimation based on different notions of multivariate sign and rank. Among them, the spatial rank based covariance matrix estimator that utilizes a robust scale estimator is especially appealing due to its high robustness, computational ease and good efficiency. Also, it is orthogonally equivariant under any distribution and affinely equivariant under elliptically symmetric distributions. In this paper, we study robustness properties of the estimator with respective to two measures: breakdown point and influence function. More specifically, the upper bound of the finite sample breakdown point can be achieved by a proper choice of univariate robust scale estimator. The influence functions for eigenvalues and eigenvectors of the estimator are derived. They are found to be bounded under some assumptions. Moreover, finite sample efficiency comparisons to popular robust MCD, M and S estimators are reported.

Key words and phrases: Scatter estimator; spatial rank; breakdown point; influence function; affine equivariance; spectral decomposition.

1 Introduction

Classical multivariate statistical inference methods including multivariate analysis of variance, principle component analysis, factor analysis, canonical correlation analysis are based on sample covariance matrix. Those moment-based techniques are optimal (most efficient) under the normality distributional assumption. They are, however, extremely sensitive to outlying observations, and poor in the efficiency for heavy-tailed distributions. A straightforward treatment is to replace the sample covariance matrix with a robust one. A variety of robust estimates of scatter matrix have been proposed. Among them are M-estimates (Maronna, 1976), S-estimates (Davies, 1987), MCD-estimates (Rousseeuw, 1985), depth(projection)-based estimates (Zuo and Cui, 2005; Zhou and Dang, 2010), as well as sign and rank covariance estimates (Visuri, Oja and Koivunen, 2000). In this paper, we focus on one of the last type.

Although component-wise sign and rank covariance matrices are simple and intuitive, they are less interesting because they are neither rotation nor scale equivariant, hence the geometry of the data cloud is not presented. Also, they lack the efficiency at the normal model. A series of papers investigate covariance matrices based on Oja sign and rank functions. For example, Ollila, Oja and Croux (2003) explored the influence function and efficiency of the Oja sign covariance matrix. With its application to a regression model was dealt by Ollila, Oja and Hettmansperger (2002). Ollila, Croux and Oja (2004) treated the Oja rank covariance matrix. In Ollila, Oja and Koivunen (2003), the regression coefficient estimators based on Oja rank covariance matrix are considered. The advantages of those methods are their affine equivariance, high efficiency in the multi-normal case and superior performance in heavy-tailed distributions. They are not robust in the usual sense. The influence function is unbounded, and the breakdown point is 0. Moreover, the computation cost is prohibitively high for large sample size in high dimensions.

Spatial sign and rank covariance matrices and methods (Oja, 2010) become more attractive due to its computational ease, high robustness and good efficiency. The major drawback of these procedures is that they are only orthogonally equivariant. In order to gain full affine equivariance property, one approach is to utilize transformation-retransformation(TR) technique, which serves as standardization of multivariate data. More details can be found in Serfling (2010). The well-known scatter functional of Tyler (1987) is a TR version of spatial sign covariance matrix. In the same spirit, Dümbgen (1998) considered symmetrized TR spatial sign covariance matrix. Oja and Randles (2004) constructed nonparametric tests based on TR spatial rank covariance matrix. Indeed, the above mentioned TR scatter functionals are of the form of M-functionals, hence inherit features of M-functionals. The estimators can

be easily computed by the usual iterative algorithm. The robustness is disappointingly low in high dimensions. Dümbgen and Tyler (2005) studied the breakdown properties of those estimators: the breakdown point for Tyler's estimator is $1/d$ and $1-\sqrt{1-1/d} \in (1/2d, 1/d)$ for Dümbgen's estimator, where d is the dimension. K-step versions of the above estimators are studied in Croux, Dehon and Yadine (2010) and Taskinen, Sirkiä and Oja (2010). A related approach is a spatial trimming technique used by Mazumder and Serfling (2010), in which a scatter estimator is obtained based on the trimmed data with the TR version spatial outlyingness less than some threshold. The robustness depends on the value of threshold, the trimming parameter. With authors' suggestion on the parameter, the breakdown point is $1/(d+2)$.

In this paper, we use different approach to obtain equivariance property of spatial sign and rank covariance matrices under elliptical models without sacrifice of robustness. The basic idea is to take advantage of the fact that the spatial sign and rank functions preserve directional information but lose some measure on distance. Consequently, eigenvectors of the spatial sign and rank covariance matrices are able to capture principle components (orientation) of a data cloud (or underlying distribution), but eigenvalues no longer reflect variation on those directions even for the rank covariance matrix in which some distance information is present in spatial rank function. This is the result of [16] (Marden, 1999). Our strategy is to replace each eigenvalue with an univariate scale estimator on the corresponding direction such that it depicts the proper variability. This is also the approach that Visuri, Oja and Koivunen (2000) take. For consideration of robustness, the univariate scale functional must be robust, e.g. MAD (median of absolute deviation). We favor spatial rank covariance matrix over spatial sign covariance matrix because it is more efficient and there is no initial location estimator needed for computing rank vectors. We call the resulting covariance matrix the modified spatial rank covariance matrix (MRCM).

The main contributions of this paper are that we study the robustness properties of MRCM by the breakdown point and influence function. The finite sample breakdown point is obtained. We show that the finite sample breakdown point can attain the upper bound by a proper choice of univariate scale estimator. The influence functions of eigenvalues and eigenvectors of the covariance matrix are derived and found to be bounded.

The paper is organized as follows. Section 2 gives the definition of the scatter estimator based on spatial rank covariance matrix and its affine equivariance property under elliptical models. Section 3 contains the robustness results in terms of the finite breakdown point and influence function. Simulation results on finite sample efficiency comparison are presented in Section 4. Section 5 contains further discussion and remarks. All formal proofs are presented in the Appendix.

2 Spatial rank covariance matrix and its equivariant version

2.1 Spatial rank function and spatial rank covariance matrix

If $\boldsymbol{Z}$ is a random variable from a distribution with cdf F in $\mathbb{R}^d$, the expected Euclidean distance from $\boldsymbol{z}$ to $\boldsymbol{Z}$ is $D(\boldsymbol{z}, F) = \mathbb{E}_F\|\boldsymbol{z} - \boldsymbol{Z}\|$, and the spatial median of F minimizes the criterion function D. The multivariate centered *spatial rank function* is defined as the gradient of D:

$$\boldsymbol{R}(\boldsymbol{z}, F) = \nabla_{\boldsymbol{z}} D(\boldsymbol{z}, F) = \mathbb{E}_F \frac{\boldsymbol{z} - \boldsymbol{Z}}{\|\boldsymbol{z} - \boldsymbol{Z}\|} = \mathbb{E}_F\{\boldsymbol{S}(\boldsymbol{z} - \boldsymbol{Z})\},$$

where $\boldsymbol{S}(\boldsymbol{x}) = \boldsymbol{x}/\|\boldsymbol{x}\|$ ($\boldsymbol{S}(\boldsymbol{0}) = \boldsymbol{0}$) is the *vector (spatial) sign function* in $\mathbb{R}^d$. The spatial rank function is the *expected direction* to $\boldsymbol{z}$ from $\boldsymbol{Z}$. We call it centered because the random ranks are centered at $\boldsymbol{0}$, that is, $\mathbb{E}_F\boldsymbol{R}(\boldsymbol{Z}, F) = \boldsymbol{0}$. In the univariate case, $D(z, F) = \mathbb{E}_F|z - Z|$ leads to the univariate centered rank function $R(z, F) = \mathbb{E}_F\text{sign}(z - Z) = 2F(z) - 1 \in [-1, 1]$. The univariate median has centered rank 0.

The spatial rank function has many nice properties. The rank function $\boldsymbol{R}(\boldsymbol{z}, F)$ characterizes the distribution F (up to a location shift)(See Koltchinskii (1997)), which means that if we know the rank function, we know the distribution (up to a location shift). Under very weak assumptions on F, $\boldsymbol{R}(\boldsymbol{z}, F)$ is a one-to-one mapping from $\boldsymbol{z} \in \mathbb{R}^d$ to a vector inside the unit ball with the magnitude $\|\boldsymbol{R}(\boldsymbol{z}, F)\| \in [0, 1]$ being a measure of outlyingness of $\boldsymbol{z}$ with respective to F. The inverse mapping is called spatial quantile function. See Serfling (2010) for further discussions and references.

The corresponding *spatial rank covariance matrix* (RCM), denoted as $\Sigma_R(F)$ or $\Sigma_R(\boldsymbol{Z})$, is defined as

$$\Sigma_R(F) = \Sigma_R(\boldsymbol{Z}) = \mathbb{E}_F\{\boldsymbol{R}(\boldsymbol{Z}, F)\boldsymbol{R}^T(\boldsymbol{Z}, F)\}.$$

Since the rank is centered, the RCM is nothing more than the covariance matrix of the rank of $\boldsymbol{Z}$, which is $\text{Cov}(\boldsymbol{R}(\boldsymbol{Z}, F))$. Recall that $\|\boldsymbol{R}(\boldsymbol{Z}, F)\| \leq 1$, hence the assumptions on F for existence of Σ_R are much weaker than the ones for existence of $\text{Cov}(\boldsymbol{Z})$.

Note that with different choices of D, different notions of rank function and rank covariance matrix can be defined accordingly. For instance, $D(\boldsymbol{z}, F) = \mathbb{E}_F\|\boldsymbol{z} - \boldsymbol{Z}\|_1$ ($\|\boldsymbol{a}\|_1 = |a_1| + ..., |a_d|$) leads to component-wise rank function and the corresponding covariance matrix. For D being the expected volume of the simplex formed by $\boldsymbol{z}$ and d independent random vectors $\boldsymbol{Z}_1, ..., \boldsymbol{Z}_d$ from F, it yields Oja rank function and Oja rank covariance matrix (See Ollila *et al.* 2004). With D being expected squared Euclidean distance from $\boldsymbol{z}$ to $\boldsymbol{Z}$, the usual covariance matrix is recovered.

