

Robust estimation of principal components from depth-based multivariate rank covariance matrix

Author(s)

Affiliation(s)

Abstract: Analyzing principal components for multivariate data from its spatial sign covariance matrix (SCM) has been proposed as a computationally simple and robust alternative to normal PCA, but it suffers from poor efficiency properties and is actually inadmissible with respect to the maximum likelihood estimator. Here we use data depth-based spatial ranks in place of spatial signs to obtain the orthogonally equivariant Depth Covariance Matrix (DCM) and use its eigenvector estimates for PCA. We derive asymptotic properties of the sample DCM and influence functions of its eigenvectors. The shapes of these influence functions indicate robustness of estimated principal components, and good efficiency properties compared to the SCM. Finite sample simulation studies show that principal components of the sample DCM are robust with respect to deviations from normality, as well as are more efficient than the SCM and its affine equivariant version, Tyler's shape matrix. Through two real data examples, we also show the effectiveness of DCM-based PCA in analyzing high-dimensional data and outlier detection, and compare it with other methods of robust PCA.

Key words and phrases: Data depth; Principal components analysis; Robustness; Sign covariance matrix; Multivariate ranking

1. Introduction

In multivariate analysis, the study of principal components is important since it provides a small number of uncorrelated variables from a potentially larger number of variables, so that these new components explain most of the underlying variability in the original data. In case of multivariate normal distribution, the sample covariance matrix provides the most asymptotically efficient estimates of eigenvectors/ principal components, but it is extremely sensitive to outliers as well as relaxations of the normality assumption. To address this issue, several robust estimators of the population covariance or correlation matrix have been proposed which can be used for Principal Components Analysis (PCA). They can be roughly put into these categories: robust, high breakdown point estimators that are computation-intensive [Rousseeuw, 1985, Maronna et al., 1976]; M-estimators that are calculated by simple iterative algorithms but do not necessarily possess high breakdown point [Huber, 1977, Tyler, 1987]; and symmetrised estimators that are highly efficient and robust to deviations from

normality, but sensitive to outliers and computationally demanding [Dümbgen, 1998, Sirkiä et al., 2007].

When principal components are of interest, one can also estimate the population eigenvectors by analyzing the spatial sign of a multivariate vector: the vector divided by its magnitude, instead of the original data. The covariance matrix of these sign vectors, namely Sign Covariance Matrix (SCM) has the same set of eigenvectors as the covariance matrix of the original population, thus the multivariate sign transformation yields computationally simple and high-breakdown estimates of principal components [Locantore et al., 1999, Visuri et al., 2000]. Although the SCM is not affine equivariant, its orthogonal equivariance suffices for the purpose of PCA. However, the resulting estimates are not very efficient, and are in fact asymptotically inadmissible [Magyar and Tyler, 2014], in the sense that there is an estimator (Tyler’s M-estimate of scatter, to be precise) that has uniformly lower asymptotic risk than the SCM.

The nonparametric concept of data-depth had first been proposed by Tukey [1975] when he introduced the halfspace depth. Given a dataset, the depth of a given point in the sample space measures how far inside the data cloud the point exists. An overview of statistical depth functions can be found in [Zuo and Serfling, 2000]. Depth-based methods have recently been popular for robust nonparametric classification [Jornsten, 2004, Ghosh and Chaudhuri, 2005, Dutta and Ghosh, 2012, Sguera et al., 2014]. In parametric estimation, depth-weighted means [Zuo et al., 2004] and covariance matrices [Zuo and Cui, 2005] provide high-breakdown point as well as efficient estimators, although they do involve choice of a suitable weight function and tuning parameters. In this paper we study the covariance matrix of the multivariate rank vector that is obtained from the data-depth of a point and its spatial sign, paying special attention to its eigenvectors. Specifically, we develop a robust version of principal components analysis for elliptically symmetric distributions based on the eigenvectors of this covariance matrix, and compare it with normal PCA and spherical PCA, i.e. PCA based on eigenvectors of the SCM.

