The spatial sign covariance matrix with unknown location

Alexander Dürre [a], Daniel Vogel [b,*], David E. Tyler [c]

[a] *Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany*
[b] *Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany*
[c] *Department of Statistics, Rutgers University, Piscataway, NJ 08854, USA*

ABSTRACT

The consistency and asymptotic normality of the spatial sign covariance matrix with unknown location are shown. Simulations illustrate the different asymptotic behaviors when using the mean and the spatial median as a location estimator.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

We define the *spatial sign* of $x \in \mathbb{R}^p$ as $s(x) = x/|x|$ for $x \neq 0$ and $s(0) = 0$, where $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^p$. Let X be a p-dimensional random vector, $p \geq 2$, having distribution F. For $t \in \mathbb{R}^p$, we call

$$S(F, t) = E\left\{ s(X - t)s(X - t)^T \right\}$$

the *spatial sign covariance matrix (SSCM) of the distribution F (or random variable X) with location t*. Letting further $\mathbb{X}_n = (X_1, \ldots, X_n)^T$, where $X_1, \ldots, X_n$ represents a random sample from the distribution F, we call

$$S_n(\mathbb{X}_n, t) = \frac{1}{n} \sum_{i=1}^{n} s(X_i - t)s(X_i - t)^T$$

the *spatial sign covariance matrix of the sample $\mathbb{X}_n$ with location t*. The term spatial sign covariance matrix was coined by Visuri, Koivunen and Oja [31], but the estimator has a longer history in the statistics literature. It has excellent robustness properties: its influence function is bounded, and the asymptotic breakdown point is 1/2 [2]. Together with its simplicity, this makes the SSCM a popular scatter estimator in multivariate data analysis.

Within the theory of multivariate scatter estimation, affine equivariance plays an important role. Let $\mathbb{Y}_n = \mathbb{X}_n A^T + 1_n b^T$ denote the $n \times p$ data matrix obtained from $\mathbb{X}_n$ by applying the affine linear transformation $x \mapsto Ax + b$ to each data point (where 1_n denotes the n-vector consisting of ones). An *affine equivariant scatter estimator*, V_n say, satisfies

$$\hat{V}_n(\mathbb{Y}_n) = A\hat{V}_n(\mathbb{X}_n)A^T \tag{1}$$

---

* Corresponding author.
  E-mail addresses: Daniel.Vogel@udo.edu, vogeldts@rub.de (D. Vogel).

for any $b \in \mathbb{R}^p$ and any full rank square matrix A, i.e. it behaves like the covariance matrix under linear transformations of the data. The problem of robust, affine equivariant scatter estimation has received much attention in the last decades; see e.g. [19] or [34] for an overview. However, the SSCM lacks this property. It fulfills the weaker condition of *orthogonal invariance*, that is, it satisfies (1) for all orthogonal matrices A. This is closely related to the fact that, at elliptical distributions, the SSCM shares the eigenvectors (and the ranking of the eigenvalues) with the covariance matrix, but the exact connection between the eigenvalues of the SSCM and covariance matrix is, even under ellipticity, rather tricky. An explicit expression is known only for $p = 2$ (e.g. [5]). Thus, the SSCM gives information about the *orientation* of the data (cf. [1]), and its use has primarily been proposed for analyses that are based on this information only, most notably principal component analysis [18,14,3,8]. Other such applications are direction-of-arrival estimation [32] or testing for sphericity in the elliptical model [27]. The latter makes use of the fact that under the null hypothesis that X is spherical, $s(X)$ is uniformly distributed on the p-dimensional unit sphere.

The scatter estimator proposed by Tyler [28] can be regarded as an affine equivariant version of the SSCM. It lacks, however, the high breakdown point and the simplicity of the SSCM, its computation requiring an iterative algorithm. The estimate obtained when stopping the algorithm after a finite number of steps has been considered and called the k-step SSCM by Croux et al. [2]. It keeps the high breakdown point of the SSCM, but has an asymptotic variance close to that of Tyler's estimator.

In this paper, we are concerned with the asymptotic properties of the SSCM. The strong law of large numbers and the central limit theorem immediately yield

(I) $S_n(\mathbb{X}_n, t) \xrightarrow{\text{a.s.}} S(F, t)$ and

(II) $\sqrt{n} \, \mathrm{vec}\{S_n(\mathbb{X}_n, t) - S(F, t)\} \xrightarrow{d} N_{p^2}(0, W)$,

for every distribution F and $t \in \mathbb{R}^p$, where $W = \mathrm{Var}(Z)$ with $Z = Z(X, t) = \mathrm{vec}\{s(X - t)s(X - t)^T\}$ and $X \sim F$.

