

k -step shape estimators based on spatial signs and ranks

S. Taskinen^{a,*}, S. Sirkiä^b, H. Oja^c

^a*Department of Mathematics and Statistics, University of Jyväskylä*

^b*Finnish Forest Research Institute, Joensuu*

^c*Tampere School of Public Health, University of Tampere*

Abstract

In this paper, the shape matrix estimators based on spatial sign and rank vectors are considered. The estimators considered here are slight modifications of the estimators introduced in Dümbgen (1998) and Oja and Randles (2004) and further studied for example in Sirkiä et al. (2009). The shape estimators are computed using pairwise differences of the observed data, therefore there is no need to estimate the location center of the data. When the estimator is based on signs, the use of differences also implies that the estimators have the so called independence property if the estimator, that is used as an initial estimator, has it. The influence functions and limiting distributions of the estimators are derived at the multivariate elliptical case. The estimators are shown to be highly efficient in the multinormal case, and for heavy-tailed distributions they outperform the shape estimator based on sample covariance matrix.

Key words: Affine equivariance, efficiency, influence function, spatial sign, spatial rank

1. Introduction and some notations

A p -variate random vector \mathbf{x} is elliptically symmetric if its density function is of the form

$$f(\mathbf{x}) = |\Sigma|^{-1/2} g[(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})], \quad (1)$$

*Corresponding author: Sara Taskinen, Department of Mathematics and Statistics, P.O. Box 35, University of Jyväskylä, FIN-40014 Jyväskylä, Finland; email: sara.l.taskinen@maths.jyu.fi

where a p -vector $\boldsymbol{\mu}$ denotes the location of the distribution and a positive definite symmetric $p \times p$ matrix (PDS(p)) Σ , the so called scatter matrix, defines the shape and scale of its contours. If \boldsymbol{x} is a random variable from an elliptical distribution, then the standardized variable $\boldsymbol{z} = \Sigma^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})$ has a spherical distribution with location $\mathbf{0}$, and \boldsymbol{z} can be decomposed as $\boldsymbol{z} = r\boldsymbol{u}$, where $r = \|\boldsymbol{z}\|$ is the Euclidean length of \boldsymbol{z} , and r and $\boldsymbol{u} = \|\boldsymbol{z}\|^{-1}\boldsymbol{z}$ are independent with \boldsymbol{u} being uniformly distributed on the unit sphere.

The scatter matrix Σ can be decomposed into two parts, that is, $\Sigma = \sigma^2\Lambda$, where σ^2 is the scale parameter and PDS(p) matrix Λ with $Tr(\Lambda) = p$ is the shape matrix. Note that the condition $Tr(\Lambda) = p$ is sometimes replaced by $Det(\Lambda) = 1$ or $\Lambda_{11} = 1$ (Paindaveine, 2008). A shape matrix includes information only of the shape and orientation of the data and in several applications it is enough to estimate the shape matrix only.

In this paper, we denote the scatter matrix functional as $C(\boldsymbol{x})$, where \boldsymbol{x} is a random vector with cumulative distribution function F . $C(\boldsymbol{x})$ is defined to be a scatter matrix if it is PDS(p) and affine equivariant so that

$$C(A\boldsymbol{x} + \boldsymbol{b}) = AC(\boldsymbol{x})A^T, \quad (2)$$

for every nonsingular $p \times p$ matrix A and p -vector \boldsymbol{b} . Applying scatter matrix functional to the empirical distribution function F_n gives the scatter matrix estimate \widehat{C} . Further, the shape matrix functional is denoted by $V(\boldsymbol{x})$ and it is PDS(p) and affine equivariant in the sense that

$$V(A\boldsymbol{x} + \boldsymbol{b}) = \frac{p}{Tr(AV(\boldsymbol{x})A^T)}AV(\boldsymbol{x})A^T. \quad (3)$$

The corresponding estimate \widehat{V} is again obtained by applying shape functional to the empirical distribution function F_n . Note that in the elliptical model, different scatter matrices are not comparable, since a correction factor is needed in order to guarantee the Fisher consistency towards Σ . All shape matrices, however, estimate the same population quantity Λ and are directly comparable without any modifications.

Tyler (1987) introduced a simple shape matrix estimator based on spatial sign vectors. His estimator is distribution-free under elliptically symmetric distributions and highly efficient in case of heavy-tailed distributions. Under multivariate normal model the efficiencies are, however, quite low. More efficient estimators are obtained by using pairwise differences instead of the original observations in Tyler's approach (Dümbgen, 1998), or spatial rank

vectors as in Oja and Randles (2004) and Sirkiä et al. (2009). Since the Dümbgen’s shape matrix and the spatial rank covariance matrix use the pairwise differences of the observed data, there is no need to estimate the location center. The use of differences in Dümbgen’s estimator also implies that the estimator has important independence property. It means that $V(\mathbf{x})$ is a diagonal matrix for all \mathbf{x} with independent components. This property is useful in several multivariate analysis problems, for example in independent component analysis (Oja et al., 2006; Tyler et al., 2009).

The estimators derived in Dümbgen (1998) and Oja and Randles (2004) are computed using iterative algorithms. Due to this property and the use of pairwise differences, the estimation procedure is rather time-consuming when the sample size is very large. In this paper, we therefore consider so called k -step versions of these estimators, that is, the estimators obtained after k -th iteration step. Similar estimators have been considered for example in Oja et al. (2006) and Tyler et al. (2009), where 1-step M-estimators and 1-step W-estimators were used in invariant coordinate selection and independent component analysis. Other examples include 1-step signed-rank scatter matrix that was introduced by Hallin et al. (2006), and 1-step Tyler’s shape matrix that was used in the context of principal axis analysis by Chritchley et al. (2008).

The outline of the paper is as follows. In Section 2, the corresponding estimators are defined and in Sections 3 and 4, the robustness and efficiency properties of them are studied. In Section 5, the properties of the estimators are examined using simple simulation studies. The paper is concluded with some final comments in Section 6. All the proofs are found in the Appendix.

2. Shape estimators based on spatial sign and rank vectors

The spatial sign function is defined as

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

where $\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2}$ is the Euclidean length of the vector \mathbf{x} . Applying spatial sign function to the data points $\mathbf{x}_1, \dots, \mathbf{x}_n$ produces spatial sign vectors $\mathbf{S}(\mathbf{x}_i)$. The centered spatial rank function is defined as

$$\mathbf{R}(\mathbf{x}) = ave\{\mathbf{S}(\mathbf{x} - \mathbf{x}_i)\},$$

thus unlike the sign function, the rank function depends on the data. If $\mathbf{x}_1, \dots, \mathbf{x}_n$ is distributed according to F , then $\mathbf{R}(\mathbf{x})$ converges uniformly in probability to the theoretical rank function $\mathbf{R}_F(\mathbf{x}) = E_F[\mathbf{S}(\mathbf{x} - \mathbf{x}_i)]$ (Möttönen et al., 1995). At spherical distribution,

$$\mathbf{R}_F(r\mathbf{u}) = q_F(r)\mathbf{u}, \quad (4)$$

for some bounded increasing function $q_F(r)$. In the following, the spatial rank vectors are denoted by $\mathbf{R}(\mathbf{x}_i) = \text{ave}_j \mathbf{S}(\mathbf{x}_i - \mathbf{x}_j)$.

In Visuri et al. (2000), the symmetrised spatial sign covariance matrix and the spatial rank covariance matrix were introduced. These estimates are given by

$$\widehat{C}_S = \text{ave}\{\mathbf{S}(\mathbf{x}_i - \mathbf{x}_j)\mathbf{S}^T(\mathbf{x}_i - \mathbf{x}_j)\}$$

and

$$\widehat{C}_R = \text{ave}\{\mathbf{R}(\mathbf{x}_i)\mathbf{R}^T(\mathbf{x}_i)\} = \text{ave}\{\mathbf{S}(\mathbf{x}_i - \mathbf{x}_j)\mathbf{S}^T(\mathbf{x}_i - \mathbf{x}_k)\}.$$

Since the estimates are based on differences, they can be computed without the need for the location estimate $\widehat{\boldsymbol{\mu}}$. The main disadvantage is that these estimates are only orthogonal equivariant, that is, they satisfy (2) only for orthogonal $p \times p$ matrices A . The estimators are therefore not genuine scatter matrix estimators.