When a random sample $\mathbb{Z} = \{\boldsymbol{z}_1, ..., \boldsymbol{z}_n\}$ from F is available, the sample counterparts are obtained by replacing F with the empirical distribution F_n, that is, the sample spatial rank function is

$$\boldsymbol{R}(\boldsymbol{z}, F_n) = \boldsymbol{R}(\boldsymbol{z}, \mathbb{Z}) = \frac{1}{n}\sum_i \boldsymbol{S}(\boldsymbol{z} - \boldsymbol{z}_i),$$

and the sample spatial rank covariance matrix is

$$\Sigma_R(F_n) = \Sigma_R(\mathbb{Z}) = \frac{1}{n}\sum_i \boldsymbol{R}(\boldsymbol{z}_i, F_n)\boldsymbol{R}^T(\boldsymbol{z}_i, F_n) = \frac{1}{n(n-1)^2}\sum_{j,k\neq i} \boldsymbol{S}(\boldsymbol{z}_i - \boldsymbol{z}_j)\boldsymbol{S}^T(\boldsymbol{z}_i - \boldsymbol{z}_k). \quad (1)$$

From the last term in the (1), it seems that the computation complexity of $\Sigma_R(F_n)$ is $O(n^3)$. However, utilizing the middle term to compute $\Sigma_R(F_n)$ only needs $O(n^2)$ computing time. That is, we compute all rank functions $R(\boldsymbol{z}_i, F_n)$ first with $O(n^2)$ time and then calculate the matrix with $O(dn)$ time, where d is the dimension and usually much less than n.

It is worthwhile to remark that $\Sigma_R(F_n)$ is asymptotically equivalent to a matrix-valued U-statistic with the kernel of size 3, hence convergence of the sample version to the population one can be established by the practice of U-statistic theory. Sirkiä et al (2009) developed the asymptotical properties for the RCM under spherically symmetrical distributions and tests of shape based on RCM was proposed. Under mild assumptions, Möttönen *et al.* (1997) proved that the sample ranks converge uniformly in probability to the theoretical functions.

As any spatial procedure, spatial signs and spatial ranks are orthogonally equivariant in the sense that for any $d \times d$ orthogonal matrix O ($O^T = O^{-1}$), d-dimensional vector $\boldsymbol{b}$ and nonzero scaler c, letting $\boldsymbol{z}^* = cO\boldsymbol{z} + \boldsymbol{b}$ and $\boldsymbol{Z}^* = cO\boldsymbol{Z} + \boldsymbol{b}$ with the distribution $F_{\mathbf{Z}^*}$,

$$\boldsymbol{S}(\boldsymbol{z}^*) = \text{sign}(c)O\boldsymbol{S}(\boldsymbol{z}), \quad \text{and} \quad \boldsymbol{R}(\boldsymbol{z}^*, F_{\mathbf{Z}^*}) = \text{sign}(c)O\boldsymbol{R}(\boldsymbol{z}, F_{\mathbf{Z}}).$$

Therefore, $\Sigma_R(F)$ is orthogonally equivariant, meaning that,

$$\Sigma_R(\boldsymbol{Z}^*) = c^2 O\Sigma_R(\boldsymbol{Z})O^T.$$

Orthogonal equivariance ensures that under rotation, translation and homogeneous scale change, the quantities are transformed accordingly. However, it does not allow heterogeneous scale changes. The above equations do not hold for a general $d \times d$ nonsingular matrix A. Hence, they are not fully affine equivariant. In order to achieve full affine equivariance, we confine our attentions to one class of distributions $\mathcal{F}$ and modify RCM such that it is affine equivariant for every distribution $F \in \mathcal{F}$. Specifically, we consider $\mathcal{F}$ to be the set of all elliptically symmetric distributions.

2.2 Elliptical models

A distribution is called *elliptical* if it has a density of the form

$$f(\boldsymbol{z};\boldsymbol{\mu},\Sigma) = \{\det(\Sigma)\}^{-1/2} h\{(\boldsymbol{z}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{z}-\boldsymbol{\mu})\}, \tag{2}$$

for some $\boldsymbol{\mu} \in \mathbb{R}^d$, a positive definite symmetric $d \times d$ matrix Σ, and a nonnegative function h with $\int_0^\infty t^{d/2-1}h(t)dt < \infty$ independent to $\boldsymbol{\mu}$ and Σ. The parameter $\boldsymbol{\mu}$ is the symmetric center of the distribution and it equals the first moment $\mathbb{E}\boldsymbol{Z}$ if it exists, while the scatter parameter Σ is proportional to the covariance matrix $\mathrm{Cov}(\boldsymbol{Z})$ when it exists. In the case of the multivariate t distribution with degrees of freedom $\nu > 0$, h at (2) is of the form $h(t) = c(\nu,d)(1+t/\nu)^{-(d+\nu)/2}$, where $c(\nu,d)$ is the normalization constant. For $\nu > 2$, the covariance matrix $\mathrm{Cov}(\boldsymbol{Z}) = \nu/(\nu-2)\Sigma$. For $\nu = 1$, it is called d-variate Cauchy distribution which has very heavy tails so that even the first moment doesn't exist. When $\nu \to \infty$, it yields the normal distribution with $h(t) = (2\pi)^{-d/2}e^{-t/2}$, in which all moments exist and $\mathrm{Cov}(\boldsymbol{Z}) = \Sigma$. If $\Sigma = I_{d\times d}$ (the identity matrix), the distribution is spherically symmetry.

It is well known that if the distribution of $\boldsymbol{Z}$ is elliptically symmetric with parameters $\boldsymbol{\mu}, \Sigma$ and corresponding function h, and let $\boldsymbol{Z}^* = A\boldsymbol{Z} + \boldsymbol{b}$ be an affine transformation with A nonsingular and $\boldsymbol{b} \in \mathbb{R}^d$, $\boldsymbol{Z}^*$ is elliptically distributed with the same h and the parameters being $A\boldsymbol{\mu} + \boldsymbol{b}$ and $A\Sigma A^T$. Hence, affine equivariance property for scatter functionals or estimators is of fundamental importance under elliptical symmetry. The key result from Marden (1999) provides a way to modify spatial rank covariance matrix such that it is affine equivariant under elliptical models. In fact, we may remove the assumption of Marden (1999) that requires the existence of the covariance matrix. The result and proofs are still valid with only difference being the interpretation of eigenvalues.

Lemma 2.1 *If $\boldsymbol{Z}$ is elliptically distributed from F with the scatter parameter Σ having the spectral decomposition $V\Lambda V^T$, then $\Sigma_R(F) = V\Lambda_R V^T$, where Λ_R is the diagonal matrix of eigenvalues of Σ_R.*

The main consequence is that the same orthogonal matrix V diagonalizes Σ and Σ_R. In other words, spatial rank covariance matrix Σ_R has the same eigenvectors as Σ. The eigenvalue λ of Σ is some spread measure of $\boldsymbol{v}^T\boldsymbol{Z}$, where $\boldsymbol{v}$ is the corresponding eigenvector of Σ. If $\mathrm{Cov}(\boldsymbol{Z})$ exists, eigenvalue λ is proportional to the variance of $\boldsymbol{v}^T\boldsymbol{Z}$. Therefore, it is natural to estimate V by the eigenvectors of $\Sigma_R(F_n)$, then estimate Λ by some scatter estimates on the univariate projected data.

2.3 Modified RCM

The way we construct the modified spatial rank covariance matrix MRCM (sample version), denoted as $\tilde{\Sigma}(F_n)$ or $\tilde{\Sigma}(\mathbb{Z})$, is as follows.

1 Compute the sample spatial covariance matrix $\Sigma_R(F_n)$ using (1).

2 Construct eigenvector estimates. Find the corresponding eigenvector estimates $\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_d$ by spectral decomposition of $\Sigma_R(F_n)$, denoted by the matrix S. That is, $S = [\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_d]$

3 Find the scale estimates (eigenvalues, principal values) of $\mathbb{Z}$ on directions of $\mathbf{s}_i$'s, by using univariate robust scale estimate σ, say MAD. Let $\hat{\lambda}_i = \{\sigma(\mathbf{s}_i^T\mathbb{Z})\}^2$ and denote $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, ..., \hat{\lambda}_d)$.

4 The scatter estimate is $\tilde{\Sigma}(F_n) = S\hat{\Lambda}S^T$.

Letting $F_{\boldsymbol{v}}$ be the distribution of $\boldsymbol{v}^T\boldsymbol{Z}$, we may obtain the population version of MRCM by finding eigenvector $\boldsymbol{v}_i$ of $\Sigma_R(F)$ and $\tilde{\lambda}_i = \sigma^2(F_{\boldsymbol{v}_i})$ for $i = 1, ..., d$, then $\tilde{\Sigma}(F) = V\tilde{\Lambda}V^T$, where $V = [\boldsymbol{v}_1, ..., \boldsymbol{v}_d]$ and $\tilde{\Lambda} = \text{diag}(\tilde{\lambda}_1, .., \tilde{\lambda}_d)$.

Theorem 2.1 *Under an elliptical model F with scatter parameter Σ, $\tilde{\Sigma}(F)$ is an affine equivariant scatter functional.*

An immediate consequence of Lemma 2.1 and the construction way of $\tilde{\Sigma}$ result that $\tilde{\Sigma}(F)$ is proportional to Σ, that is,

$$\tilde{\Sigma}(F) = c(h_F, \sigma)\Sigma, \tag{3}$$

with $c(h_F, \sigma)$ being a constant only depending on h function of the distribution F and the choice of univariate scale functional σ. For example, taking $\sigma = MAD$, if F is the multivariate normal distribution, then $c(h_F, \sigma) = \{\Phi^{-1}(3/4)\}^2 \approx 0.455$, where Φ^{-1} is the quantile function of the standard normal distribution. If F is a t-distribution with $\nu > 2$, $c(h_F, \sigma) \approx 0.455\nu/(\nu - 2)$. If F is the Cauchy distribution, $c(h_F, \sigma) = 1$. Later on, for consideration of highest possible robustness in terms of the breakdown point, we will use a variation of MAD, MAD_k. Its definition and some discussion are given in the next section.