The paper is arranged in the following fashion. Section 2 provides preliminary theoretical concepts required for developments in the subsequent sections. Section 3 introduces the Depth Covariance Matrix (DCM) and states some basic results related to this. Section 4 provides asymptotic results regarding the sample DCM, calculated using data depths with respect to the empirical distribution function, as well as its eigenvectors and eigenvalues. Section 5 focuses solely on principal component estimation using the sample DCM. We obtain influence functions and asymptotic efficiencies for eigenvectors of the DCM. We also compare their finite sample efficiencies for several multinormal and multivariate t -distributions with those of the SCM, Tyler’s scatter matrix and its

depth-weighted version through a simulation study. Section 6 presents two applications of the methods we develop on real data. Finally, we wrap up our discussion in Section 7 by giving a summary of our findings and providing some potential future areas of research. Proofs of all the results we derive are given in the appendix.

2. Preliminaries

Given a vector $\mathbf{x} \in \mathbb{R}^p$, its spatial sign is defined as the vector valued function [Locantore et al., 1999]:

$$\mathbf{S}(\mathbf{x}) = \begin{cases} \mathbf{x} \|\mathbf{x}\|^{-1} & \text{if } \mathbf{x} \neq \mathbf{0} \\ \mathbf{0} & \text{if } \mathbf{x} = \mathbf{0} \end{cases}$$

When \mathbf{x} is a random vector that follows an elliptic distribution $|\Sigma|^{-1/2} f((\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}))$, with a mean vector $\boldsymbol{\mu}$ and covariance matrix Σ , the sign vectors $\mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$ reside on the surface of a p -dimensional unit ball centered at $\boldsymbol{\mu}$. Denote by $\Sigma_S(\mathbf{X}) = E\mathbf{S}(\mathbf{X} - \boldsymbol{\mu})\mathbf{S}(\mathbf{X} - \boldsymbol{\mu})^T$ the covariance matrix of spatial signs, or the *Sign Covariance Matrix* (SCM). The transformation $\mathbf{x} \mapsto \mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$ keeps eigenvectors of population covariance matrix unchanged, and eigenvectors of the sample SCM $\hat{\Sigma}_S = \sum_{i=1}^n \mathbf{S}(\mathbf{x}_i - \boldsymbol{\mu})\mathbf{S}(\mathbf{x}_i - \boldsymbol{\mu})^T / n$ are \sqrt{n} -consistent estimators of their population counterparts [Taskinen et al., 2012].

The sign transformation is rotation equivariant, i.e. $\mathbf{S}(P(\mathbf{x} - \boldsymbol{\mu})) = P(\mathbf{x} - \boldsymbol{\mu}) / \|P(\mathbf{x} - \boldsymbol{\mu})\| = P(\mathbf{x} - \boldsymbol{\mu}) / \|\mathbf{x} - \boldsymbol{\mu}\| = P\mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$ for any orthogonal matrix P , and as a result the SCM is rotation equivariant too, in the sense that $\Sigma_S(P\mathbf{X}) = P\Sigma_S(\mathbf{X})P^T$. This is not necessarily true in general if P is replaced by any non-singular matrix. An affine equivariant version of the sample SCM is obtained as the solution $\hat{\Sigma}_T$ of the following equation:

$$\hat{\Sigma}_T(\mathbf{X}) = \frac{p}{n} \sum_{i=1}^n \frac{(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T}{(\mathbf{x}_i - \boldsymbol{\mu})^T \hat{\Sigma}_T(\mathbf{X})^{-1} (\mathbf{x}_i - \boldsymbol{\mu})}$$

which turns out to be Tyler's M-estimator of scatter [Tyler, 1987]. In this context, one should note that for scatter matrices, affine equivariance will mean any affine transformation on the original random variable $\mathbf{X} \mapsto \mathbf{X}^* = A\mathbf{X} + \mathbf{b}$ (A non-singular, $\mathbf{b} \in \mathbb{R}^p$) being carried over to the covariance matrix estimate upto a scalar multiple: $\hat{\Sigma}_T(\mathbf{X}^*) = k.A\hat{\Sigma}_T(\mathbf{X})A^T$ for some $k > 0$.