To derive affine equivariant versions of these estimates, one has to proceed as in Hettmansperger and Randles (2003) and apply the spatial sign and rank functions to the transformed data points. Let V be a PDS(p) matrix with $\text{Tr}(V) = p$ and write $\mathbf{z}_i = V^{-1/2}\mathbf{x}_i$, $i = 1, \dots, n$, for the standardized observations. Then the affine equivariant version of the symmetrised spatial sign covariance matrix is such \widehat{V} that satisfies

$$p \text{ave}\{\mathbf{S}(\mathbf{z}_i - \mathbf{z}_j)\mathbf{S}^T(\mathbf{z}_i - \mathbf{z}_j)\} = I_p, \quad (5)$$

and for which $\text{Tr}(\widehat{V}) = p$. The resulting estimate is also known as the Dümbgen's estimate (Dümbgen, 1998) and it is a symmetrised version of the Tyler's M-estimate (Tyler, 1987) that is found as a solution to $p \text{ave}\{\mathbf{S}(\mathbf{z}_i)\mathbf{S}^T(\mathbf{z}_i)\} = I_p$. The properties of the Dümbgen's estimator are studied in Dümbgen (1998), Dümbgen and Tyler (2005) and Sirkiä et al. (2009) among others. The estimating equation (5) yields to the iteration step

$$\widehat{V}_{k+1} \leftarrow \widehat{V}_k^{1/2} \text{ave}\{\mathbf{S}(\mathbf{z}_i - \mathbf{z}_j)\mathbf{S}^T(\mathbf{z}_i - \mathbf{z}_j)\} \widehat{V}_k^{1/2}, \quad (6)$$

which is repeated until the sequence converges. In the end, the resulting estimate is scaled so that $Tr(\widehat{V}) = p$. This algorithm is similar to that considered in Tyler (1987), only computed using pairwise differences instead of the original observations. For the proof of convergence, see Tyler (1987).

The affine equivariant version of the spatial rank covariance matrix is obtained by solving

$$p \text{ ave}\{\mathbf{R}(\mathbf{z}_i)\mathbf{R}^T(\mathbf{z}_i)\} = \text{ave}\{\mathbf{R}^T(\mathbf{z}_i)\mathbf{R}(\mathbf{z}_i)\} I_p. \quad (7)$$

The resulting estimate is again scaled so that $Tr(\widehat{V}) = p$. This estimate is considered for example in Oja and Randles (2004) and in Sirkiä et al. (2009). An iteration step suggested by the estimating equation (7) is now

$$\widehat{V}_{k+1} \leftarrow \widehat{V}_k^{1/2} \text{ave}\{\mathbf{R}(\mathbf{z}_i)\mathbf{R}^T(\mathbf{z}_i)\} \widehat{V}_k^{1/2}. \quad (8)$$

Note that the above equation does not include any standardization term, since the resulting estimate is scaled to have trace equal to p . The algorithm (8) seems always to converge in practice, however, the proof for this is still an open question.

In Sirkiä et al. (2009), it is shown that the above estimates are highly efficient even in the multinormal case. The main drawback is that since these estimates are based on pairwise differences, they are rather slow to compute when the sample size is very large. Motivated by this computational inconvenience and the fact that there is no proof for the convergence of (8) so far, we consider in the following so called k -step versions of the estimates. These are simply obtained by starting with some \sqrt{n} -consistent shape matrix estimate, for example Tyler's M-estimate, and repeating steps in (6) and (8) k times.

Definition 1. Let \widehat{V}_0 be some \sqrt{n} -consistent shape matrix estimate, which is used as a starting value. The k -step sign estimate is then computed with

1. $\mathbf{z}_i \leftarrow \widehat{V}_{k-1}^{-1/2} \mathbf{x}_i$, for $i = 1, \dots, n$
2. $\widehat{V}_k \leftarrow \widehat{V}_{k-1}^{1/2} \text{ave}\{\mathbf{S}(\mathbf{z}_i - \mathbf{z}_j)\mathbf{S}^T(\mathbf{z}_i - \mathbf{z}_j)\} \widehat{V}_{k-1}^{1/2}$,

and then standardized so that $Tr(\widehat{V}_k) = p$.

Definition 2. Let \widehat{V}_0 be some \sqrt{n} -consistent shape matrix estimate, which is used as a starting value. The k -step rank estimate is then computed with

1. $\mathbf{z}_i \leftarrow \widehat{V}_{k-1}^{-1/2} \mathbf{x}_i$, for $i = 1, \dots, n$
2. $\widehat{V}_k \leftarrow \widehat{V}_{k-1}^{1/2} \text{ave}\{\mathbf{R}(\mathbf{z}_i)\mathbf{R}^T(\mathbf{z}_i)\} \widehat{V}_{k-1}^{1/2}$,

and then standardized so that $\text{Tr}(\widehat{V}_k) = p$.

To compute influence functions of k -step estimators, we need to define the corresponding functionals. Write $V_k = V_k(\mathbf{x})$, where $\mathbf{x} \sim F$, for the functional corresponding to k -step sign estimate \widehat{V}_k . Then V_k is given by the implicit equation

$$V_k = p V_{k-1}^{1/2} E_F [\mathbf{S}(\mathbf{z}_1 - \mathbf{z}_2)\mathbf{S}^T(\mathbf{z}_1 - \mathbf{z}_2)] V_{k-1}^{1/2}, \quad (9)$$

where $\mathbf{z}_i = V_{k-1}^{-1/2} \mathbf{x}_i$, for $i = 1, 2$, and \mathbf{x}_i 's are independent and distributed according to F .

For k -step rank estimator, the functional corresponding to the estimator is given by

$$V_k = p E_F^{-1}[\mathbf{R}_F^T(\mathbf{z}_1)\mathbf{R}_F(\mathbf{z}_1)] V_{k-1}^{1/2} E_F[\mathbf{R}_F(\mathbf{z}_1)\mathbf{R}_F^T(\mathbf{z}_1)] V_{k-1}^{1/2} \quad (10)$$

where again $\mathbf{z}_i = V_{k-1}^{-1/2} \mathbf{x}_i$ and \mathbf{x}_i 's are independently distributed according to F . Note that to guarantee that $\text{Tr}(V_k) = p$, (10) now includes a scaling term.

The above k -step estimators inherit now some important properties of the initial estimator. Estimators are naturally affine equivariant shape estimators and satisfy (3), as the initial estimator V_0 is affine equivariant. The independence property of k -step sign estimators is also inherited from the initial estimator as stated in the following theorem. It is still unclear whether the rank estimators have the same property.

Theorem 1. Assume that \mathbf{x} is a random p -vector with independent components and $V_0(\mathbf{x})$ has the independence property, that is, $V_0(\mathbf{x})$ is diagonal. Then the k -step sign functional has the independence property.

In the following sections, we examine the robustness and efficiency properties of the k -step estimators. It is shown that when for example Tyler's M-estimate is used as a starting value, then even the resulting 1-step estimators are highly robust and efficient shape matrix estimators.

3. Influence functions

The robustness of a functional T against a single outlier \mathbf{x} can be measured using the influence functions (Hampel et al., 1986). Let

$$F_\epsilon = (1 - \epsilon)F + \epsilon\Delta_{\mathbf{x}},$$

denote the contaminated distribution, where $\Delta_{\mathbf{x}}$ is the cdf of a distribution with probability mass one at point \mathbf{x} . The influence function of T is then given by

$$IF(\mathbf{x}; T, F) = \lim_{\epsilon \rightarrow 0} \frac{T(F_\epsilon) - T(F)}{\epsilon}.$$

Hampel et al. (1986) showed that, for any scatter functional $C(F)$, the influence function of C at a spherical F_0 , symmetric around the origin and with $C(F_0) = I_k$, is given by

$$IF(\mathbf{x}; C, F_0) = \alpha_C(r)\mathbf{u}\mathbf{u}^T - \beta_C(r)I_k,$$

where $r = \|\mathbf{x}\|$ and $\mathbf{u} = \|\mathbf{x}\|^{-1}\mathbf{x}$, and weight functions α_C and β_C depend on the functional and the distribution F_0 . For shape functionals $V(F)$, the influence function reduces to

$$IF(\mathbf{x}; V, F_0) = \alpha_V(r) \left[\mathbf{u}\mathbf{u}^T - \frac{1}{p}I_p \right], \quad (11)$$

where α_V is again a weight function that includes all information of the influence of \mathbf{x} on the estimator. For robust estimators, α_V is continuous and bounded and for comparisons between different shape estimators, we only need to compare functions α_V .

The influence function of k -step sign functional defined in (9) is now given by the following theorem.