The proof of affine equivariance of $\tilde{\Sigma}$ is straightforward. Let $\boldsymbol{Z}$ be a random vector elliptically distributed from F with h and scatter parameter Σ, then $\boldsymbol{Z}^* = A\boldsymbol{Z} + \boldsymbol{b}$ has the elliptical distribution with the same h and scatter parameter $A\Sigma A^T$. So by (3), we have

$$\tilde{\Sigma}(\boldsymbol{Z}^*) = c(h_F, \sigma)A\Sigma A^T = c(h_F, \sigma)Ac(h_F, \sigma)^{-1}\tilde{\Sigma}(\boldsymbol{Z})A^T = A\tilde{\Sigma}(\boldsymbol{Z})A^T.$$

The modified spatial rank covariance matrix is affine equivariant under elliptical models. For any distribution, it is orthogonally equivariant, the property inherited from the spatial rank covariance matrix.

Theorem 2.2 *For any d-variate random vector* $\boldsymbol{Z}$, $\tilde{\Sigma}(\boldsymbol{Z})$ *is orthogonally equivariant.*

Robustness is one of the characteristics of procedures based on ranks. MRCM is essentially robust. For example, the covariance matrix of Cauchy distribution doesn't exist, but MRCM is still meaningful and equal to the scatter parameter Σ. In the next section, we conduct robustness analysis on the MRCM through two approaches: influence function and breakdown point.

3 Robustness property of MRCM

3.1 Influence function approach

For any fixed point $\boldsymbol{x} \in \mathbb{R}^d$, let $\Delta_{\boldsymbol{x}}$ be the point-mass probability distribution at $\boldsymbol{x}$ and the ε-*contamination distribution* be $F_\varepsilon = (1-\varepsilon)F + \varepsilon\Delta_{\boldsymbol{x}}$. Then the influence function of a functional $T(\cdot)$ at the given distribution F is defined as

$$\mathrm{IF}(\boldsymbol{x}, T; F) = \lim_{\varepsilon \to 0^+} \frac{T(F_\varepsilon) - T(F)}{\varepsilon} = \left.\frac{\partial T(F_\varepsilon)}{\partial \varepsilon}\right|_{\varepsilon=0}.$$

The influence function measures the effect on T of infinitesimal point mass contamination of the distribution F. Clearly, it is desired to be small or at least bounded. A functional T with a bounded influence function is regarded as robust.

Lemma 3.1 *For any given random vector* $\boldsymbol{Z}$ *with distribution* F*, the influence functions of the spatial rank at the point* $\boldsymbol{z}$ *and RCM are given by*

$$\mathrm{IF}(\boldsymbol{x}, \boldsymbol{R}(\boldsymbol{z}, F); F) = \boldsymbol{S}(\boldsymbol{z} - \boldsymbol{x}) - \boldsymbol{R}(\boldsymbol{z}, F),$$

and

$$\begin{aligned}\mathrm{IF}(\boldsymbol{x}, \Sigma_R; F) &= \mathbb{E}_F \boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})^T + \boldsymbol{R}(\boldsymbol{x}, F)\boldsymbol{R}(\boldsymbol{x}, F)^T \\ &\quad -2\Sigma_R(F) - \mathbb{E}_F \mathrm{IF}(\boldsymbol{x}, \boldsymbol{R}(\boldsymbol{Z}, F); F)\mathrm{IF}(\boldsymbol{x}, \boldsymbol{R}(\boldsymbol{Z}, F); F)^T.\end{aligned}$$

Remark 3.1 (i) The influence function of the rank function $\boldsymbol{R}(\boldsymbol{z}, F)$ is bounded with $\sup_{\boldsymbol{x}} \|\mathrm{IF}(\boldsymbol{x}, \boldsymbol{R}(\boldsymbol{z}, F); F)\| = \|\boldsymbol{R}(\boldsymbol{z}, F)\| + 1 < 2$ and the supermum is achieved at $\boldsymbol{x} = \boldsymbol{z} + c\boldsymbol{R}(\boldsymbol{z}, F)$, where c is any positive scalar. In that case, $\mathrm{IF}(\boldsymbol{x}, \boldsymbol{R}(\boldsymbol{z}, F); F) = (1 + \frac{1}{\|R(\boldsymbol{z}, F)\|})R(\boldsymbol{z}, F)$ and its norm is equal to the supermum $\|\boldsymbol{R}(\boldsymbol{z}, F)\| + 1$.
(ii) The influence function for the RCM is bounded due to the boundedness of the spatial sign function, rank function and influence function of the rank function. We say a matrix to

be bounded if all of its elements are bounded.
(iii) The IF of the RCM for a spherically symmetrical distribution F can be obtained from Sirkiä et al. (2009), in which it was derived through the U-theory. However, the result can not be extended to an elliptical distribution by using Lemma 1 of Croux and Haesbroeck (2000), since RCM is not affine equivariant.

The modified rank covariance matrix is determined by the eigenvectors of RCM and robust scale estimator of the univariate projection on each eigenvector. For $\sigma = \text{MAD}$, we first conduct the perturbation analysis for eigenvalues and eigenvectors of $\tilde{\Sigma}(F)$. Their influence functions are given by the following theorem.

Theorem 3.1 *Let $\tilde{\Sigma}(F)$ be the modified spatial rank covariance functional on an elliptical distribution F with $\sigma = \text{MAD}$. Suppose the spatial rank covariance matrix $\Sigma_R(F)$ has distinct eigenvalues $\lambda_1 > ... > \lambda_d > 0$ and the corresponding eigenvectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_d$. Denote $\tilde{\lambda}_j$ and $\tilde{\boldsymbol{v}}_j$ as the j^{th} eigenvalue and corresponding eigenvector of $\tilde{\Sigma}(F)$, respectively. Then the influence functions of $\tilde{\boldsymbol{v}}_j$ and $\tilde{\lambda}_j$ $(j = 1, ..., d)$ are given by*

$$\text{IF}(\boldsymbol{x}, \tilde{\boldsymbol{v}}_j; F) = \sum_{k \neq j}^{d} \frac{1}{\lambda_j - \lambda_k} \{\boldsymbol{v}_k^T \text{IF}(\boldsymbol{x}, \Sigma_R; F) \boldsymbol{v}_j\} \boldsymbol{v}_k$$

and

$$\text{IF}(\boldsymbol{x}, \tilde{\lambda}_j; F) = \frac{1}{4h(1)} \mathbf{sgn}^T (|\boldsymbol{v}_j \circ \boldsymbol{x}| - \mathbf{1}_d) \text{IF}(\boldsymbol{x}, \tilde{\boldsymbol{v}}_j; F), \tag{4}$$

where $\boldsymbol{a} \circ \boldsymbol{b}$ is the component-wise product of $\mathbf{a}$ and $\mathbf{b}$, $|\boldsymbol{a}| = (|a_1|, ..., |a_d|)^T$ is the compoent-wise absolute value, $\mathbf{1_d}$ is the d-vector with all entries 1, and $\mathbf{sgn}(\boldsymbol{a})$ is the component-wise sign vector and equal to $(\text{sgn}(a_1), ..., \text{sgn}(a_d))^T$.

Remark 3.2 (i) The boundedness of IF of RCM implies the boundedness of influence functions for the eigenvectors of $\tilde{\Sigma}(F)$. This feature is distinct from that of eigenvector estimates using projection-pursuit techniques, whose IF's are unbounded. See Croux and Ruiz-Gazen (2005).
(ii) With a robust choice of univariate scale functional (MAD), the influence functions for eigenvalues of $\tilde{\Sigma}(F)$ are also kept bounded.
(iii) It provides an immediate application to robust principle component analysis for dimension reduction.
(iv) The result is obtained under the assumption of distinct eigenvalues. A generalization to multiple eigenvalues is possible as in Tanaka (1988).
(v) The assumption on elliptical symmetry is not necessary. However, for a general model, the representations of IF for MAD and eigenvectors may be in length.

For demonstration and comparison, we compute and plot the influence functions of eigenvalues and eigenvectors for the MRCM, as well as for the classical covariance matrix. At $F = N(\mathbf{0}, \text{diag}(1,4))$, both scatter functionals and RCM have the same eigenvectors $\boldsymbol{v}_1 = (0,1)^T$ and $\boldsymbol{v}_2 = (1,0)^T$. Since IF's of $\boldsymbol{v}_1$ for both functionals are of the form $b(\boldsymbol{x})\boldsymbol{v}_2$, hence the second components are always zero. We plot the first components of IF for $\boldsymbol{v}_1$ in Fig. 3.1 (c) and (d). The curve for our MRCM is saddle-shaped and bounded. Note that if we turn the curve upside down or rotate the curve 90 degree, we obtain the curve of the second component of IF for $\boldsymbol{v}_2$ since it equals the negative of the first component of IF of $\boldsymbol{v}_1$. Surprisingly, the influence function of the largest eigenvalue for MRCM is the first component of IF for $\boldsymbol{v}_1$ multiplying a factor $-1/4h(1) = -2.5898$. The curve of IF of $\tilde{\lambda}_1$ is plotted in Fig. 3.1 (b). The component-wise sign function in the formula (4) doesn't play any role in this case because the first component of $\boldsymbol{v}_1$ is 0 and the second component of IF for $\boldsymbol{v}_1$ is also 0. However, for a general distribution, we shall anticipate that the curve of IF for eigenvalue has more jumps, hence more local valleys and peaks than the curve of IF for eigenvector because of the component-wise sign function. As expected, the curves for our MRCM are kept bounded, while unbounded for the classical covariance matrix.