2.1 Data depth and outlyingness

For any multivariate distribution $F = F_{\mathbf{X}}$, the depth of a point $\mathbf{x} \in \mathbb{R}^p$, say $D(\mathbf{x}, F_{\mathbf{X}})$ is any real-valued function that provides a 'center outward ordering' of \mathbf{x} with respect to F [Zuo and Serfling, 2000]. Liu [1990] outlines the desirable properties of a statistical depth function:

- (P1) Affine invariance: $D(A\mathbf{x} + \mathbf{b}, F_{A\mathbf{X}+\mathbf{b}}) = D(\mathbf{x}, F_{\mathbf{X}})$
- (P2) Maximality at center: $D(\boldsymbol{\theta}, F_{\mathbf{X}}) = \sup_{\mathbf{x} \in \mathbb{R}^p} D(\mathbf{x}, F_{\mathbf{X}})$ for $F_{\mathbf{X}}$ having center of symmetry $\boldsymbol{\theta}$. This point is called the *deepest point* of the distribution.
- (P3) Monotonicity with respect to deepest point: $D(\mathbf{x}; F_{\mathbf{X}}) \leq D(\boldsymbol{\theta} + a(\mathbf{x} - \boldsymbol{\theta}), F_{\mathbf{X}})$, $\boldsymbol{\theta}$ being deepest point of $F_{\mathbf{X}}$.
- (P4) Vanishing at infinity: $D(\mathbf{x}; F_{\mathbf{X}}) \rightarrow \mathbf{0}$ as $\|\mathbf{x}\| \rightarrow \infty$.

In (P2) the types of symmetry considered can be central symmetry, angular symmetry and halfspace symmetry. Also for multimodal probability distributions, i.e. distributions with multiple local maxima in their probability density functions, properties (P2) and (P3) are actually restrictive towards the formulation of a reasonable depth function that captures the shape of the data cloud. In our derivations that follow, we replace these two by a slightly weaker condition: (P2*) The *existence* of a maximal point, i.e. $\sup_{\mathbf{x} \in \mathbb{R}^p} D(\mathbf{x}, F_{\mathbf{X}}) < \infty$. We denote this point by $M_D(F)$.

A real-valued function measuring the outlyingness of a point with respect to the data cloud can be seen as the opposite of what data depth does. Indeed, such functions have been used to define several depth functions, for example simplicial volume depth, projection depth and L_p -depth. Here we give a general definition of such functions as a transformation on any depth function:

Definition 2.0.1. Given a random variable \mathbf{X} following a probability distribution F , and a depth function $D(., .)$, we define Htped of a point \mathbf{x} as: $\tilde{D}(\mathbf{x}, F) = h(d_{\mathbf{x}})$ as any function of the data depth $D(\mathbf{x}, F) = d_{\mathbf{x}}$ so that $h(d_{\mathbf{x}})$ is bounded, monotonically decreasing in $d_{\mathbf{x}}$ and $\sup_{\mathbf{x}} \tilde{D}(\mathbf{x}, F) < \infty$.

For a fixed depth function, there are several choices of a corresponding htped. We develop our theory assuming a general htped function, but for the plots and simulations, fix our htped as $\tilde{D}(\mathbf{x}, F) = M_D(F) - D(\mathbf{x}, F)$, i.e. simply subtract the depth of a point from the maximum possible depth over all points in sample space.

We will be using the following 3 measures of data-depth to obtain our DCMs and compare their performances:

- **Halfspace depth** (HD) [Tukey, 1975] is defined as the minimum probability of all halfspaces containing a point. In our notations,

$$HD(\mathbf{x}, F) = \inf_{\mathbf{u} \in \mathbb{R}^p; \mathbf{u} \neq \mathbf{0}} P(\mathbf{u}^T \mathbf{X} \geq \mathbf{u}^T \mathbf{x})$$

- **Mahalanobis depth** (MhD) [Liu et al., 1999] is based on the Mahalanobis distance of \mathbf{x} to $\boldsymbol{\mu}$ with respect to Σ : $d_{\Sigma}(\mathbf{x}, \boldsymbol{\mu}) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}$. It is defined as

$$MhD(\mathbf{X}, F) = \frac{1}{1 + d_{\Sigma}^2(\mathbf{x} - \boldsymbol{\mu})}$$

note here that $d_{\Sigma}(\mathbf{x}, \boldsymbol{\mu})$ can be seen as a valid htped function of \mathbf{x} with respect to F .