However, the estimator $S_n(\mathbb{X}_n, t)$ itself is rarely applicable, since the central location t is usually unknown and needs to be estimated. Instead, $S_n(\mathbb{X}_n, t_n)$ is used, where $t_n = t_n(\mathbb{X}_n)$ is a suitable location estimator. Often, (I) and (II) are mentioned as a theoretical justification for the use of $S_n(\mathbb{X}_n, t_n)$, accompanied by a more or less explicit remark that $S_n(\mathbb{X}_n, t)$ and $S_n(\mathbb{X}_n, t_n)$ possess the same asymptotic behavior as long as t_n is consistent for t in some suitable sense. The purpose of this article is to close this gap and rigorously prove that, under weak conditions on F and t_n, assertions (I) and (II) still hold true if t is replaced with t_n. We do not study the asymptotic variance and efficiency of the SSCM in this article, but refer the reader to Magyar and Tyler [17] and Dürre et al. [5]. Roughly speaking, the SSCM achieves the same asymptotic efficiency as Tyler's estimator at spherical distributions, which is $1 + 2/p$ relative to the (suitably scaled) sample covariance matrix, but its asymptotic variance may get arbitrarily large for heteroscedastic data.

The canonical location estimator for the SSCM is the *spatial median*

$$\mu_n = \mu_n(\mathbb{X}_n) = \arg\min_{\mu \in \mathbb{R}^p} \sum_{i=1}^{n} |X_i - \mu|.$$

If the data points do not lie on a straight line and none of them coincides with μ_n, the spatial signs with respect to the spatial median are centered [12, p. 228], i.e. $\sum_{i=1}^{n} s(X_i - \mu_n) = 0$. Thus the SSCM $S_n(\mathbb{X}_n, \mu_n)$ is indeed the sample covariance matrix of the spatial signs of the centered observations, if the latter are centered by the spatial median. The theoretical counterpart, the *spatial median of the distribution F*, is

$$\mu(F) = \arg\min_{\mu \in \mathbb{R}^p} E\left(|X - \mu| - |X|\right).$$

The spatial median always exists, and, if F is not concentrated on a straight line, it is unique. For further details see, e.g., [12,20,13]. If the first moments of F are finite, then the spatial median allows the more descriptive characterization as the minimizing point of $E|X - \mu|$. The spatial median falls within the class of M-estimators, for which an elaborate asymptotic theory exists [11, Section 6]. See [16] for the asymptotic distribution and finite sample efficiencies. The computation of the spatial median is also a thoroughly studied problem (e.g. [33,9,30]). Alternative names are L_1 median, median center and space median. For a recent review see [22].

An alternative definition of the sample spatial sign covariance matrix with unknown location is

$$S_n^*(\mathbb{X}_n, t_n) = \frac{1}{n^*} \sum_{i=1}^{n} s(X_i - t_n)s(X_i - t_n)^T$$

with $n^* = \#\{1 \leq i \leq n \mid X_i \neq t_n\}$. This definition is preferable for practical purposes. For instance, the trace of $S_n^*(\mathbb{X}_n, t_n)$ is always 1 (except for the uninteresting case that all observations X_i coincide with t_n). The estimator $S_n(\mathbb{X}_n, t)$, on the other hand, also contains the information how many of the data points X_i coincide with t. This information is generally of little interest. More importantly, when $n^* < n$ is observed, it is in most cases not due to properties of the underlying distribution (about which one seeks to draw inference), but is an artifact of the measurement or the location estimator. When $X_i = t_n$, it is usually either caused by a coarse rounding of the data or an estimator t_n that attains one of the data points with positive probability, which is common for many robust estimators, among them the spatial median. All asymptotic results of Section 2 concerning the estimator $S_n(\mathbb{X}_n, t_n)$ are also true for $S_n^*(\mathbb{X}_n, t_n)$. The difference between both estimators is, except for very small n, negligible for continuous population distributions.