Based on the influence functions of eigenvalues and eigenvectors, we are able to derive the influence function for our modified spatial rank covariance matrix, which is given in the following theorem.

Theorem 3.2 *For an elliptical distribution F, let the eigenvalues and eigenvectors of $\Sigma_R(F)$ be $\lambda_1 > ... > \lambda_d > 0$ and $\boldsymbol{v}_1, ..., \boldsymbol{v}_d$ respectively. Then the influence function of $\tilde{\Sigma}$ at F is given by*

$$\text{IF}(\boldsymbol{x}, \tilde{\Sigma}; F) = \text{IF}(\boldsymbol{x}, \Sigma_R; F) + \sum_{j=1}^{d} a_j(\boldsymbol{x}) \boldsymbol{v}_j \boldsymbol{v}_j^T,$$

where $a_j(\boldsymbol{x}) = \frac{1}{4h(1)} \sum_{k \neq j}^{d} \frac{1}{\lambda_j - \lambda_k} \boldsymbol{v}_k^T \text{IF}(\boldsymbol{x}, \Sigma_R; F) \boldsymbol{v}_j \mathbf{sgn}^T(|\boldsymbol{v}_j \circ \boldsymbol{x}| - \mathbf{1}_d) \boldsymbol{v}_k - \boldsymbol{v}_j^T \text{IF}(\boldsymbol{x}, \Sigma_R; F) \boldsymbol{v}_j$

Remark 3.3 (i) $a_j(\boldsymbol{x})$ is bounded in $\boldsymbol{x}$, therefore the influence function of $\tilde{\Sigma}(F)$ is bounded. (ii) Even though MRCM is affine equivariant under elliptical distributions, its influence function could not be written as the form in the result of Lemma 1 in Croux and Haesbroeck (2000). This is due to the construction way of MRCM. It is based on RCM, which is the covariance matrix of nonlinearly transformed ranks.

3.2 Breakdown point approach

The influence function measures the local robustness of a functional $T(F)$, while the breakdown point captures the global robustness of estimator $T(F_n)$. Roughly speaking, the finite

sample breakdown point is the minimum fraction of the "bad" data points in a data set $\mathbb{Z} = \{\boldsymbol{z}_1, ..., \boldsymbol{z}_n\}$ that can render the estimator useless. For scale estimators, they can break down in two ways: when they become arbitrarily large (explosion) or when they become arbitrarily close to zero (implosion). In terms of scatter estimators, they break down if one of its eigenvalues approaches 0 or ∞. So the *replacement breakdown point* denoted by ε (Donoho and Huber [7]) of an scatter estimator T at $\mathbb{Z}$ is defined as

$$\varepsilon(T, \mathbb{Z}) = \min\{\frac{m}{n} : \sup_{\mathbb{Z}_m} |\det\{T(\mathbb{Z})T(\mathbb{Z}_m)^{-1}\} + \det\{T(\mathbb{Z})^{-1}T(\mathbb{Z}_m)\}| = \infty\},$$

where $\mathbb{Z}_m$ denotes a contaminated sample resulting from replacing m points of $\mathbb{Z}$ with arbitrary values; $\det(A)$ is the determinant of A.

Before we derive the finite sample breakdown point for our MRCM, a variation of MAD, MAD_k is introduced. Later on, the univariate scale estimator σ, which is used to obtain MRCM, is taken to be the MAD_k. This can lead to a slightly higher breakdown point. Similar ideas were adopted by several authors in the literature, for example by Tyler (1994), Gather and Hilker (1997), Zhou and Dang (2010). More specifically, let $\mathbb{X} = \{x_1, ..., x_n\}$ be a sample of n points in $\mathbb{R}^1$ with ordered values $x_{(1)} \leq \cdots \leq x_{(n)}$. Define

$$\text{MAD}_k(\mathbb{X}) = \text{Med}_k(|x_1 - \text{Med}(\mathbb{X})|, ..., |x_n - \text{Med}(\mathbb{X})|),$$

where $\text{Med}_k(\mathbb{X}) = (x_{(\lfloor (n+k)/2 \rfloor)} + x_{(\lfloor (n+k+1)/2 \rfloor)})/2$, $1 \leq k \leq n$, and $\lfloor a \rfloor$ is the largest integer no larger than a. The usual Med and MAD correspond to Med_k and MAD_k with $k = 1$, respectively.

Theorem 3.3 *For any $d-$variate random sample $\mathbb{Z} = \{\boldsymbol{z}_1, ..., \boldsymbol{z}_n\}$, let $\sigma = \text{MAD}_k$ and $c_1(\mathbb{Z})$ be the maximum number of points of $\mathbb{Z}$ in any $(d-1)$-dimensional hyperplane. If $n > 2c_1(\mathbb{Z}) - k + 1$, then*

$$\varepsilon(\tilde{\Sigma}, \mathbb{Z}) = \begin{cases} \lfloor (n - 2c_1(\mathbb{Z}) + k + 1)/2 \rfloor / n & \text{if } 1 \leq k \leq c_1(\mathbb{Z}) \\ \lfloor (n - k + 2)/2 \rfloor / n & \text{if } c_1(\mathbb{Z}) + 1 \leq k \leq n. \end{cases}$$

Remark 3.4 (i) The breakdown point of $\tilde{\Sigma}$ depends only on the sample size n and $c_1(\mathbb{Z})$, but is independent of other configurations of $\mathbb{Z}$. For the breakdown point as a global robustness assessment of estimators, this 'sample-free' property is definitely desirable.
(ii) The optimal choices of k are $c_1(\mathbb{Z})$ or $c_1(\mathbb{Z}) + 1$ so that $\varepsilon(\tilde{\Sigma}, \mathbb{Z}) = \lfloor (n - c_1(\mathbb{Z}) + 1)/2 \rfloor / n$ attains the maximum value. This is the upper bound of breakdown point for any affine equivariant scatter estimators, see Tyler (1994). Although MRCM is only orthogonally invariant in general, it is affine equivariant for elliptical models. Clearly, this attainability is preferred.

(iii) If $\mathbb{Z}$ is in general position, that is, $c_1(\mathbb{Z}) = d$, then the breakdown point of $\tilde{\Sigma}$ equals $\lfloor (n-d+1)/2 \rfloor / n$ when $k = d$, reaching the upper bound given by Davies (1987).
(iv) Theorem 3.3 focuses on the case $\sigma = \mathrm{MAD}_k$. The result, however, can be extended to any scale estimator with the same breakdown point as MAD_k.
(v) On the discussion of Davies and Gather (2005) [6], Tyler mentioned an example to construct a high breakdown point covariance matrix estimator based on the sample covariance matrix by replacing the eigenvalues of the sample covariance matrix with robust variances for the sample principle component variables. Then the resulting covariance matrix has a high breakdown point. His intention was to call a reasonable concept of breakdown on principle component vectors that are some estimators in a compact set. It is out of scope of this paper to do so. However, it is worthwhile to note that MRCM has different robustness properties comparing to that example. It has bounded influence functions for eigenvectors. Such property is closely related to the bounded breakdown function concept proposed by He and Simpson (1992).

4 Finite sample efficiency and robustness

In this section, we ran a simulation to investigate the finite sample efficiency and robustness of MRCM and compare with other well-known robust scatter estimators. The following estimators are considered:

RCM The function *spatial.rank* in the R package ICSNP is used for computing the spatial rank vector. It is only orthogonally equivalent, hence RCM is only suitable for spherically symmetric distributions.

Mest Tyler's M-estimator is obtained by the function *HR.Mest* in the R package ICSNP. It simultaneously yields affine equivariant spatial median (Hettmansberger and Randles, 2003) and Tyler's shape matrix.

Mcd Minimum covariance determinant estimator is computed by the R package rrcov. The MCD method looks for the h observations (out of n) whose classical covariance matrix has the lowest possible determinant. Then MCD scatter estimator is the covariance matrix based on those h observations. h value is set to be default value $n/2$ of the function *CovMcd* which leads the breakdown point close to 1/2.

Sest Re-weighted S-estimator (Sest) is calculated by the R package riv using Tukey bi-weighted ρ function with $c = 2.661$ for $d = 2$ and $c = 4.652$ for $d = 5$. Such c values provided as output of the function *slc* yield the breakdown point close to 1/2.

Cov Non-robust sample covariance matrix.

		$\varepsilon = 0$		$\varepsilon = 0.1$		$\varepsilon = 0.2$	
		$d = 2$	$d = 5$	$d = 2$	$d = 5$	$d = 2$	$d = 5$
$n = 50$	MRCM(RE)	0.72	0.83	1.07	1.26	1.21	1.39
	RCM(RE)	0.35	0.89	0.56	1.40	0.59	1.47
	Mest(RE)	0.71	0.86	1.08	1.34	1.22	1.39
	Mcd(RE)	0.55	0.65	0.86	1.08	1.03	1.16
	Sest(RE)	0.59	0.87	0.89	1.36	1.06	1.38
	Cov(MLCN)	0.26	0.68	0.40	1.07	0.43	1.12
$n = 200$	MRCM(RE)	0.74	0.82	1.24	1.36	1.22	1.48
	RCM(RE)	0.26	0.67	0.31	0.92	0.32	0.96
	Mest(RE)	0.75	0.86	1.13	1.38	1.12	1.46
	Mcd(RE)	0.65	0.78	1.05	1.35	0.96	1.44
	Sest(RE)	0.64	0.90	1.03	1.42	1.03	1.51
	Cov(MLCN)	0.12	0.33	0.20	0.54	0.21	0.58

Table 1: Mean of log condition numbers (MLCN) of the sample covariance matrix (Cov) and relative efficiencies (RE) of other estimators relative to Cov under $F = (1-\varepsilon)\mathcal{N}(\mathbf{0}, \Sigma_{d\times d}) + \varepsilon\mathcal{N}(10\mathbf{1}_d, \Sigma_{d\times d})$, where $\Sigma_{d\times d} = \text{diag}(4, \mathbf{1}_{d-1}^T)$.