- **Projection depth** (PD) [Zuo, 2003] is another depth function based on an outlyingness function. Here that function is

$$O(\mathbf{x}, F) = \sup_{\|\mathbf{u}\|=1} \frac{|\mathbf{u}^T \mathbf{x} - m(\mathbf{u}^T \mathbf{X})|}{s(\mathbf{u}^T \mathbf{X})}$$

where m and s are some univariate measures location and scale, respectively. Given this the depth at \mathbf{x} is defined as $PD(\mathbf{x}, F) = 1/(1 + O(\mathbf{x}, F))$.

Computation-wise, MhD is easy to calculate since the sample mean and covariance matrix are generally used as estimates of μ and Σ , respectively. However this makes MhD less robust with respect to outliers. PD is generally approximated by taking maximum over a number of random projections. There have been several approaches for calculating HD. A recent unpublished paper [Rainer and Mozharovskiy, 2014] provides a general algorithm that computes exact HD in $O(n^{p-1} \log n)$ time. In this paper, we shall use inbuilt functions in the R package `fda.usc` for calculating the above depth functions.

3. Depth-based rank covariance matrix

Consider a vector-valued random variable $\mathbf{X} \in \mathbb{R}^p$. Data depth is as much a property of the random variable as it is of the underlying distribution, so for ease of notation while working with transformed random variables, from now on we shall be using $D_{\mathbf{X}}(\mathbf{x}) = D(\mathbf{x}, F)$ to denote the depth of a point \mathbf{x} . Now, given a depth function $D_{\mathbf{X}}(\mathbf{x})$ (equivalently, an htped function $\tilde{D}_{\mathbf{X}}(\mathbf{x}) = \tilde{D}(\mathbf{x}, F)$), transform the original random variable as: $\tilde{\mathbf{x}} = \tilde{D}_{\mathbf{X}}(\mathbf{x})\mathbf{S}(\mathbf{x} - \boldsymbol{\mu})$, $\mathbf{S}(\cdot)$ being the spatial sign functional. The transformed random variable $\tilde{\mathbf{X}}$ can be seen as the multivariate rank corresponding to \mathbf{X} (e.g. Serfling [2006]). Figure 1 gives an idea of how $\tilde{\mathbf{X}}$ is distributed when \mathbf{X} has a bivariate normal distribution. Compared to the spatial sign, which are distributed on the surface of p -dimensional unit ball centered at $\boldsymbol{\mu}$, these spatial ranks have the same direction as original data and reside *inside* the p -dimensional ball around $\boldsymbol{\mu}$ that has radius $M_D(F)$ (which, for the case of halfspace depth, equals 0.5).

Now consider the spectral decomposition for the covariance matrix of F : $\Sigma = \Gamma \Lambda \Gamma^T$, Γ being orthogonal and Λ diagonal with positive diagonal elements.

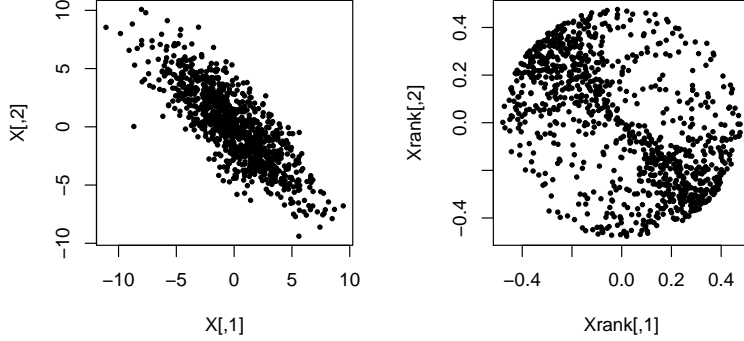


Figure 1: (Left) 1000 points randomly drawn from $\mathcal{N}_2((0,0)^T, \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix})$ and (Right) their multivariate ranks based on halfspace depth