Theorem 2. *At spherical F_0 , the influence function of k -step sign functional V_k with initial shape matrix V_0 is given by (11) with*

$$\alpha_{V_k}(r) = c_p^k \alpha_{V_0}(r) + (1 - c_p^k) 2(p + 2)(1 - pg(r)),$$

where $c_p = 2/(p + 2)$ and

$$g(r) = E_{\mathbf{x}_1} \left[\frac{(\mathbf{x}_1 - r\mathbf{e}_1)_2^2}{\|\mathbf{x}_1 - r\mathbf{e}_1\|^2} \right],$$

where $\mathbf{e}_1 = (1, 0, \dots, 0)^T$, $r = \|\mathbf{x}\|$ and \mathbf{x} and \mathbf{x}_1 are independent and spherically distributed.

Note that when $k \rightarrow \infty$, the weight function reduces to $\alpha_V(r) = 2(p+2)(1-pg(r))$, that is, the weight function of the regular Dümbgen's shape matrix (Sirkiä et al., 2009).

For k -step rank functional defined in (10) the following result holds:

Theorem 3. *At spherical F_0 , the influence function of k -step rank functional V_k with initial shape matrix V_0 is given by (11) with*

$$\alpha_{V_k}(r) = c_p^k \alpha_{V_0}(r) + (1 - c_p^k) \frac{p+2}{c_F^2} (g_1(r) - g_2(r)),$$

where $c_p = 2/(p+2)$, $c_F^2 = E_F[q_F^2(r)]$ and $q_F(r)$ is defined in (4). Further,

$$g_1(r) = E_{\mathbf{x}_2, \mathbf{x}_3} \left[\frac{(r\mathbf{e}_1 - \mathbf{x}_2)_1 (r\mathbf{e}_1 - \mathbf{x}_3)_1}{\|r\mathbf{e}_1 - \mathbf{x}_2\| \|r\mathbf{e}_1 - \mathbf{x}_3\|} \right] + 2E_{\mathbf{x}_2, \mathbf{x}_3} \left[\frac{(\mathbf{x}_2 - r\mathbf{e}_1)_1 (\mathbf{x}_2 - \mathbf{x}_3)_1}{\|\mathbf{x}_2 - r\mathbf{e}_1\| \|\mathbf{x}_2 - \mathbf{x}_3\|} \right]$$

and

$$g_2(r) = E_{\mathbf{x}_2, \mathbf{x}_3} \left[\frac{(\mathbf{x}_2 - r\mathbf{e}_1)_2 (\mathbf{x}_3 - r\mathbf{e}_1)_2}{\|\mathbf{x}_2 - r\mathbf{e}_1\| \|\mathbf{x}_3 - r\mathbf{e}_1\|} \right] + 2E_{\mathbf{x}_2, \mathbf{x}_3} \left[\frac{(\mathbf{x}_2 - r\mathbf{e}_1)_2 (\mathbf{x}_2 - \mathbf{x}_3)_2}{\|\mathbf{x}_2 - r\mathbf{e}_1\| \|\mathbf{x}_2 - \mathbf{x}_3\|} \right],$$

where $\mathbf{e}_1 = (1, 0, \dots, 0)^T$, $r = \|\mathbf{x}\|$ and \mathbf{x} , \mathbf{x}_2 and \mathbf{x}_3 are independent and spherically distributed.

When $k \rightarrow \infty$, the influence function of regular rank covariance functional is obtained. In that case, $\alpha_V(r) = (p+2)(g_1(r) - g_2(r))/c_F^2$.

Figure 1 illustrates functions α_{V_k} at bivariate standard normal distribution case for k -step sign and rank estimators with $k = 1, 2, 3$ and ∞ and when the Tyler's M-estimator with $\alpha_{V_0} = p+2$ (first row) or the sample covariance matrix with $\alpha_{V_0} = r^2$ (second row) is used as a starting value.

As seen in the figure, when $k = 3$ and the Tyler's M-estimator is used as a starting value, the α_{V_k} functions are very similar to those obtained as $k \rightarrow \infty$. For all estimators, the functions α_{V_k} are clearly continuous. Influence functions are also bounded as stated in the following corollary. When the sample covariance matrix is used as an initial estimator, the influence functions of k -step estimators are naturally unbounded. However, as k increases, the resulting estimators are clearly more robust than the sample covariance matrix as the outlying observations do not have that much influence on the estimators.

Corollary 1. *For k -step sign and rank estimators, the functions α_{V_k} are bounded when α_{V_0} is bounded.*

4. Asymptotic distributions and efficiencies

In the next, we will denote

$$C_{p,p}(V) = (I_{p^2} + K_{p,p})(V \otimes V) - \frac{2}{p} \text{vec}(V) \text{vec}^T(V),$$

where $K_{p,p}$ is the commutation matrix, that is, a $p^2 \times p^2$ block matrix with (i, j) -block being equal to a $p \times p$ matrix that has 1 at entry (j, i) and zero elsewhere. By $\text{vec}(V)$ we mean the vectorization of a matrix V , which is obtained by stacking the columns of V on top of each other. If $V = I_p$, then

$$C_{p,p}(I_p) = I_{p^2} + K_{p,p} - \frac{2}{p} J_{p,p},$$

where $J_{p,p} = \text{vec}(I_p) \text{vec}(I_p)^T$.

The limiting distributions of k -step sign and rank estimators are given by the following theorem.

Theorem 4. *Let F_0 be a spherical distribution. Assume that the initial estimate \widehat{V}_0 is \sqrt{n} -consistent and follows the multinormal distribution. Then the limiting distribution of $\sqrt{n} \text{vec}(\widehat{V}_k - I_p)$ is also multinormal with mean $\mathbf{0}$ and asymptotic covariance matrix $\tau C_{p,p}(I_p)$, where*

$$\tau = ASV(\widehat{V}_{k,12}; F_0) = \frac{E_{F_0}[\alpha_{V_k}^2(r)]}{p(p+2)}$$

denotes the asymptotic variance of any off-diagonal element of \widehat{V}_k and functions $\alpha_{V_k}(r)$ are defined in Theorems 2 and 3.

The limiting distributions at elliptical case can be derived using the affine equivariance properties of the estimators.

Corollary 2. *Let F be an elliptical distribution with population shape parameter V . Then the limiting distribution of $\sqrt{n} \text{vec}(\widehat{V}_k - V)$ is multivariate normal with mean $\mathbf{0}$ and asymptotic covariance matrix*

$$\tau \left(I_{p^2} - \frac{1}{p} \text{vec}(V) \text{vec}(I_p)^T \right) (I_{p^2} + K_{p,p})(V \otimes V) \left(I_{p^2} - \frac{1}{p} \text{vec}(I_p) \text{vec}(V)^T \right),$$

where τ is given in Theorem 4.

Note that only the scalars τ are needed to characterize the limiting distributions of the shape estimators. Therefore, in order to compare asymptotic efficiencies of different estimators, one only has to compare these scalars.

In Table 1, we list the asymptotic relative efficiencies of k -step sign and rank estimators with $k = 1, 2, 3$ and ∞ with respect to the regular shape estimator. By regular shape estimator, we mean the estimator, which is obtained by standardizing the sample covariance matrix. The efficiencies are computed at different p -variate t -distributions with selected values of dimensions p and degrees of freedom ν , where $\nu = \infty$ refers to the multivariate normal case. The variance of the off-diagonal element of regular shape estimator at the t -distribution is $(\nu - 2)/(\nu - 4)$. In the table, the notation $k = 0$ refers to the Tyler's M-estimator, which is used as a starting value. The variance of the off-diagonal element of it is $(p + 2)/p$. When $k = \infty$, the efficiencies of regular Dümbgen's estimator and the rank covariance estimator are listed.

As seen in Table 1, when p is small, by taking just one step in this estimation procedure, a significant increase in the efficiencies is seen as compared to the initial Tyler's M-estimator ($k = 0$). After three steps, the efficiencies of k -step estimators are already very close to those of the Dümbgen's or rank covariance estimator ($k = \infty$). When p increases, there are not much differences seen in the efficiencies of different k -step estimators. In fact, for high-dimensional data, 1-step sign estimator seems to provide a robust and highly efficient alternative to the regular shape estimator.

In case of multivariate normal data, the limiting efficiencies of sign and rank covariance estimators are very close to each other. When high-dimensional heavy-tailed data is encountered, the sign estimators seem to outperform the rank estimators. In other cases, the rank estimators are slightly more efficient than the sign estimators.

5. Simulation studies

In this section, we will use simple simulation studies to compare the finite-sample efficiencies of different k -step shape estimators with respect to those of the regular shape estimator.