As performance criteria for matrices, two quantities are used. One is the condition number of $\Sigma^{-1}V$, where Σ is the true scatter matrix and V is one of the above mentioned estimators. It is the ratio of the largest eigenvalue to the smallest eigenvalue of $\Sigma^{-1}V$. A good estimator V estimates Σ well such that $\Sigma^{-1}V$ is close to the identity matrix, hence the mean of log condition number (MLCN) of $\Sigma^{-1}V$ is expected to be close to 0. That is,

$$MLCN = \frac{1}{M}\sum_{i=1}^{M} \log \frac{\lambda_1^{(i)}}{\lambda_d^{(i)}},$$

where $\lambda_1^{(i)}$ and $\lambda_d^{(i)}$ are the largest and smallest eigenvalues of $\Sigma^{-1}V^{(i)}$ with $V^{(i)}$ being the estimator of Σ in the i^{th} sample. A good estimator tends to have a small value of MLCN. Such a criterion was also utilized in Maronna and Yohai (1995) and Gervini (2003). The other one is called the MSAD, the mean squared angle difference in the direction of the first eigenvectors to measure the accuracy on estimating the first principle component, that is, $(\cos^{-1}(|\boldsymbol{v}_1^T\hat{\boldsymbol{v}}_1|))^2$, where $\boldsymbol{v}_1$ is the first eigenvector of the theoretical scatter matrix Σ and $\hat{\boldsymbol{v}}_1$ is the first eigenvector of the scatter estimator V. See Croux, Ollila and Oja (2002) for asymptotical properties of MSAD.

		$\nu=1$		$\nu=3$		$\nu=5$		$\nu=\infty$	
		$d=2$	$d=5$	$d=2$	$d=5$	$d=2$	$d=5$	$d=2$	$d=5$
$n=50$	MRCM(RE)	4.53	3.87	1.35	1.38	1.01	1.14	0.72	0.83
	RCM(RE)	3.57	4.68	0.99	1.72	0.71	1.43	0.35	0.89
	Mest(RE)	4.81	4.42	1.36	1.41	1.02	1.10	0.71	0.86
	Mcd(RE)	3.32	2.48	0.99	0.96	0.72	0.79	0.55	0.65
	Sest(RE)	3.54	3.11	1.09	1.25	0.80	1.01	0.59	0.87
	Cov(MLCN)	2.51	4.65	0.68	1.57	0.50	1.28	0.26	0.68
$n=200$	MRCM(RE)	8.73	7.89	1.53	1.82	1.11	1.17	0.74	0.82
	RCM(RE)	3.46	6.97	0.63	1.61	0.39	1.06	0.26	0.67
	Mest(RE)	9.89	8.25	1.51	1.94	1.09	1.20	0.75	0.86
	Mcd(RE)	7.21	5.43	1.22	1.50	0.87	0.92	0.65	0.78
	Sest(RE)	7.58	6.39	1.43	1.78	0.95	1.14	0.64	0.90
	Cov(MLCN)	2.43	4.47	0.41	1.03	0.27	0.65	0.12	0.33

Table 2: Mean of log condition numbers (MLCN) for Cov and relative efficiencies (RE) for other estimators relative to Cov under t_ν-distributions with $\mu = \mathbf{0}, \Sigma = \text{diag}(4, \mathbf{1}_{d-1}^T)$.

We generate $M = 1000$ samples with two sample sizes $n = 50$ and $n = 200$ from the first two scenarios and with sample sizes $n = 100$ and $n = 400$ from the third scenario.

Case I Standard normal distributions with contaminations on shifted locations. i.e $(1-\varepsilon)\mathcal{N}(\mathbf{0}, \Sigma_{d\times d}) + \varepsilon\mathcal{N}(\mathbf{10}_d, \Sigma_{d\times d})$, where $\mathbf{10}_d$ is the d-vector with all elements 10, $\Sigma_{d\times d} = \text{diag}(4, \mathbf{1}_{d-1}^T)$, ε to be 0, 0.1, 0.2 and $d = 2, 5$.

Case II Heavy-tailed $t_\nu(\mathbf{0}, \Sigma_{d\times d})$ distributions for different degrees of freedom $\nu = 1, 3, 5$ and ∞ with dimension $d = 2$ and $d = 5$. Note that $\nu = \infty$ corresponds to Case I with $\varepsilon = 0$, the standard normal distributions.

Case III Normal mixtures with contaminations on rotation. i.e. $(1-\varepsilon)\mathcal{N}(\mathbf{0}, \Sigma_{d\times d}) + \varepsilon\mathcal{N}(\mathbf{0}, \Sigma^*_{d\times d})$, where $\Sigma_{d\times d} = \text{diag}(4, \mathbf{1}_{d-1}^T)$ and $\Sigma^*_{d\times d} = 10 \times \text{diag}(\mathbf{1}_{d-1}^T, 4)$. We take $\varepsilon = 0, 0.05, 0.1$.

In the first two cases, we check the efficiency and robustness on the eigenvalues of estimators, hence use the mean of log(cond) (MLCN) as criterion to measure non-sphericity of $\Sigma^{-1}V$. The finite sample relative efficiency (RE) of estimator V is obtained by the ratio of MLCN of sample covariance to that of V. Reported in Tables 1 and 2 are MLCN for Cov and RE for other estimators. In the third case, contaminations have totally different orientation and we are interested in estimating the first principle component. Hence the

		$\varepsilon = 0$		$\varepsilon = 0.05$		$\varepsilon = 0.1$	
		$d = 2$	$d = 5$	$d = 2$	$d = 5$	$d = 2$	$d = 5$
$n = 100$	MRCM(RE)	0.93	0.92	42.27	20.04	120.83	61.15
	Mest(RE)	0.54	0.70	32.74	17.82	135.49	66.48
	Mcd(RE)	0.38	0.61	31.46	16.70	136.21	71.36
	Sest(RE)	0.39	0.76	29.00	21.04	151.69	85.30
	$\mathrm{Cov}((\cos^{-1}(\boldsymbol{v}_1^T\hat{\boldsymbol{v}}_1))^2)$	0.00	0.01	0.19	0.24	1.01	0.99
$n = 400$	MRCM(RE)	0.91	0.93	46.27	16.13	382.35	165.03
	Mest(RE)	0.53	0.73	34.45	13.80	447.80	177.38
	Mcd(RE)	0.43	0.71	31.94	14.81	513.11	219.25
	Sest(RE)	0.38	0.79	30.94	16.42	511.27	236.25
	$\mathrm{Cov}((\cos^{-1}(\boldsymbol{v}_1^T\hat{\boldsymbol{v}}_1))^2)$	0.00	0.00	0.07	0.08	1.13	1.07

Table 3: Mean of squared angle differences in the direction (MSAD measured in squared radians) for Cov and relative efficiencies for other estimators relative to Cov under $F = (1-\varepsilon)\mathcal{N}(\mathbf{0}, \Sigma_{d\times d}) + \varepsilon\mathcal{N}(\mathbf{0}, \Sigma^*_{d\times d})$, where $\Sigma_{d\times d} = \mathrm{diag}(4, \mathbf{1}^T_{d-1})$, and $\Sigma^*_{d\times d} = 10 \times \mathrm{diag}(\mathbf{1}^T_{d-1}, 4)$.

mean of squared angle differences (MSAD) in the direction between the first eigenvector of the true scatter matrix Σ and that of V (MSAD) are computed, and the relative efficiency of V is the ratio of MSAD of Cov to that of V. Results of MSAD of Cov and RE's of other estimators are listed in Table 3. RCM is skipped since it yields the exactly same results as the MRCM.

For the non-standard normal distribution, RCM performs poorly due to its failure on estimating eigenvalues of the true covariance matrix. The MRCM seems to be better than other estimators, especially in estimating orientation of data clouds with the efficiencies at least 0.90. This results agree well with the conclusions from Croux, *et al.* (2002). The finite sample efficiencies of M-estimator are also close to its asymptotic one $d/(d+2)$, where d is the dimension. Efficiency on shape (eigenvalues) of MRCM is lower than it on the orientation, while the other estimators have almost the same RE on the shape as it on the orientation. This phenomenon can be explained by the separated steps in the construction of MRCM and relatively low efficiency on MAD_k. In the contaminated normal models, the MRCM has a comparable or better performance comparing with other robust estimators. MRCM and Tyler M-estimator have very similar behaviors since both of them are based on spatial procedures with some treatments for affine equivariance property. Without a surprise, Tyler M-estimator is superior to MRCM and others under the heavy-tailed t-distribution of the degree of freedom $\nu = 1$, since it is the limiting form of MLE for scatter as $\nu \to 0$. In summary, MRCM has a competitive performance on efficiency as well as robustness.

5 Discussion

In this paper, we studied robustness properties of the modified spatial rank covariance matrix proposed by Visuri, Oja and Koivunen (2000). It is robust locally in terms of the influence function and highly robust globally in terms of finite sample breakdown point. We derived the influence functions for the eigenvectors and eigenvalues of the MRCM, then the influence function for the MRCM. They are bounded under the assumption that the scatter parameter has distinct eigenvalues. A generalization to multiple eigenvalues is possible as in Tanaka (1988). The breakdown point attains the upper bound by the choice of robust univariate scale functional $\sigma = \text{MAD}_k$ with some optimal values for k. Comparing with other high breakdown point estimators such as the MCD, the S-estimates and the projection based estimates, our MRCM is easy to compute with the complexity $O(n^2 + d^3)$. Even for large data sets in high dimensions, using MRCM is still practical. Also, MRCM is highly efficient under normal models and heavy-tailed distributions.