Also normalize the original random variable as $\mathbf{z} = \Gamma^T \Lambda^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$. In this setup, we can represent the transformed random variable as

$$\begin{aligned}
 \tilde{\mathbf{x}} &= \tilde{D}_{\mathbf{X}}(\mathbf{x})\mathbf{S}(\mathbf{X} - \boldsymbol{\mu}) \\
 &= \tilde{D}_{\Gamma\Lambda^{1/2}\mathbf{z} + \boldsymbol{\mu}}(\Gamma\Lambda^{1/2}\mathbf{z} + \boldsymbol{\mu})\mathbf{S}(\Gamma\Lambda^{1/2}\mathbf{z}) \\
 &= \Gamma\tilde{D}_{\mathbf{Z}}(\mathbf{z})\mathbf{S}(\Lambda^{1/2}\mathbf{z}) \\
 &= \Gamma\Lambda^{1/2}\tilde{D}_{\mathbf{Z}}(\mathbf{z})\mathbf{S}(\mathbf{z})\frac{\|\mathbf{z}\|}{\|\Lambda^{1/2}\mathbf{z}\|}
 \end{aligned} \tag{3.1}$$

$\tilde{D}_{\mathbf{Z}}(\mathbf{z})$ is an even function in \mathbf{z} because of affine invariance, as is $\|\mathbf{z}\|/\|\Lambda^{1/2}\mathbf{z}\|$. Since $\mathbf{S}(\mathbf{z})$ is odd in \mathbf{z} for circularly symmetric \mathbf{z} , it follows that $E(\tilde{\mathbf{X}}) = \mathbf{0}$, and consequently we obtain an expression for the covariance matrix of $\tilde{\mathbf{X}}$:

Theorem 3.0.2. *Let the random variable $\mathbf{X} \in \mathbb{R}^p$ follow an elliptical distribution with center $\boldsymbol{\mu}$ and covariance matrix $\Sigma = \Gamma\Lambda\Gamma^T$, its spectral decomposition. Then, given a depth function $D_{\mathbf{X}}(\cdot)$ the covariance matrix of the transformed random variable $\tilde{\mathbf{X}}$ is*

$$\text{Cov}(\tilde{\mathbf{X}}) = \Gamma\Lambda_{D,S}\Gamma^T, \quad \text{with} \quad \Lambda_{D,S} = \mathbb{E}_{\mathbf{Z}} \left[(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 \frac{\Lambda^{1/2}\mathbf{z}\mathbf{z}^T\Lambda^{1/2}}{\mathbf{z}^T\Lambda\mathbf{z}} \right] \tag{3.2}$$

where $\mathbf{z} = (z_1, \dots, z_p)^T \sim N(\mathbf{0}, I_p)$, so that $\Lambda_{D,S}$ a diagonal matrix with diagonal entries

$$\lambda_{D,S,i} = \mathbb{E}_{\mathbf{Z}} \left[\frac{(\tilde{D}_{\mathbf{Z}}(\mathbf{z}))^2 \lambda_i z_i^2}{\sum_{j=1}^p \lambda_j z_j^2} \right]$$

The matrix of eigenvectors of the covariance matrix, Γ , remains unchanged in the transformation $\mathbf{X} \rightarrow \tilde{\mathbf{X}}$. As a result, the multivariate rank vectors can be used for robust principal component analysis, which will be outlined in the following sections. However, as one can see in the above expression, the diagonal entries of $\Lambda_{D,S}$ do not change if a scale change is done on all entries of Λ , meaning the $\Lambda_{D,S}$ matrices corresponding to F and cF for some $c \neq 0$ will be same. This is the reason for lack of affine equivariance of the DCM. Following the case of multivariate sign covariance matrices [Taskinen et al., 2012] one can get back the shape components, i.e. original *standardized* eigenvalues Λ^* from $\Lambda_{D,S}$ by an iterative algorithm:

1. Set $k = 0$, and start with an initial value $\Lambda^{*(0)}$.
2. Calculate the next iterate

$$\Lambda^{*(k+1)} = \left[\mathbb{E}_{\mathbf{z}} \left(\frac{(\tilde{D}_{\mathbf{z}}(\mathbf{z}))^2 \mathbf{z} \mathbf{z}^T}{\mathbf{z}^T \Lambda^{*(k)} \mathbf{z}} \right) \right]^{-1} \Lambda_{D,S}$$

and standardize its eigenvalues:

$$\Lambda^{*(k+1)} = \frac{\Lambda^{*(k+1)}}{\det(\Lambda^{*(k+1)})^{1/p}}$$

3. Stop if convergence criterion is satisfied. Otherwise set $k \rightarrow k + 1$ and go to step 2.

Unlike sign covariance matrices and symmetrized sign covariance matrices [Dümbgen, 1998], however, attempting to derive an affine equivariant counterpart (as opposed to only orthogonal equivariance) of the DCM through an iterative approach analogous to Tyler [1987] will not result in anything new. This is because Tyler's scatter matrix Σ_T is defined as the implicit solution to the following equation:

$$\Sigma_T = \mathbb{E} \left[\frac{(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T}{(\mathbf{x} - \boldsymbol{\mu})^T \Sigma_T^{-1} (\mathbf{x} - \boldsymbol{\mu})} \right] \quad (3.3)$$

and simply replacing \mathbf{x} by its multivariate rank counterpart $\tilde{\mathbf{x}}$ will not change the estimate Σ_T as \mathbf{x} and $\tilde{\mathbf{x}}$ have the same directions. Instead we consider a depth-weighted version of Tyler's scatter matrix (i.e. weights $(\tilde{D}_{\mathbf{x}}(\mathbf{x}))^2$ in right side of (3.3)) in the simulations in Section ???. The simulations show that it has slightly better finite-sample efficiency than Σ_T but has same asymptotic performance. We conjecture that its concentration properties can be obtained by taking an approach similar to Soloveychik and Wiesel [2014].

$$\text{The 1st display equation of the second section.} \quad (3.4)$$

The 2nd display equation of the second section. (3.5)

3. The Third Section

Contents of the third section.

Contents of the third section.

Contents of the third section.

Contents of the third section.

Contents of the third section.

The 1st display equation of the third section. (3.1)

The 2nd display equation of the third section. (3.2)

Table 1: Caption of the table.

Bandwidth		$h = 3$			$h = 4$			$h = 5$		
Estimates		\hat{N}_H^{LC}	\hat{N}_H^{LC}	\hat{N}_H^{LL}	\hat{N}_H^{LC}	\hat{N}_H^{LC}	\hat{N}_H^{LL}	\hat{N}_H^{LC}	\hat{N}_H^{LC}	\hat{N}_H^{LL}
$beta(10, 10)$	BIAS	-22.5	-14.8	14.0	-13.3	-6.9	12.5	-8.2	-4.7	11.5
$\bar{p} = 0.500$	S.E.	13.8	14.9	12.1	12.6	14.6	11.7	15.1	19.3	15.5
$cv = 0.218$	RMSE	26.4	21.0	18.5	18.3	16.2	17.1	17.2	19.9	19.3
$beta(5, 5)$	BIAS	-32.2	-19.9	5.4	-21.9	-11.8	4.9	-15.6	-8.4	5.1
$\bar{p} = 0.500$	S.E.	15.8	16.2	12.3	14.3	19.8	12.9	16.4	19.8	14.6
$cv = 0.302$	RMSE	35.9	25.7	13.4	26.1	15.9	11.9	22.6	21.5	15.4
$beta(4, 8)$	BIAS	-53.7	-29.1	-10.8	-42.0	-19.3	-8.4	-34.4	-13.9	-7.6
$\bar{p} = 0.333$	S.E.	20.8	19.4	18.1	18.5	18.8	15.6	19.2	20.6	16.4
$cv = 0.392$	RMSE	57.6	34.9	21.1	45.9	26.9	17.8	39.4	24.9	18.1
$beta(3, 5)$	BIAS	-57.4	-32.4	-15.5	-45.6	-22.8	-13.1	-37.4	-17.1	-11.4
$\bar{p} = 0.375$	S.E.	21.2	19.9	17.1	18.9	18.8	15.0	19.6	21.6	15.6
$cv = 0.430$	RMSE	61.2	38.1	23.1	49.4	29.5	19.9	42.3	27.5	19.3

Supplementary Materials

Contain the brief description of the online supplementary materials.