$M = 2500$ samples with sample sizes $n = 50, 100$ and 200 and dimensions $p = 3$ and 5 were drawn from t -distributions with $\nu = 5, 8$ and ∞ degrees of freedom, where $\nu = \infty$ corresponds to multinormal samples. For every estimator and distribution, the mean squared errors (MSE) were computed

Table 1: AREs of k -step sign and rank (in brackets) estimators as compared to the regular shape estimator at different p -variate t -distributions with selected values of dimension p and degrees of freedom ν . $k = 0$ refers to the Tyler's M-estimator that is used as a starting value. $k = \infty$ refers to the Dümbgen's estimator and the regular rank estimator.

	p	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = \infty$	
$\nu = 5$	2	1.50	2.03	2.24	2.30	2.33	
		1.50	(2.04)	(2.25)	(2.32)	(2.34)	
	3	1.80	2.28	2.38	2.39	2.40	
		1.80	(2.29)	(2.39)	(2.40)	(2.40)	
	4	2.00	2.41	2.45	2.45	2.45	
		2.00	(2.41)	(2.45)	(2.45)	(2.45)	
	5	2.14	2.48	2.49	2.49	2.49	
		2.14	(2.47)	(2.47)	(2.47)	(2.47)	
	$\nu = 8$	2	0.75	1.04	1.16	1.20	1.23
			0.75	(1.04)	(1.17)	(1.21)	(1.24)
3		0.90	1.17	1.24	1.25	1.26	
		0.90	(1.18)	(1.25)	(1.27)	(1.28)	
4		1.00	1.23	1.27	1.28	1.28	
		1.00	(1.24)	(1.28)	(1.29)	(1.29)	
5		1.07	1.27	1.30	1.30	1.30	
		1.07	(1.27)	(1.30)	(1.30)	(1.30)	
$\nu = \infty$		2	0.50	0.72	0.83	0.88	0.91
			0.50	(0.72)	(0.83)	(0.88)	(0.91)
	3	0.60	0.82	0.89	0.91	0.92	
		0.60	(0.82)	(0.89)	(0.91)	(0.92)	
	4	0.67	0.86	0.92	0.93	0.93	
		0.67	(0.86)	(0.92)	(0.93)	(0.93)	
	5	0.71	0.89	0.93	0.94	0.94	
		0.71	(0.89)	(0.93)	(0.94)	(0.94)	

using

$$\text{MSE}(\widehat{V}_{k,ij}) = \frac{1}{M} \sum_{k=1}^{2500} \widehat{V}_{k,ij}^{(m)2},$$

where $\widehat{V}_{k,ij}$, with $i \neq j$, denotes the off-diagonal element of corresponding k -step estimate computed from the m -th generated sample. The estimated finite sample efficiencies were then computed as ratios of the simulated MSEs and are listed in Tables 2 and 3 for $p = 3$ and $p = 5$ respectively. The corresponding asymptotic relative efficiencies (denoted by $n = \infty$) from Table 1 are also listed for easy reference.

The results in Tables 2 and 3 show that when samples are generated from the multinormal distributions, the finite sample and the limiting efficiencies are almost the same even when $n = 50$. When sampling is done from the heavy-tailed distributions, the convergence to the limiting efficiency is clear but much slower. Especially for small sample sizes, the loss in efficiency is remarkable, but also in the case $n = 200$, the efficiencies are far from the asymptotical ones.

In Figure 2 examples of the average computation times (in seconds) of different k -step sign (left column) and rank (right column) estimators are illustrated as a function of n for $p = 2, 3$ and 5. The average was taken over 100 trials. Tyler's M-estimator was again used as an initial estimator. The estimators were computed on an ordinary workstation using C-functions called from R, similar to the implementations found in **R**-package **SpatialNP** (Sirkiä et al., 2008).

Since these estimators operate on pairs of observations instead of observations themselves it is clear that in relative terms they are much more time consuming than the ordinary covariance matrix, for example. However, as shown in Figure 2, in absolute terms the increase is not unbearable, at least in a simple application. As the number of dimensions increase the time consumption of the sign estimator increases as a function of n faster than that of the rank estimator. This is because the computation of an outer product becomes heavier in higher dimension, and the number of those to be computed for the sign estimator is of order n^2 and only n for the rank estimator.

The use of pairs also implies that the increase in time consumption as a function of sample size is essentially quadratic. This together with the iterative nature of the original estimators can indeed lead to rather long computing times but it should be kept in mind that in practice the iteration is just a repetition of the step until some convergence criteria is met. This

Table 2: Finite sample efficiencies of the k -step sign (left column) and rank (right column) estimators as compared to the regular shape estimator at different 3-variate t -distributions with selected values of ν .

		sign				rank			
	n	$k = 1$	$k = 2$	$k = 3$	$k = \infty$	$k = 1$	$k = 2$	$k = 3$	$k = \infty$
$\nu = \infty$	50	0.81	0.88	0.91	0.92	0.81	0.89	0.91	0.93
	100	0.81	0.88	0.91	0.92	0.81	0.88	0.91	0.92
	200	0.82	0.90	0.92	0.93	0.82	0.89	0.92	0.93
	∞	0.82	0.89	0.91	0.92	0.82	0.89	0.91	0.92
$\nu = 8$	50	1.07	1.14	1.16	1.17	1.08	1.15	1.17	1.17
	100	1.10	1.17	1.19	1.20	1.11	1.18	1.20	1.20
	200	1.14	1.21	1.23	1.24	1.14	1.22	1.23	1.24
	∞	1.17	1.24	1.25	1.26	1.18	1.25	1.27	1.28
$\nu = 5$	50	1.41	1.47	1.49	1.49	1.42	1.48	1.50	1.50
	100	1.60	1.67	1.68	1.68	1.61	1.67	1.69	1.69
	200	1.70	1.77	1.79	1.79	1.70	1.78	1.79	1.79
	∞	2.28	2.38	2.40	2.40	2.29	2.39	2.40	2.40

Table 3: Finite sample efficiencies of the k -step sign (left column) and rank (right column) estimators as compared to the regular shape estimator at different 5-variate t -distributions with selected values of ν .

		sign				rank			
	n	$k = 1$	$k = 2$	$k = 3$	$k = \infty$	$k = 1$	$k = 2$	$k = 3$	$k = \infty$
$\nu = \infty$	50	0.88	0.92	0.93	0.94	0.88	0.92	0.93	0.94
	100	0.89	0.93	0.94	0.94	0.89	0.93	0.94	0.94
	200	0.90	0.94	0.94	0.95	0.89	0.93	0.94	0.94
	∞	0.89	0.93	0.94	0.94	0.89	0.93	0.94	0.94
$\nu = 8$	50	1.17	1.20	1.20	1.20	1.17	1.20	1.20	1.20
	100	1.23	1.25	1.25	1.25	1.23	1.25	1.25	1.26
	200	1.24	1.26	1.26	1.26	1.24	1.26	1.26	1.26
	∞	1.27	1.30	1.30	1.30	1.27	1.30	1.30	1.30
$\nu = 5$	50	1.55	1.57	1.57	1.57	1.56	1.58	1.57	1.57
	100	1.73	1.75	1.75	1.75	1.73	1.75	1.75	1.74
	200	1.98	2.00	2.00	2.00	1.98	1.99	1.99	1.99
	∞	2.48	2.49	2.49	2.49	2.47	2.47	2.47	2.47

criteria is basically a small number which is usually chosen somewhat arbitrarily but it has a severe impact on how many times the iteration step is repeated. On the other hand it has been previously shown that even with a very modest number of steps the properties of the k -step estimators are already close to those of the (theoretical) limit. Using the k -step versions removes the need for choosing the convergence criteria, and as can be seen from the figure, can greatly reduce the total time consumption.

6. Conclusions

In this paper, we proposed so called k -step versions of the shape estimators introduced by Dümbgen (1998) and Oja and Randles (2004). The motivation for this study arose mainly from the fact that at the moment there is no proof for the existence and uniqueness of the estimator based on spatial rank vectors (Oja and Randles, 2004). From the practical point of view, the long computation times of the estimators may also cause problems in some applications.

The estimators proposed in this paper appear to be very practical with known asymptotical properties. We showed that the k -step estimators inherit some important properties of the initial estimators, namely, affine equivariance, robustness and limiting normality. The k -step sign estimator was also shown to have the important independence property if the initial estimator has it. This property is needed for example if the independent component analysis problem is solved using the technique utilizing two scatter (or shape) matrices (Tyler et al., 2009). Note that in our robustness and simulation studies we used as a starting values the Tyler’s M-estimator (Tyler, 1987), which does not possess the independence property. A simple example of the robust initial estimator with independence property is obtained as the Tyler’s M-estimator is computed using $n/2$ pairwise differences $\mathbf{x}_{2i} - \mathbf{x}_{2i-1}$, $i = 1, \dots, n/2$.