Although we focused on the spatial rank covariance matrix, the general approach and results can be applied to any estimate of covariance matrix which is orthogonally equivariant but not necessarily affine equivariant. That is, for any such estimate if one replaces the eigenvalues with an estimate of scale applied to the corresponding principle component variable, then the resulting modified estimate is consistent for the shape matrix of an elliptical distribution; is affine equivariant at elliptical models; has influence functions which have essentially the same form as that given in Theorem 3.1 and Theorem 3.2 when the MAD is used; and has the same breakdown points as that given in Theorem 3.3 when MAD_k is used. Also, the results can be easily extended to any robust estimate of scale. For example, using alternatives to the MAD proposed by Rousseeuw and Croux [27], influence functions and breakdown point can be similarly obtained for the corresponding covariance matrix.

Under elliptical symmetry, MRCM is affine equivariant and proportional to the scatter parameter. Indeed, the class of distributions for which MRCM has such properties is broader than the class of elliptical distributions. For example, it can be shown that for $F_{\boldsymbol{Z}^*}$ where $\boldsymbol{Z}^* = A\boldsymbol{Z} + \boldsymbol{\mu}$ for $\boldsymbol{Z}$ having an exchangeable and symmetric distribution. That is, $\boldsymbol{Z}$ and $DJ\boldsymbol{Z}$ have the same distribution for any permutation matrix J and any diagonal matrix D with diagonal elements ± 1. Elliptical models belong to this class, since the distribution of the corresponding $\boldsymbol{Z}$ is independent and symmetric in each component. Under a distribution in this class, we may obtain affine equivariant location estimator using the spatial median along with the transformation and retransformation technique. More specifically, one can find the spatial median of the transformed data $(\tilde{\Sigma}^{-1/2}(\mathbb{Z}^*)\boldsymbol{z}_1^*, \cdots, \tilde{\Sigma}^{-1/2}(\mathbb{Z}^*)\boldsymbol{z}_n^*)$, denoted as M_s, then transform it back to the original coordinate system, i.e., $\tilde{\Sigma}^{1/2}(\mathbb{Z}^*)M_s$ gives an affine

equivariant location estimator.

Tyler *et al.* (2009) presented a general method for exploring multivariate data based on the spectrum decomposition of one scatter matrix relative to another. For a distribution which lies outside of the above mentioned class, different scatter statistics may estimate different quantities of underly distributions, hence their method may reveal interesting features in data structure. Our MRCM, an easy-computed high breakdown point scatter matrix, certainly deserves further investigation in application to their method and other multivariate methods.

6 Appendix

Proof of Theorem 2.2: Let $\boldsymbol{Z}^* = cO\boldsymbol{Z} + \boldsymbol{b}$ for any orthogonal matrix O, d-vector $\boldsymbol{b}$ and nonzero scalar c. Let the spectral decomposition of $\Sigma_R(\boldsymbol{Z})$ be $U\Lambda_R U^T$. By the orthogonal equivariance of spatial rank covariance matrix, we have

$$\Sigma_R(\boldsymbol{Z}^*) = c^2 O\Sigma_R(\boldsymbol{Z})O^T = c^2 OU\Lambda_R U^T O^T = c^2(OU)\Lambda_R(OU)^T,$$

so the eigenvector matrix of $\Sigma_R(\boldsymbol{Z}^*)$ is OU, which is $[O\boldsymbol{u}_1, ..., O\boldsymbol{u}_d]$. Then for each scale estimate $\tilde{\lambda}_i(\boldsymbol{Z}^*) = \sigma^2((O\boldsymbol{u}_i)^T\boldsymbol{Z}^*) = \sigma^2((O\boldsymbol{u}_i)^T(cO\boldsymbol{Z}+\boldsymbol{b})) = c^2\sigma^2(\boldsymbol{u}_i^T O^T O\boldsymbol{Z}) = c^2\sigma^2(\boldsymbol{u}_i^T\boldsymbol{Z}) = c^2\tilde{\lambda}_i(\boldsymbol{Z})$ for $i = 1, ..., d$. Therefore, by the way we construct $\tilde{\Sigma}$, we have

$$\tilde{\Sigma}(\boldsymbol{Z}^*) = (OU)c^2\tilde{\Lambda}(\boldsymbol{Z})(OU)^T = c^2 O\tilde{\Sigma}(\boldsymbol{Z})O^T.$$

□

Proof of Lemma 3.1: We have

$$\begin{aligned}
\mathrm{IF}(\boldsymbol{x}, \boldsymbol{R}(\boldsymbol{z}, F); F) &= \frac{\partial}{\partial\varepsilon}\boldsymbol{R}(\boldsymbol{z}, (1-\varepsilon)F + \varepsilon\Delta_{\boldsymbol{x}})\Big|_{\varepsilon=0} \\
&= \frac{\partial}{\partial\varepsilon}\left[(1-\varepsilon)\boldsymbol{R}(\boldsymbol{z}, F) + \varepsilon\frac{\boldsymbol{z}-\boldsymbol{x}}{\|\boldsymbol{z}-\boldsymbol{x}\|}\right]\Bigg|_{\varepsilon=0} \\
&= \boldsymbol{S}(\boldsymbol{z}-\boldsymbol{x}) - \boldsymbol{R}(\boldsymbol{z}, F).
\end{aligned}$$

Because

$$\begin{aligned}
\mathrm{IF}(\boldsymbol{x}, \Sigma_R; F) &= \frac{\partial}{\partial\varepsilon}\mathbb{E}_{F_\varepsilon}\{\boldsymbol{R}(\boldsymbol{Z}, F_\varepsilon)\boldsymbol{R}(\boldsymbol{Z}, F_\varepsilon)^T\}\Big|_{\varepsilon=0} \\
&= \frac{\partial}{\partial\varepsilon}\mathbb{E}_{F_\varepsilon}\{[(1-\varepsilon)\boldsymbol{R}(\boldsymbol{Z}, F) + \varepsilon\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})][(1-\varepsilon)\boldsymbol{R}(\boldsymbol{Z}, F) + \varepsilon\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})]^T\}\Big|_{\varepsilon=0} \\
&= \frac{\partial}{\partial\varepsilon}(1-\varepsilon)\mathbb{E}_F\mathbf{M} + \varepsilon 1_{[\boldsymbol{Z}=\boldsymbol{x}]}\mathbf{M}\Big|_{\varepsilon=0}
\end{aligned}$$

where 1_A is the indicator function being 1 when A is true and 0 otherwise, and

$$\mathbf{M} = [(1-\varepsilon)\boldsymbol{R}(\boldsymbol{Z},F) + \boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})][(1-\varepsilon)\boldsymbol{R}(\boldsymbol{Z},F) + \boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})]^T\}.$$

Simplify further, we have

$$\begin{aligned}
\mathrm{IF}(\boldsymbol{x},\Sigma_R;F) &= -3\Sigma_R(F) + \mathbb{E}_F\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})\boldsymbol{R}(\boldsymbol{Z},F)^T + \mathbb{E}_F\boldsymbol{R}(\boldsymbol{Z},F)\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})^T \\
&\quad +\boldsymbol{R}(\boldsymbol{x},F)\boldsymbol{R}(\boldsymbol{x},F)^T \\
&= -3\Sigma_R(F) - \mathbb{E}_F[\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x}) - \boldsymbol{R}(\boldsymbol{Z},F)][\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x}) - \boldsymbol{R}(\boldsymbol{Z},F)]^T \\
&\quad +\mathbb{E}_F\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})^T + \Sigma_R(F) + \boldsymbol{R}(\boldsymbol{x},F)\boldsymbol{R}(\boldsymbol{x},F)^T \\
&= \mathbb{E}_F\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})\boldsymbol{S}(\boldsymbol{Z}-\boldsymbol{x})^T + \boldsymbol{R}(\boldsymbol{x},F)\boldsymbol{R}(\boldsymbol{x},F)^T - 2\Sigma_R(F) \\
&\quad -\mathbb{E}_F\mathrm{IF}(\boldsymbol{x},\boldsymbol{R}(\mathbf{Z},F);F)\mathrm{IF}(\boldsymbol{x},\boldsymbol{R}(\boldsymbol{Z},F);F)^T.
\end{aligned}$$

□

The proof of Theorem 3.1 relies on the following lemma, which is Lemma 3 of Croux and Haesbroeck (2000).

Lemma 1 *Let $S : \mathcal{F} \to SPD(d)$ be a statistical functional and F a d-dimensional distribution. Suppose that $\mathrm{IF}(\boldsymbol{x},S;F)$ exists. Denote $\boldsymbol{v}_1,...,\boldsymbol{v}_d$ and $\lambda_1,...,\lambda_d$ the eigenvectors and eigenvalues of $S(F)$. Then the influence functions of $\boldsymbol{v}_j$ and λ_j are given by*

$$\mathrm{IF}(\boldsymbol{x},\lambda_j;F) = \boldsymbol{v}_j^T\mathrm{IF}(\boldsymbol{x},S;F)\boldsymbol{v}_j,$$

$$\mathrm{IF}(\boldsymbol{x},\boldsymbol{v}_j;F) = \sum_{k\neq j}^{d}\frac{1}{\lambda_j-\lambda_k}\{\boldsymbol{v}_k^T\mathrm{IF}(\boldsymbol{x},S;F)\boldsymbol{v}_j\}\boldsymbol{v}_k.$$

Proof of Theorem 3.1: Since $\tilde{\boldsymbol{v}}_j = \boldsymbol{v}_j$ for $j = 1,...,d$, the influence function of $\tilde{\boldsymbol{v}}_j$ is directly followed from Lemma 1 with $S = \Sigma_R$.