Acknowledgements

Write the acknowledgements here.

References

- L. Dümbgen. On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.*, 50:471–491, 1998.
- S. Dutta and A.K. Ghosh. On robust classification using projection depth. *Ann. Inst. Stat. Math.*, 64-3:657–676, 2012.
- A.K. Ghosh and P. Chaudhuri. On Maximum Depth and Related Classifiers. *Scand. J. Statist.*, 32:327–350, 2005.
- P.J. Huber. *Robust Statistical Procedures*. Society for Industrial and Applied Mathematics, Philadelphia, 1977.
- R. Jörnsten. Clustering and classification based on the l_1 depth. *J. Mult. Anal.*, 90-1:67–89, 2004.
- R.Y. Liu. On a notion of data depth based on random simplices. *Ann. of Statist.*, 18:405–414, 1990.
- R.Y. Liu, J.M. Parelius, and K. Singh. Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Ann. of Statist.*, 27:783–858, 1999.
- N. Locantore, J.S. Marron, D.G. Simpson, N. Tripoli, J.T. Zhang, and K.L. Cohen. Robust principal components of functional data. *TEST*, 8:1–73, 1999.

- A.F. Magyar and D.E. Tyler. The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions. *Biometrika*, 101:673–688, 2014.
- R.A. Maronna, W.A. Staehl, and V.Y. Yohai. Bias-Robust Estimators of Multivariate Scatter Based on Projections. *J. Mult. Anal.*, 42:141–161, 1976.
- D. Rainer and P. Mozharovskyi. Exact computation of the halfspace depth. <http://arxiv.org/abs/1411.6927>, 2014.
- P. Rousseeuw. Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications, Volume B (Proc. 4th Pannonian Symp. Math. Statist., Bad Tatzmannsdorf, 1983)*, pages 283–97, Dordrecht, 1985. D. Reidel.
- R. Serfling. Depth Functions in Nonparametric Multivariate Inference. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 72, pages 1–16, 2006.
- C. Sguera, P. Galeano, and R. Lillo. Spatial depth-based classification for functional data. *TEST*, 23-4:725–750, 2014.
- S. Sirkiä, S. Taskinen, and H. Oja. Symmetrised M-estimators of scatter. *J. Mult. Anal.*, 98: 1611–1629, 2007.
- I. Soloveychik and A. Wiesel. Performance Analysis of Tyler’s Covariance Estimator. *IEEE Trans. Sig. Process.*, 63-2:418–426, 2014.
- S. Taskinen, I. Koch, and H. Oja. Robustifying principal component analysis with spatial sign vectors. *Statist. and Prob. Lett.*, 82:765–774, 2012.
- J.W. Tukey. Mathematics and picturing data. In R.D. James, editor, *Proceedings of the International Congress on Mathematics*, volume 2, pages 523–531, 1975.
- D.E. Tyler. A distribution-free M-estimator of multivariate scatter. *Ann. Statist.*, 15:234–251, 1987.
- S. Visuri, V. Koivunen, and H. Oja. Sign and rank covariance matrices. *J. Statist. Plan. Inf.*, 91:557–575, 2000.
- Y. Zuo. Projection-based depth functions and associated medians. *Ann. Statist.*, 31:1460–1490, 2003.
- Y. Zuo and M. Cui. Depth weighted scatter estimators. *Ann. Statist.*, 33-1:381–413, 2005.
- Y. Zuo and R. Serfling. General notions of statistical depth functions. *Ann. Statist.*, 28-2: 461–482, 2000.
- Y. Zuo, M. Cui, and X. He. On the Staehl-Donoho estimator and depth-weighted means of multivariate data. *Ann. Statist.*, 32-1:167–188, 2004.

first author affiliation

E-mail: (first author email)

second author affiliation

E-mail: (second author email)