In our simulation studies of Section 5 we showed that with taking just few steps in the estimation procedure, robust and highly efficient alternatives to traditional spatial sign and rank covariance estimators are obtained. The computation times of these estimators are very low as compared to those of Dümbgen’s estimator and spatial rank covariance matrix. Programs are available in the **R**-package **SpatialNP**.

References

- Chritchley, F., Pires, A. and Amado, C., 2008. Principal axis analysis. Manuscript.
- Dümbgen, L., 1998. On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.*, 50, 471–491.
- Dümbgen, L. and Tyler, D., 2005. On the Breakdown Properties of Some Multivariate M-Functionals. *Scand. J. Statist.*, 32, 247–264
- Hallin, M., Oja, H. and Paindaveine, D., 2006. Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape, *Ann. Statist.*, 34, 2757–2789.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.J., 1986. *Robust Statistics: the Approach Based on Influence Functions*. Wiley, New York.
- Hettmansperger, T.P. and Randles, R.H., 2003. A practical affine equivariant multivariate median. *Biometrika*, 89, 851–860.
- Möttönen J., Oja, H. and Tienari J., 1997. On the efficiency of multivariate spatial sign and rank tests. *Ann. Statist.*, 25, 542–552.
- Oja, H. and Randles, R., 2004. Multivariate nonparametric tests, *Statist. Sci.*, 19, 598–605.
- Oja, H., Sirkiä, S. and Eriksson, J., 2006. Scatter matrices and independent component analysis, *Austrian J. Statist.*, 19, 175–189.
- Paindaveine, D., 2008. A canonical definition of shape. *Statist. Probab. Lett.*, 78, 2240–2247.
- Sirkiä, S., Taskinen, S and Oja, H. (2007). Symmetrised M-estimators of multivariate scatter. *Journal of Multivariate Analysis* 98, 1611–1629.
- Sirkiä, S., Taskinen, S., and Oja, H., 2007. Symmetrised M-estimators of scatter. *J. Multivariate Anal.*, 98, 1611–1629.
- Sirkiä, S., Taskinen, S., Nevalainen, J. and Oja, H., 2008. Multivariate non-parametrical methods based on spatial signs and ranks: The R package SpatialNP, conditionally accepted in *J. Statist. Software*.

Sirkiä, S., Taskinen, S., Oja H. and Tyler, D., 2009. Multivariate tests and estimates for shape based on spatial signs and ranks, *J. Nonp. Statist.*, 21, 155–176.

Tyler, D.E., 1987. A distribution-free M-estimator of multivariate scatter. *Ann. Statist.*, 15, 234–251.

Tyler, D.E., Critchley, F., Dümbgen, L. and Oja, H., 2009. Invariant coordinate selection, *J. Royal Statist. Soc. B*, 71, 549–592.

Visuri, S., Oja, H. and Koivunen, V., 2000. Sign and rank covariane matrices. *J. Statist. Plann. Inference*, 91, 557–575.

Appendix: Proofs of the results

PROOF OF THEOREM 1. As in Sirkiä et al. (2007), write I_i^- for the $p \times p$ diagonal matrix that has -1 as the i th element and 1 as all other diagonal elements. Now as the components of \mathbf{x} are independent, then the components of $\mathbf{x}_1 - \mathbf{x}_2$ are independent as well and symmetrically distributed around zero. Moreover, $\mathbf{x}_1 - \mathbf{x}_2$ and $I_i^-(\mathbf{x}_1 - \mathbf{x}_2)$ have the same distribution and the corresponding shape functionals are equivalent.

For simplicity, write $V_1(\mathbf{x}_1 - \mathbf{x}_2)$ and $V_1(I_i^-(\mathbf{x}_1 - \mathbf{x}_2))$ for the 1-step sign functional obtained using $\mathbf{x}_1 - \mathbf{x}_2$ and $I_i^-(\mathbf{x}_1 - \mathbf{x}_2)$, respectively. Then if V_0 is diagonal, we get

$$\begin{aligned} V_1(\mathbf{x}_1 - \mathbf{x}_2) &= V_1(I_i^-(\mathbf{x}_1 - \mathbf{x}_2)) = p E \left[\frac{I_i^-(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)^T I_i^-}{(\mathbf{x}_1 - \mathbf{x}_2)^T I_i^- V_0^{-1} I_i^- (\mathbf{x}_1 - \mathbf{x}_2)} \right] \\ &= p E \left[\frac{I_i^-(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)^T I_i^-}{(\mathbf{x}_1 - \mathbf{x}_2)^T V_0^{-1} (\mathbf{x}_1 - \mathbf{x}_2)} \right] = I_i^- V_1(\mathbf{x}_1 - \mathbf{x}_2) I_i^-, \end{aligned}$$

for all $i = 1, \dots, p$. This implies that $[V_1(\mathbf{x}_1 - \mathbf{x}_2)]_{ij} = -[V_1(\mathbf{x}_1 - \mathbf{x}_2)]_{ij}$, for all $i \neq j$, that is, $V_1(\mathbf{x}_1 - \mathbf{x}_2)$ and further the subsequent k -step functionals are diagonal.

PROOF OF THEOREM 2. Applying functional (9) to $F_\epsilon = (1 - \epsilon)F_0 + \epsilon\Delta_{\mathbf{x}}$

and taking the derivative with respect to ϵ at $\epsilon = 0$ yields to

$$\begin{aligned} IF(\mathbf{x}; V_k, F_0) &= \frac{\partial}{\partial \epsilon} p E \left[\frac{(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)^T}{(\mathbf{x}_1 - \mathbf{x}_2)^T V_{k-1}^{-1}(F_\epsilon)(\mathbf{x}_1 - \mathbf{x}_2)} \right] \Big|_{\epsilon=0} \\ &\quad - 2p E \left[\frac{(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)^T}{(\mathbf{x}_1 - \mathbf{x}_2)^T V_{k-1}^{-1}(F_\epsilon)(\mathbf{x}_1 - \mathbf{x}_2)} \right] \Big|_{\epsilon=0} \\ &\quad + 2p E_{\mathbf{x}_1} \left[\frac{(\mathbf{x}_1 - \mathbf{x})(\mathbf{x}_1 - \mathbf{x})^T}{(\mathbf{x}_1 - \mathbf{x})^T V_{k-1}^{-1}(F_\epsilon)(\mathbf{x}_1 - \mathbf{x})} \right] \Big|_{\epsilon=0}. \end{aligned}$$

Then by writing $\mathbf{u}_{12} = \|\mathbf{x}_1 - \mathbf{x}_2\|^{-1}(\mathbf{x}_1 - \mathbf{x}_2)$, and $\mathbf{u}_{1\mathbf{x}} = \|\mathbf{x}_1 - \mathbf{x}\|^{-1}(\mathbf{x}_1 - \mathbf{x})$ and assuming that the order of differentiation and integration can be changed, we get that

$$IF(\mathbf{x}; V_k, F_0) = p E[\mathbf{u}_{12} \mathbf{u}_{12}^T \mathbf{u}_{12}^T IF(\mathbf{x}; V_{k-1}, F_0) \mathbf{u}_{12}] - 2p E[\mathbf{u}_{12} \mathbf{u}_{12}^T] + 2p E_{\mathbf{x}_1}[\mathbf{u}_{1\mathbf{x}} \mathbf{u}_{1\mathbf{x}}^T].$$

Now using

$$E[\mathbf{u}_{12} \mathbf{u}_{12}^T \mathbf{u}_{12}^T IF(\mathbf{x}; V_{k-1}, F_0) \mathbf{u}_{12}] = 2 [p(p+2)]^{-1} IF(\mathbf{x}; V_{k-1}, F_0)$$

and

$$E_{\mathbf{x}_1}[\mathbf{u}_{1\mathbf{x}} \mathbf{u}_{1\mathbf{x}}^T] = (1 - pg(r)) \mathbf{u} \mathbf{u}^T + g(r) I_p,$$

where

$$g(r) = E_{\mathbf{x}_1} \left[\frac{(\mathbf{x}_1 - r \mathbf{e}_1)_2^2}{\|\mathbf{x}_1 - r \mathbf{e}_1\|^2} \right]$$

(Sirkiä et al., 2009), the influence function reduces to

$$\begin{aligned} IF(\mathbf{x}; V_k, F_0) &= \frac{2}{p+2} IF(\mathbf{x}; V_{k-1}, F_0) + 2p(1 - pg(r)) \left(\mathbf{u} \mathbf{u}^T - \frac{1}{p} I_p \right) \\ &= \left(\frac{2}{p+2} \alpha_{V_{k-1}}(r) + 2p(1 - pg(r)) \right) \left(\mathbf{u} \mathbf{u}^T - \frac{1}{p} I_p \right) \\ &= \alpha_{V_k}(r) \left(\mathbf{u} \mathbf{u}^T - \frac{1}{p} I_p \right). \end{aligned}$$

Since the influence functions of k -step estimators for all k are of the same type, we get

$$\begin{aligned} \alpha_{V_k}(r) &= \left(\frac{2}{p+2} \right)^k \alpha_{V_0} + \sum_{q=0}^{k-1} \left(\frac{2}{p+2} \right)^q 2p(1 - pg(r)) \\ &= c_p^k \alpha_{V_0}(r) + (1 - c_p^k) 2(p+2)(1 - pg(r)), \end{aligned}$$

where $c_p = 2(p+2)^{-1}$.