Treating Median and MAD as simultaneous M-estimators as in Page 135 of [14] (Huber and Ronchetti, 2009), it is easy to prove that for any unit vector $\boldsymbol{u} \in R^d$ under an elliptical model

$$\mathrm{IF}(\boldsymbol{u}^T\boldsymbol{x},\mathrm{MAD};F_{\boldsymbol{u}}) = \frac{\mathrm{sgn}(|\boldsymbol{u}^T\boldsymbol{x}|-1)}{4h(1)}. \tag{5}$$

Now $\tilde{\lambda}_j = \mathrm{MAD}(F_{\boldsymbol{v}_j})$. By the chain rule of vector derivatives,

$$\frac{\partial\tilde{\lambda}_j(F_\varepsilon)}{\partial\varepsilon} = \frac{\partial\tilde{\lambda}_j(F_\varepsilon)}{\partial\boldsymbol{v}_j(F_\varepsilon)}\frac{\partial\boldsymbol{v}_j(F_\varepsilon)}{\partial\varepsilon}. \tag{6}$$

The evaluation at $\varepsilon = 0$ of the second derivative on the right hand side of (6) is the column vector that is the influence function of $\tilde{\boldsymbol{v}}_j$. The evaluation of the first derivative at $\varepsilon = 0$ is

the row vector whose i^{th} element is the influence function of MAD at $(0,...0,\boldsymbol{v}_{j_i},0,...0)^T\boldsymbol{x}$. By (5), the influence function of eigenvalue of $\tilde{\Sigma}$ follows.

□

Proof of Theorem 3.2: The results of Theorem 3.1 imply that

$$\tilde{\lambda}_j(F_\varepsilon) = \tilde{\lambda}_j(F) + \varepsilon \mathrm{IF}(\boldsymbol{x}, \tilde{\lambda}_j; F) + O(\varepsilon^2),$$

and

$$\boldsymbol{v}_j(F_\varepsilon) = \boldsymbol{v}_j(F) + \varepsilon \mathrm{IF}(\boldsymbol{x}, \boldsymbol{v}_j; F) + O(\varepsilon^2).$$

Then

$$\begin{aligned}
\tilde{\Sigma}(F_\varepsilon) &= \sum_{j=1}^{d} \tilde{\lambda}_j(F_\varepsilon)\boldsymbol{v}_j(F_\varepsilon)\boldsymbol{v}_j^T(F_\varepsilon) \\
&= \sum_{j=1}^{d} \{\tilde{\lambda}_j(F)\boldsymbol{v}_j(F)\boldsymbol{v}_j^T(F) + \varepsilon \mathrm{IF}(\boldsymbol{x}, \tilde{\lambda}_j; F)\boldsymbol{v}_j(F)\boldsymbol{v}_j^T(F) \\
&\quad + \varepsilon\tilde{\lambda}_j(F)\mathrm{IF}(\boldsymbol{x}, \boldsymbol{v}_j; F)\boldsymbol{v}_j^T(F) + \varepsilon\tilde{\lambda}_j(F)\boldsymbol{v}_j(F)\mathrm{IF}(\boldsymbol{x}, \boldsymbol{v}_j; F)^T\} + O(\varepsilon^2).
\end{aligned}$$

Hence,

$$\mathrm{IF}(\boldsymbol{x}, \tilde{\Sigma}; F) = \sum_{j=1}^{d} \{\mathrm{IF}(\boldsymbol{x}, \tilde{\lambda}_j; F)\boldsymbol{v}_j(F)\boldsymbol{v}_j^T(F) + \tilde{\lambda}_j(F)[\mathrm{IF}(\boldsymbol{x}, \boldsymbol{v}_j; F)\boldsymbol{v}_j^T(F) + \boldsymbol{v}_j(F)\mathrm{IF}(\boldsymbol{x}, \boldsymbol{v}_j; F)^T]\}. \tag{7}$$

The summation of the last two terms is

$$\sum_{j=1}^{d}\sum_{k\neq j}^{d} \boldsymbol{v}_j^T \mathrm{IF}(\boldsymbol{x}, \Sigma_R; F)\boldsymbol{v}_k\boldsymbol{v}_k\boldsymbol{v}_j^T = \mathrm{IF}(\boldsymbol{x}, \Sigma_R; F) - \sum_{j=1}^{d} \boldsymbol{v}_j^T \mathrm{IF}(\boldsymbol{x}, \Sigma_R; F)\boldsymbol{v}_j\boldsymbol{v}_j\boldsymbol{v}_j^T.$$

Plugging the influence functions of eigenvectors into the first term of (7) in the right side yields the stated expressions.

□

Proof of Theorem 3.3: Let $\varepsilon^*(\sigma, \mathbb{Z})$ represent the uniform finite sample replacement breakdown point of σ at $\mathbb{Z}$ as defined by Tyler (1994) when all univariate projections of the data are considered. That is,

$$\varepsilon^*(\sigma, \mathbb{Z}) = \min_m \{\frac{m}{n} : \sup_{\|\boldsymbol{u}\|=1} \sup_{\mathbb{Z}_m} \{\sigma(\boldsymbol{u}^T\mathbb{Z})\sigma(\boldsymbol{u}^T\mathbb{Z}_m)^{-1} + \sigma(\boldsymbol{u}^T\mathbb{Z})^{-1}\sigma(\boldsymbol{u}^T\mathbb{Z}_m)\} = \infty\}.$$

Let $\varepsilon(\sigma, \boldsymbol{u}^T\mathbb{Z})$ represent the finite sample breakdown point for σ for the projected data in direction $\boldsymbol{u}$.

The main idea of the proof as follows. The estimator $\tilde{\Sigma}(\mathbb{Z})$ breaks down only if $\sigma = \text{MAD}_k$ breaks down for some direction $\boldsymbol{u}$. Since MAD_k can be exploded ($\to \infty$) or imploded ($\to 0$), the breakdown point of $\tilde{\Sigma}(\mathbb{Z})$ is determined by two quantities corresponding to the explosion and implosion of MAD_k, respectively.

According to Lemma 1 in [10] (Gather and Hilker 1997), for $k \in [1, c_2(\boldsymbol{u}^T\mathbb{Z})]$, MAD_k in direction $\boldsymbol{u}$ will implode with the breakdown point being $\lfloor (n-2c_2(\boldsymbol{u}^T\mathbb{Z})+k+1)/2 \rfloor /n$, where $c_2(\boldsymbol{u}^T\mathbb{Z})$ represents the maximum number of data points on the hyperplane orthogonal to the direction $\boldsymbol{u}$. If $k \in [c_2(\boldsymbol{u}^T\mathbb{Z}) + 1, n]$, the finite sample explosive breakdown points for MAD_k in direction $\boldsymbol{u}$ is $\lfloor (n-k+2/2 \rfloor /n$. Tyler (1994) states that $\varepsilon^*(\sigma, \mathbb{Z}) \leq \inf_{\boldsymbol{u}} \varepsilon(\sigma, \boldsymbol{u}^T\mathbb{Z})$ and equality holds if $\sigma(\boldsymbol{u}^T\mathbb{Z})$ is a continuous function of $\boldsymbol{u}$. This is the case if σ is MAD_k. Also note that $c_2(\boldsymbol{u}^T\mathbb{Z}) \leq c_1(\mathbb{Z})$ for any $\boldsymbol{u}$ with equality holding for some $\boldsymbol{u}$. So we have

$$\varepsilon^*(\text{MAD}_k, \mathbb{Z}) = \begin{cases} \lfloor (n - 2c_1(\mathbb{Z}) + k + 1)/2 \rfloor /n & \text{if } 1 \leq k \leq c_1(\mathbb{Z}) \\ \lfloor (n-k+2)/2 \rfloor /n & \text{if } c_1(\mathbb{Z}) + 1 \leq k \leq n. \end{cases}$$

The proof will be finished if we show that $\varepsilon(\tilde{\Sigma}, \mathbb{Z}) = \varepsilon^*(\text{MAD}_k, \mathbb{Z})$. First, $\varepsilon(\tilde{\Sigma}, \mathbb{Z}) \leq \varepsilon^*(\text{MAD}_k, \mathbb{Z})$ due to orthogonal equivariance of $\tilde{\Sigma}$. This follows by noting that if $m = \lfloor (n - 2c_1(\mathbb{Z}) + k + 1)/2 \rfloor$ and the replacements all lie in the same plane as $c_1(\mathbb{Z})$ data points reside, then MAD_k equals to 0 for the univariate projection orthogonal to that plane. By the orthogonal equivariance of $\tilde{\Sigma}$, there exists orthogonal matrix O such that $\tilde{\Sigma}(O\mathbb{Z}_m)$ has an eigenvalue 0. Hence $0 = \det\{\tilde{\Sigma}(O\mathbb{Z}_m)\} = \det\{O\tilde{\Sigma}(\mathbb{Z}_m)O^T\} = \det(O)\det\{\tilde{\Sigma}(\mathbb{Z}_m)\}\det(O^T) = \det\{\tilde{\Sigma}(\mathbb{Z}_m)\}$ and $\tilde{\Sigma}$ implodes. Similarly, when MAD_k explodes, our estimator explodes.

One the other hand, we also have $\varepsilon(\tilde{\Sigma}, \mathbb{Z}) \geq \varepsilon^*(\text{MAD}_k, \mathbb{Z})$. Suppose $\varepsilon_m = \varepsilon(\tilde{\Sigma}, \mathbb{Z}) < \varepsilon^*(\text{MAD}_k, \mathbb{Z})$. Then for all ε_m-corrupted data sets $\mathbb{Z}_m$ and for all unit directions $\boldsymbol{u}$, there exist σ_0 and σ_1 such that $0 < \sigma_0 < \text{MAD}_k(\boldsymbol{u}^T\mathbb{Z}) < \sigma_1 < \infty$. This implies that for all $\mathbb{Z}_m$, $0 < \sigma_0 < \lambda_j(\tilde{\Sigma}(\mathbb{Z}_m)) < \sigma_1 < \infty$ for all $j = 1, ..., d$, where λ_j is an eigenvalue of $\tilde{\Sigma}(\mathbb{Z}_m)$. Hence $\tilde{\Sigma}$ does not break down at ε_m, contradicting the definition of $\varepsilon(\tilde{\Sigma}, \mathbb{Z})$.