A referee raised the question of the general importance of the location estimation problem and pointed out that it can be elegantly circumvented by means of symmetrization. For scatter estimation in general (univariate or multivariate), the need for a prior location estimate is usually regarded as a kind of nuisance, making thorough derivations more involved than, e.g., for the problem of location estimation alone. (We are addressing explicitly this nuisance here.) One way to avoid this problem is to use symmetrized scatter estimators. Any scatter estimator gives rise to a symmetrized estimator, which is simply the estimate computed from the pairwise differences, instead from the (suitably centered) data. To name two well-known univariate examples, the symmetrized version of the mean absolute deviation is known as Gini's mean difference (simply the mean of all pairwise distances). The Q_n scale estimator proposed by Rousseeuw and Croux [25], which is the lower sample quartile of all pairwise distances, can be regarded as a symmetrized version of the median absolute deviation (MAD, [10]). Roughly speaking, symmetrized estimators tend to be more efficient at the normal model, but less efficient at very heavy tailed distributions and less robust than the original estimator. Symmetrization is often successfully applied to highly robust estimators to increase their efficiency while retaining a satisfactory degree of robustness.

Multivariate symmetrized scatter estimators have been considered, e.g., in [4,26]. They play an important role in robust principal component analysis, since they possess the so-called independence property: symmetrized scatter functionals are always diagonal matrices at multivariate distributions with independent margins. See e.g. [23] or [29] for further details. Particularly the symmetrized version of the SSCM (which we simply want to call SSSCM here) has also been considered in [31]. It is denoted by TCM$_2$ there, and the authors call it the *spatial Kendall's τ covariance matrix*. It is indeed much more efficient than the SSCM at the normal distribution. For instance, at a bivariate spherical normal distribution, where the SSCM and the SSSCM both are consistent for $I_2/2$, the SSCM has an asymptotic relative efficiency (with respect to $\hat{\Sigma}_n/\mathrm{tr}(\hat{\Sigma}_n)$) of 50%, whereas the SSSCM achieves 91%, which can be deduced from results by Sirkiä et al. [27]. However, despite their appealing efficiency properties, symmetrized estimators also have a few drawbacks, mainly a higher computational effort, and a less tractable asymptotic variance, making inferential procedures more laborious. The SSCM, with its simplicity as a major appeal, will certainly be of continuing relevance in statistics.

The rest of the paper is organized as follows: in Section 2, strong and weak consistency of the SSCM are proven under the assumption of strong and weak consistency, respectively, of the location estimator t_n and a moment condition on $|X - t|^{-1}$. Furthermore, the asymptotic normality of $S_n(\mathbb{X}_n, t_n)$ is shown, provided that t_n converges at the usual $\sqrt{n}$ rate. Section 3 contains the results of a small simulation study demonstrating the effects of different location estimators and exploring the sensitivity of the convergence of the SSCM on the aforementioned inverse moment condition. All proofs are deferred to the Appendix.

2. Asymptotic theory

Our first result states conditions on the location estimator t_n and the population distribution F that guarantee weak and strong consistency, respectively, of the spatial sign covariance matrix with unknown location.

Theorem 1. *If there is $\alpha \geq 0$ such that*

(I) $E|X - t|^{-1/(1+\alpha)} < \infty$, *and*
(II) *there is a sequence $(t_n)_{n \in \mathbb{N}}$ of random p-vectors that satisfies*

$$n^\alpha |t_n - t| \leq T_n \quad a.s.,$$

where $(T_n)_{n \in \mathbb{N}}$ is a sequence of random variables that converges almost surely (if $\alpha > 0$) or converges almost surely to 0 (if $\alpha = 0$),

then $S_n(\mathbb{X}_n, t_n) \xrightarrow{a.s.} S(F, t)$. If $n^\alpha |t_n - t|$ is bounded in probability (for $\alpha > 0$) or converges in probability to 0 (for $\alpha = 0$), then $S_n(\mathbb{X}_n, t_n) \xrightarrow{p} S(F, t)$.

We have the following remarks concerning Theorem 1.

(I) The primary assertion of Theorem 1 is the case $\alpha = 0$. If t_n is strongly or weakly consistent for t, then $S_n(\mathbb{X}_n, t_n)$ is strongly or weakly consistent, respectively, for $S(F, t)$ if the first moment of $1/|X - t|$ is finite. Theorem 1 states further that, if there is information on the rate of convergence of t_n, i.e., (II) is fulfilled for some $\alpha > 0$, the assumption on F can be weakened, requiring less than first moments of $1/|X - t|$. Note that condition (I) of Theorem 1 gets weaker with increasing α, whereas (II) gets stronger.

(II) For any reasonable location estimate, we expect $\sqrt{n}(t_n - t)$ to converge in distribution. Then we have weak consistency of $S_n(X_n, t_n)$ if $E|X - t|^{-2/3} < \infty$.