PROOF OF THEOREM 3. The functional (10) may alternatively be written as

$$E_F [V_k^{-1/2} V_{k-1}^{1/2} \mathbf{S}(\mathbf{z}_1 - \mathbf{z}_2) \mathbf{S}^T(\mathbf{z}_1 - \mathbf{z}_3) V_{k-1}^{1/2} V_k^{-1/2} - p^{-1} \mathbf{S}(\mathbf{z}_1 - \mathbf{z}_2) \mathbf{S}^T(\mathbf{z}_1 - \mathbf{z}_3) I_p] = 0. \quad (12)$$

Then proceeding as in the proof of Theorem 2, that is, applying functional (12) to $F_\epsilon = (1 - \epsilon)F_0 + \epsilon\Delta_{\mathbf{x}}$ and taking the derivative with respect to ϵ at $\epsilon = 0$, we get

$$\begin{aligned} & E[IF(\mathbf{x}; V_k, F_0) \mathbf{u}_{12} \mathbf{u}_{13}^T] - E[\mathbf{u}_{12} \mathbf{u}_{13}^T \mathbf{u}_{13}^T IF(\mathbf{x}; V_{k-1}, F_0) \mathbf{u}_{13}] \\ & - p^{-1} E[\mathbf{u}_{12}^T IF(\mathbf{x}; V_{k-1}, F_0) \mathbf{u}_{13}] I_p + p^{-1} E[\mathbf{u}_{12}^T \mathbf{u}_{13} \mathbf{u}_{13}^T IF(\mathbf{x}; V_{k-1}, F_0) \mathbf{u}_{13}] I_p \\ & - E_{\mathbf{x}_2, \mathbf{x}_3} [\mathbf{u}_{\mathbf{x}_2} \mathbf{u}_{\mathbf{x}_3}^T - p^{-1} \mathbf{u}_{\mathbf{x}_2}^T \mathbf{u}_{\mathbf{x}_3} I_p] - 2E_{\mathbf{x}_1, \mathbf{x}_3} [\mathbf{u}_{1\mathbf{x}} \mathbf{u}_{13}^T - p^{-1} \mathbf{u}_{1\mathbf{x}}^T \mathbf{u}_{13} I_p] = 0, \end{aligned} \quad (13)$$

where \mathbf{u}_{12} , \mathbf{u}_{13} , $\mathbf{u}_{1\mathbf{x}}$, $\mathbf{u}_{\mathbf{x}_2}$ and $\mathbf{u}_{\mathbf{x}_3}$ are defined as in the previous proof. Next we will use the results given in the proof of Lemma 4 in Sirkiä et al. (2009). The terms on the second line of (13) include $Tr(IF(\mathbf{x}; V_{k-1}, F_0))$ and therefore equal to zero. Further,

$$E[IF(\mathbf{x}; V_k, F_0) \mathbf{u}_{12} \mathbf{u}_{13}^T] = p^{-1} c_F^2 IF(\mathbf{x}; V_k, F_0),$$

$$E[\mathbf{u}_{12} \mathbf{u}_{13}^T \mathbf{u}_{13}^T IF(\mathbf{x}; V_{k-1}, F_0) \mathbf{u}_{13}] = 2c_F^2 [p(p+2)]^{-1} IF(\mathbf{x}; V_{k-1}, F_0)$$

and

$$E_{\mathbf{x}_2, \mathbf{x}_3} [\mathbf{u}_{\mathbf{x}_2} \mathbf{u}_{\mathbf{x}_3}^T] + 2E_{\mathbf{x}_1, \mathbf{x}_3} [\mathbf{u}_{1\mathbf{x}} \mathbf{u}_{13}^T] = (g_1(r) - g_2(r)) (\mathbf{u} \mathbf{u}^T - p^{-1} I_p),$$

where $c_F^2 = E_F[q_F^2(r)]$ and $q_F(r)$ is defined in (4). Functions $g_1(r)$ and $g_2(r)$ are given in the proof of Lemma 3 in Sirkiä et al. (2009). Thus

$$IF(\mathbf{x}; V_k, F_0) = \frac{2}{p+2} IF(\mathbf{x}; V_{k-1}, F_0) + \frac{p}{c_F^2} (g_1(r) - g_2(r)) \left(\mathbf{u} \mathbf{u}^T - \frac{1}{p} I_p \right),$$

and further

$$\begin{aligned} IF(\mathbf{x}; V_k, F_0) &= \left(\frac{2}{p+2} \alpha_{V_{k-1}}(r) + \frac{p}{c_F^2} (g_1(r) - g_2(r)) \right) \left(\mathbf{u} \mathbf{u}^T - \frac{1}{p} I_p \right) \\ &= \alpha_{V_k}(r) \left(\mathbf{u} \mathbf{u}^T - \frac{1}{p} I_p \right). \end{aligned}$$

The result then follows by proceeding as in the proof of Theorem 2.

PROOF OF COROLLARY 1. As in Sirkiä et al. (2007), note that

$$g(r) = E_{\mathbf{x}_1} \left[\frac{(\mathbf{x}_1 - r\mathbf{e}_1)_2^2}{\|\mathbf{x}_1 - r\mathbf{e}_1\|^2} \right]$$

is clearly finite since $0 \leq (\mathbf{x}_1 - r\mathbf{e}_1)_2^2 \leq \|\mathbf{x}_1 - r\mathbf{e}_1\|^2$. By similar argument, it is seen that $g_1(r) - g_2(r)$ in the influence function of k -step rank estimator is finite. Corresponding influence functions are therefore bounded when α_{V_0} is bounded.

PROOF OF THEOREM 4. Consider first the limiting distribution of 1-step sign estimator. Write

$$\widehat{V}_1 = p \left[\text{ave} \left\{ \frac{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)}{(\mathbf{x}_i - \mathbf{x}_j)^T \widehat{V}_0^{-1} (\mathbf{x}_i - \mathbf{x}_j)} \right\} \right]^{-1} \text{ave} \left\{ \frac{(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T}{(\mathbf{x}_i - \mathbf{x}_j)^T \widehat{V}_0^{-1} (\mathbf{x}_i - \mathbf{x}_j)} \right\},$$

that is, the estimator is scaled so that $Tr(\widehat{V}_1) = p$. Let now $\mathbf{x}_i, \dots, \mathbf{x}_n$ be a sample from a spherically symmetric distribution and write $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, $r_{ij} = \|\mathbf{x}_{ij}\|$ and $\mathbf{u}_{ij} = r_{ij}^{-1} \mathbf{x}_{ij}$, for simplicity. Further, write $V_0^* = \sqrt{n}(\widehat{V}_0 - I_k)$ and note that since \widehat{V}_0 is \sqrt{n} -consistent, V_0^* is bounded in probability. Now,

$$\widehat{V}_0^{-1/2} = I_k - \frac{1}{2\sqrt{n}} V_0^* + o_p(n^{-1/2}),$$

$$\widehat{V}_0^{-1/2} \mathbf{x}_{ij} = \mathbf{x}_{ij} - \frac{1}{2\sqrt{n}} V_0^* \mathbf{x}_{ij} + o_p(n^{-1/2}),$$

and

$$[\mathbf{x}_{ij}^T \widehat{V}_0^{-1} \mathbf{x}_{ij}]^{-1} = \frac{1}{\|\mathbf{x}_{ij}\|^2} \left(1 + \frac{1}{\sqrt{n}} \mathbf{u}_{ij}^T V_0^* \mathbf{u}_{ij} + o_p(n^{-1/2}) \right).$$