□

7 Acknowledgements

The authors highly value reviewers' helpful comments and suggestions. The work of Y. Chen is supported by the National Science Foundation under Grant No. MCB-1027989.

References

[1] Croux, C., Dehon, C. and Yadine, A. (2010). The k-step spatial sign covariance matrix. *Advances in Data Analysis and Classification*, **4**, 137–150.

[2] Croux, C. and Haesbroeck, G. (2000). Principle component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika*,**87**, 603–618.

[3] Croux, C., Ollila, E. and Oja, H. (2002). Sign and rank covariance matrices: Statistical properties and application to principle component analysis. In *Statistical Data Analysis Based on the L_1 Norm and Related Methods* (ed. by Y. Dodge), Birkhäuser, Basel, 257–270.

[4] Croux, C. and Ruiz-Gazen, A. (2005). High breakdown estimators for principle components: the projection-pursuit approach revisited. *Journal of Multivariate Analysis*, **95**, 206–226.

[5] Davies, P. (1987). Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices. *Annals of Statistics*, **15**, 1269–1292.

[6] Davies, P. and Gather, U. (2005). Breakdown and groups. *Annals of Statistics*, **33**, 977–1035.

[7] Donoho, D. and Huber, P. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. Bickel, K. Doksum and J. Hodges, eds.) 157–184. Wadsworth, Belmont, CA.

[8] Dümbgen, L. (1998). On Tyler's M-functional of scatter in high dimension. *Annals of Institute of Statistical Mathematics*, **50**, 471–491.

[9] Dümbgen, L. and Tyler, D. E.(2005). On the breakdown properties of some multivariate M-functionals. *Scandinavian Journal of Statistics*, **32**, 247–264.

[10] Gather, U. and Hilker, T. (1997). A note of Tyler's modification of the MAD for the Stahel-Donoho estimator. *Annals of Statistics*, **25**, 2024–2026.

[11] Gervini, D. (2003). A robust and efficient adaptive reweighted estimator of multivariate location and scatter. *Journal of Multivariate Analysis*, **84**, 116–144.

[12] He, X. and Simpson, D. (1992). Robust direction estimation. *The Annals of Statistics*, **20**, 351–369.

[13] Hettmansperger, T. and Randles, R. (2002). A practical affine equivariant multivariate median. *Biometrika*, **89**, 851–860.

[14] Huber, P. and Ronchetti, E. M. (2009). *Robust Statistics, 2nd edition*, John Wiley & Sons, New Jersey.

[15] Koltchinskii, V. I. (1997). M-estimation, convexity and quantiles. *Annals of Statistics*, **25**, 435–477.

[16] Marden, J. (1999). Some robust estimates of principal components, *Statistics & Probability Letters*, **43**, 349–359.

[17] Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Annals of Statistics*, **4**, 51–67.

[18] Maronna, R. A. and Yohai, V. J. (1995). The behavior of the Stahel-Donoho robust multivariate estimator. *Journal of American Statistical Association*, **90**, 330–341.

[19] Mazumder, S. and Serfling, R (2010). Spatial trimming, with applications to robustify sample spatial quantile and outlyingness functions, and to construct a new robust scatter estimator. Preprint.

[20] Möttönen, J., Oja, H., and Tienari, J. (1997). On the efficiency of multivariate spatial sign and rank tests. *Annals of Statistics*, **25**, 542–552.

[21] Oja, H. (2010). *Multivariate Nonparametric Methods with R: An Approach Based on Spatial Signs and Ranks*. Springer

[22] Oja, H. and Randles, R (2004). *Multivariate nonparametric tests. Statistical Science*, **19**, 598–605.

[23] Ollila, E., Oja, H. and Hettmansperger, T. (2002). Estimates of regression coefficients based on the sign covariance matrix. *Journal of Royal Statistical Society, B*, **64**, 447–466.

[24] Ollila, E., Oja, H. and Koivunen, V. (2003). Estimates of regression coefficients based on lift covariance matrix. *Journal of the American Statistical Association*, **98**, 90–98.

[25] Ollila, E., Croux, C. and Oja, H. (2004). Influence function and asymptotic efficiency of the affine equivariant rank covariance matrix. *Statistica Sinica*, **14**, 297–316.

[26] Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications*, Vol. B, edited by W. Grossmann, G. Pug, I. Vincze and W. Wertz. Dordrecht, Reidel Publishing Company, 283–297.

[27] Rousseeuw, P. and Croux, C. (1993). Alternatives to the Median absolute deviation. *Journal of the American Statistical Association*, **88**, 1273–1283.

[28] Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization. *Journal of Nonparametric Statistics*, **22**, 915–936.

[29] Sirkiä, S., Taskinen, S., Oja, H. and Tyler, D. (2009). Tests and estimates of shape based on spatial signs and ranks. *Journal of Nonparametric Statistics*, **21**, 155–176.

[30] Tanaka, Y. (1988). Sensitivity analysis in PCA: influence on the subspace spanned by principle components. *Communication in Statistics, A*, **17**, 3157–3175.

[31] Taskinen, S., Sirkiä, S. and Oja, H. (2010). K-step shape estimators based on spatial signs and ranks. *Journal of Statistical Planning and Inference*, **140**, 3376–3388.

[32] Tyler, D. (1987). A distribution-free M-estimator of multivariate scatter. *Annals of Statistics*, **15**, 234–251.

[33] Tyler, D. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Annals of Statistics*, **22**, 1024-1044.

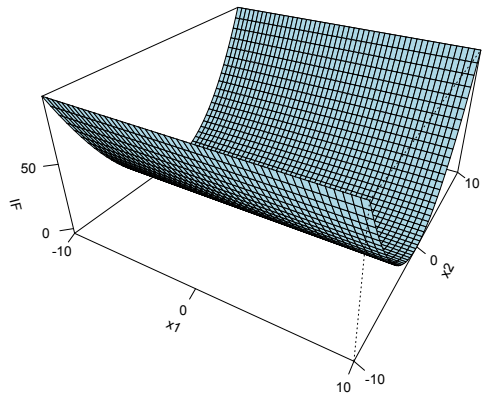
[34] Tyler, D., Critchley, F., Dümbgen, L. and Oja, H. (2009). Invariant co-ordinate selection. *Journal of the Royal Statistical Society, Series B*, **71**, 549–592.

[35] Visuri, S., Koivunen, V. and Oja, H. (2000). Sign and rank covariance matrices. *Journal of Statistical Planning and Inference*, **91**, 557–575.

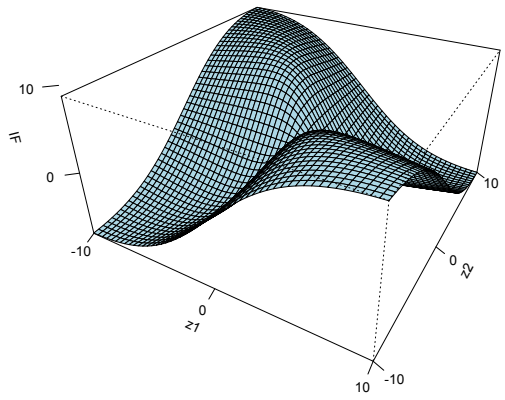
[36] Visuri, S., Ollila, E., Koivunen, V., Möttönen, J. and Oja, H. (2003). Affine equivariant multivariate rank methods. *Journal of Statistical Planning and Inference*, **114**, 161–185.

[37] Zhou, W. and Dang, X. (2010). Projection based scatter depth functions and associated scatter estimators. *Journal of Multivariate Analysis*, **101**, 138–153.

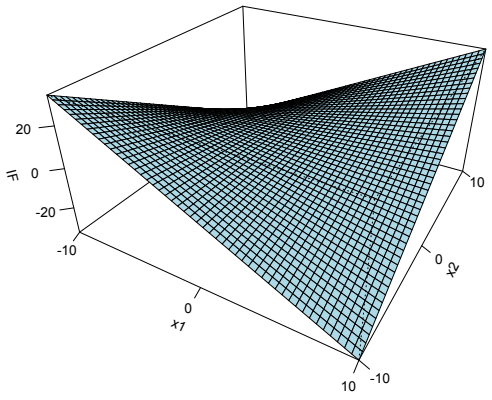
[38] Zuo, Y. and Cui, H. (2005). Depth weighted scatter estimators. *Annals of Statistics*, **33**, 381–413.



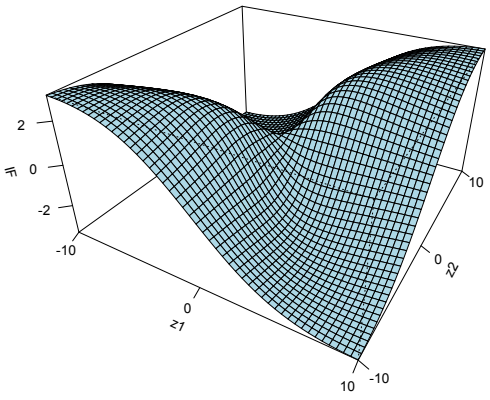
(a) IF, eigenvalue, classical covariance matrix



(b) IF, eigenvalue, modified spatial covariance matrix



(c) IF, eigenvector, classical covariance matrix



(d) IF, eigenvector, modified spatial covariance matrix

Figure 1: Influence functions of (a) the largest eigenvalue for the classical covariance matrix, (b) the largest eigenvalue for the modified spatial covariance matrix, (c) the first component of the eigenvector corresponding to the largest eigenvalue for the classical covariance matrix and (d) the first component of the eigenvector corresponding to the largest eigenvalue for the modified spatial covariance matrix at $F = N(\mathbf{0}, \mathrm{diag}(1, 4))$.