Then (omitting the $o_p(n^{-1/2})$ terms)

$$\widehat{V}_1 = p \left[1 + \frac{1}{\sqrt{n}} \text{ave} \{ \mathbf{u}_{ij}^T V_0^* \mathbf{u}_{ij} \} \right]^{-1} \left[\text{ave} \{ \mathbf{u}_{ij} \mathbf{u}_{ij}^T \} + \frac{1}{\sqrt{n}} \text{ave} \{ \mathbf{u}_{ij}^T V_0^* \mathbf{u}_{ij} \mathbf{u}_{ij} \mathbf{u}_{ij}^T \} \right]$$

and further

$$\begin{aligned} \sqrt{n}(\widehat{V}_1 - I_p) &= \left[1 + \frac{1}{\sqrt{n}} \text{ave} \{ \mathbf{u}_{ij}^T V_0^* \mathbf{u}_{ij} \} \right]^{-1} \left[p\sqrt{n} \left(\text{ave} \{ \mathbf{u}_{ij} \mathbf{u}_{ij}^T \} - \frac{1}{p} I_p \right) \right. \\ &\quad \left. + p \text{ave} \{ \mathbf{u}_{ij}^T V_0^* \mathbf{u}_{ij} \mathbf{u}_{ij} \mathbf{u}_{ij}^T \} - \text{ave} \{ \mathbf{u}_{ij}^T V_0^* \mathbf{u}_{ij} \} I_p \right]. \end{aligned} \tag{14}$$

The asymptotic normality of $\sqrt{n}(\text{ave}\{\mathbf{u}_{ij}\mathbf{u}_{ij}^T\} - p^{-1}I_p)$ follows from the U-statistic theory and is proved in Sirkiä et al. (2009). The asymptotic normality of $\sqrt{n}(\widehat{V}_1 - I_p)$ then follows from the Slutsky's theorem and the fact that

$$E[\mathbf{u}_{ij}^T V^* \mathbf{u}_{ij}] = \text{Tr}(V^*) = 0$$

and

$$E[\mathbf{u}_{ij}^T V_0^* \mathbf{u}_{ij} \mathbf{u}_{ij} \mathbf{u}_{ij}^T] = \frac{2}{p(p+2)} V_0^* = \frac{2}{p(p+2)} \sqrt{n}(\widehat{V}_0 - I_p),$$

where, by assumption, $\sqrt{n}(\widehat{V}_0 - I_p)$ is multivariate normal.

Equation (14) reduces now to

$$\sqrt{n}(\widehat{V}_1 - I_p) = p\sqrt{n} \left(\text{ave}\{\mathbf{u}_{ij}\mathbf{u}_{ij}^T\} - \frac{1}{p}I_p \right) + \frac{2}{p+2} \sqrt{n}(\widehat{V}_0 - I_p),$$

and by applying similar technique to subsequent k -step estimators, we obtain

$$\begin{aligned} & \sqrt{n}(\widehat{V}_k - I_p) \\ &= \sum_{q=0}^{k-1} \left(\frac{2}{p+2} \right)^q p\sqrt{n} \left(\text{ave}\{\mathbf{u}_{ij}\mathbf{u}_{ij}^T\} - \frac{1}{p}I_p \right) + \left(\frac{2}{p+2} \right)^k \sqrt{n}(\widehat{V}_0 - I_p) \\ &= (1 - c_p^k)(p+2)\sqrt{n} \left(\text{ave}\{\mathbf{u}_{ij}\mathbf{u}_{ij}^T\} - \frac{1}{p}I_p \right) + c_p^k \sqrt{n}(\widehat{V}_0 - I_p). \end{aligned} \quad (15)$$

The asymptotic covariance matrix may be computed by writing (15) using the influence functions as follows

$$\begin{aligned} & \sqrt{n} \text{vec}(\widehat{V}_k - I_p) \\ &= \sqrt{n} \text{vec} \left(\text{ave} \left\{ [(1 - c_p^k)2(p+2)(1 - pg(r_i)) + c_p^k \alpha_{V_0}(r_i)] \left[\mathbf{u}_i \mathbf{u}_i^T - \frac{1}{p}I_p \right] \right\} \right) \\ &= \sqrt{n} \text{vec} \left(\text{ave} \left\{ \alpha_{V_k}(r_i) \left[\mathbf{u}_i \mathbf{u}_i^T - \frac{1}{p}I_p \right] \right\} \right). \end{aligned}$$

Then the covariance matrix of $\sqrt{n} \text{vec}(\widehat{V}_k - I_p)$ equals

$$\begin{aligned} & E \left[\alpha_{V_k}^2(r) \text{vec} \left(\mathbf{u} \mathbf{u}^T - \frac{1}{p}I_p \right) \text{vec}^T \left(\mathbf{u} \mathbf{u}^T - \frac{1}{p}I_p \right) \right] \\ &= \frac{E[\alpha_{V_k}^2(r)]}{p(p+2)} (I_{p^2} + K_{p,p} - \frac{2}{p}J_p) = \frac{E[\alpha_{V_k}^2(r)]}{p(p+2)} C_{p,p}(I_p). \end{aligned}$$

The limiting distribution of 1-step rank estimator can be derived as above. In that case

$$\widehat{V}_1 = p \left[\text{ave}\{\mathbf{u}_{ij}^T \mathbf{u}_{ik}\} + \frac{1}{2\sqrt{n}} \text{ave}\{\mathbf{u}_{ik}^T V^* \mathbf{u}_{ik}^T \mathbf{u}_{ij}^T \mathbf{u}_{ik}\} + \frac{1}{2\sqrt{n}} \text{ave}\{\mathbf{u}_{ij}^T V^* \mathbf{u}_{ij}^T \mathbf{u}_{ij}^T \mathbf{u}_{ik}\} \right]^{-1} \cdot \left[\text{ave}\{\mathbf{u}_{ij} \mathbf{u}_{ik}^T\} + \frac{1}{2\sqrt{n}} \text{ave}\{\mathbf{u}_{ik}^T V^* \mathbf{u}_{ik}^T \mathbf{u}_{ij} \mathbf{u}_{ik}^T\} + \frac{1}{2\sqrt{n}} \text{ave}\{\mathbf{u}_{ij}^T V^* \mathbf{u}_{ij}^T \mathbf{u}_{ij} \mathbf{u}_{ik}^T\} \right].$$

Note next that

$$E[\mathbf{u}_{ij}^T \mathbf{u}_{ik}] = c_F^2, \quad E[\mathbf{u}_{ij}^T V^* \mathbf{u}_{ij}^T \mathbf{u}_{ij}^T \mathbf{u}_{ik}] = \frac{c_F^2}{p} \text{tr}(V^*) = 0$$

and

$$E[\mathbf{u}_{ij}^T V^* \mathbf{u}_{ij}^T \mathbf{u}_{ij}^T \mathbf{u}_{ik}] = \frac{2c_F^2}{p(p+2)} V^*,$$

where $c_F^2 = E_F[q^2(r)]$ (Sirkiä et al., 2009). Then using again the Slutsky's theorem and the U-statistics theory, we get that

$$\sqrt{n}(\widehat{V}_1 - I_p) = \frac{p\sqrt{n}}{c_F^2} \left(\text{ave}\{\mathbf{u}_{ij} \mathbf{u}_{ik}^T\} - \text{ave}\{\mathbf{u}_{ij}^T \mathbf{u}_{ik}\} \frac{1}{p} I_p \right) + \frac{2}{p+2} \sqrt{n}(\widehat{V}_0 - I_p)$$

has the limiting normal distribution. Further,

$$\begin{aligned} & \sqrt{n}(\widehat{V}_k - I_p) \\ &= \sum_{q=0}^{k-1} \left(\frac{2}{p+2} \right)^q \frac{p\sqrt{n}}{c_F^2} \left(\text{ave}\{\mathbf{u}_{ij} \mathbf{u}_{ik}^T\} - \text{ave}\{\mathbf{u}_{ij}^T \mathbf{u}_{ik}\} \frac{1}{p} I_p \right) + \left(\frac{2}{p+2} \right)^k \sqrt{n}(\widehat{V}_0 - I_p) \\ &= \frac{(1 - c_p^k)(p+2)}{c_F^2} \sqrt{n} \left(\text{ave}\{\mathbf{u}_{ij} \mathbf{u}_{ik}^T\} - \text{ave}\{\mathbf{u}_{ij}^T \mathbf{u}_{ik}\} \frac{1}{p} I_p \right) + c_p^k \sqrt{n}(\widehat{V}_0 - I_p). \end{aligned}$$

The asymptotic covariance matrix is obtained by writing

$$\begin{aligned} & \sqrt{n} \text{vec}(\widehat{V}_k - I_p) \\ &= \sqrt{n} \text{vec} \left(\text{ave} \left\{ \left[\frac{(1 - c_p^k)(p+2)}{c_F^2} (g_1(r_i) - g_2(r_i)) + c_p^k \alpha_{V_0}(r_i) \right] \left[\mathbf{u}_i \mathbf{u}_i^T - \frac{1}{p} I_p \right] \right\} \right) \\ &= \sqrt{n} \text{vec} \left(\text{ave} \left\{ \alpha_{V_k}(r_i) \left[\mathbf{u}_i \mathbf{u}_i^T - \frac{1}{p} I_p \right] \right\} \right), \end{aligned}$$

and proceeding as in the case of sign estimator.

PROOF OF COROLLARY 2. Let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ denote a random sample from a spherical distribution and write $\widehat{V}_k(X)$ for the estimator computed on the sample. Now due to affine equivariance of \widehat{V}_k , in the elliptical case

$$\widehat{V}_k(V^{1/2}X) = p [Tr(V^{1/2}\widehat{V}_k(X)V^{1/2})]^{-1}V^{1/2}\widehat{V}_k(X)V^{1/2}.$$

Then using some simple matrix algebra, we get

$$\begin{aligned} & \sqrt{n} \text{vec}(\widehat{V}_k(V^{1/2}X) - V) \\ &= \sqrt{n} \text{vec} \left(p [Tr(V^{1/2}\widehat{V}_k(X)V^{1/2})]^{-1}V^{1/2}\widehat{V}_k(X)V^{1/2} - V \right) \\ &= \sqrt{n} p [Tr(V^{1/2}\widehat{V}_k(X)V^{1/2})]^{-1} \left[I_{p^2} - \frac{1}{p} \text{vec}(V)\text{vec}^T(I_p) \right] \text{vec}(V^{1/2}\widehat{V}_k(X)V^{1/2} - V) \\ &= \sqrt{n} p [Tr(V^{1/2}\widehat{V}_k(X)V^{1/2})]^{-1} W(V^{1/2} \otimes V^{1/2}) \text{vec}(\widehat{V}_k(X) - I_p), \end{aligned}$$

where $W = [I_{p^2} - p^{-1}\text{vec}(V)\text{vec}^T(I_p)]$. By Slutsky's theorem, $\sqrt{n} \text{vec}(\widehat{V}_k(V^{1/2}X) - V)$ has now a limiting normal distribution with asymptotic covariance matrix

$$\begin{aligned} & ASC \left\{ \sqrt{n} W(V^{1/2} \otimes V^{1/2}) \text{vec}(\widehat{V}_k(X) - I_p) \right\} \\ &= W(V^{1/2} \otimes V^{1/2}) ASC \{ \text{vec}(\widehat{V}_k(X) - I_p) \} (V^{1/2} \otimes V^{1/2})^T W^T \\ &= ASV(\widehat{V}_{k,12}; F_0) W(V^{1/2} \otimes V^{1/2}) C_{p,p}(I_p) (V^{1/2} \otimes V^{1/2})^T W^T \\ &= ASV(\widehat{V}_{k,12}; F_0) W C_{p,p}(V) W^T. \end{aligned}$$

Finally note that $W 2p^{-1}\text{vec}(V)\text{vec}^T(V) W^T = 0$, the result thus follows.

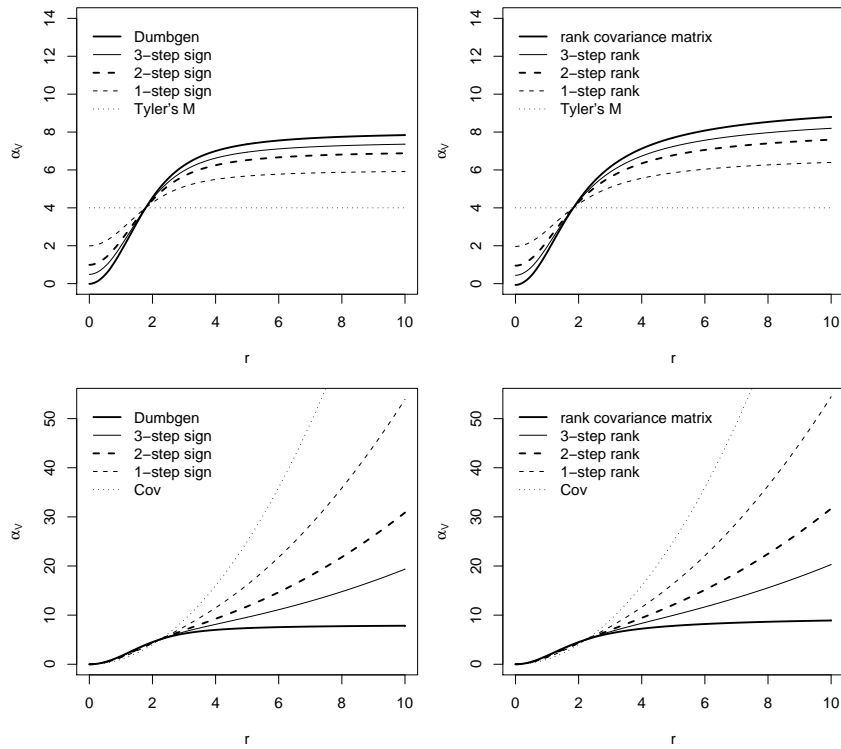


Figure 1: Functions α_{V_k} for some k -step sign and rank estimators when the Tyler's M-estimator (first row) and the sample covariance matrix (second row) is used as a starting value. The functions are computed at the bivariate standard normal distribution case.

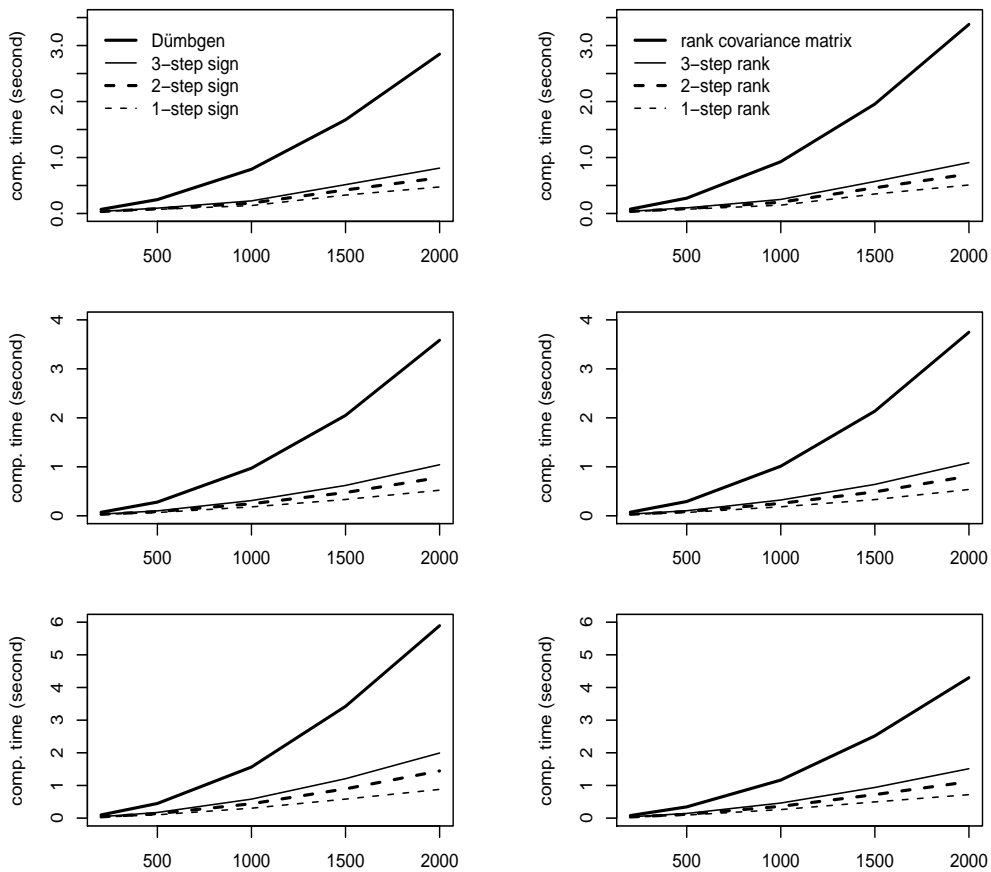


Figure 2: Average computation times of the k -step sign (left column) and rank (right column) estimators as a function of n at 2-dimensional (first row), 3-dimensional (second row) and 5-dimensional (third row) distribution cases. Tyler's M-estimate is used as a starting value.