(III) The situation is slightly different for strong consistency. If we take the mean as the location estimator, i.e. $t_n = n^{-1} \sum_{i=1}^n X_i$ and $t = E(X)$, we know by the law of the iterated logarithm that $\sqrt{n}|t_n - t|$ does not converge almost surely, but $n^\alpha |t_n - t| \xrightarrow{a.s.} 0$ for any $\alpha < 1/2$. Thus $E|X - t|^{-2/3-\delta} < \infty$ for some $\delta > 0$ is required for strong consistency.

(IV) Assumption (I) of Theorem 1 requires that the probability mass is not too strongly concentrated in the vicinity of t. This seems intuitive: for many observations X_i being very close to t, the spatial signs $s(X_i - t_n)$ and $s(X_i - t)$ will vastly differ, even if t_n is close to t. In this sense, assumption (I) accounts for the discontinuity of the sign function $s(\cdot)$ at 0. However, it is a very mild condition; it is fulfilled, e.g., if the density of F is bounded at t.

(V) A continuous, elliptical distribution F is characterized by a density f of the form

$$f(x) = \det(V)^{-\frac{1}{2}} g\{(x - \mu)^T V^{-1}(x - \mu)\} \tag{2}$$

for a function $g : [0, \infty) \to [0, \infty)$, a positive definite $p \times p$ matrix V and a p-vector μ. The parameter μ coincides with the spatial median of F as well as the mean, provided that the latter exists. Then $E|X - \mu|^{-1/(1+\alpha)} < \infty$ is fulfilled if

$$g(z) = O\big(z^{\{1/(1+\alpha)-p\}/2+\delta}\big) \quad \text{as } z \to 0$$

for some $\delta > 0$. In particular, it is always fulfilled if g is bounded at the origin, thus for instance for all normal and elliptical t distributions in any dimension. Note that the boundedness is not a necessary condition.

The next result gathers conditions that ensure the asymptotic normality of the SSCM with unknown location.

Theorem 2. *If*

(I) $E|t_n - t|^4 = O(n^{-2})$,
(II) $E|X - t|^{-3/2} < \infty$,
(III) $E\left\{\frac{X-t}{|X-t|^2}\right\} = 0$ *and* $E\left\{\frac{(X-t)^{(i)}(X-t)^{(j)}(X-t)^{(k)}}{|X-t|^4}\right\} = 0$ *for* $i, j, k = 1, \ldots, p$,

where $(X - t)^{(i)}$ *denotes the ith component of the random vector* $X - t$, *then*

$$\sqrt{n}\,\text{vec}\{S_n(\mathbb{X}_n, t_n) - S(F, t)\} \xrightarrow{d} N_{p^2}(0, W),$$

with W being defined in Section 1.

Assumption (III) of Theorem 2 imposes some form of symmetry of the distribution F around the point t. It is fulfilled, e.g., if $(X - t) \sim -(X - t)$, thus in particular for elliptical distributions with center t. If F is not symmetric around t, asymptotic normality can nevertheless be shown, but we require $S_n(\mathbb{X}_n, t)$ and t_n to converge jointly, and the asymptotic covariance matrix of $S_n(\mathbb{X}_n, t_n)$ does in general not coincide with W. This is the situation of our last theorem.

Theorem 3. *If*

(I) $E|t_n - t|^4 = O(n^{-2})$,
(II) $E|X - t|^{-3/2} < \infty$, *and*
(III) $\sqrt{n}\begin{pmatrix} t_n - t \\ \text{vec}\{S_n(\mathbb{X}_n, t) - S(F, t)\} \end{pmatrix} \xrightarrow{d} N_{p+p^2}(0, \Xi)$ *for a symmetric* $p(p + 1) \times p(p + 1)$ *matrix* Ξ,

then

$$\sqrt{n}\,\text{vec}\{S_n(\mathbb{X}_n, t_n) - S(F, t)\} \xrightarrow{d} N_{p^2}(0, A\Xi A^T)$$

with $A = (B, I_{p^2}) \in \mathbb{R}^{p^2 \times p(p+1)}$, *where* I_{p^2} *is the* p^2-*dimensional identity matrix and* $B \in \mathbb{R}^{p^2 \times p}$ *is given by*

$$B = 2E\left[\frac{\{(X - t) \otimes (X - t)\}(X - t)^T}{|X - t|^4}\right] - E\left\{\frac{X - t}{|X - t|^2}\right\} \otimes I_p - I_p \otimes E\left\{\frac{X - t}{|X - t|^2}\right\}.$$

Again, the assumptions are rather weak. If the fixed-location SSCM and the location estimator converge individually, the joint convergence (III) is usually also fulfilled, although a thorough proof may be tedious. For the sample mean, it can be seen fairly easily, and hence Theorem 3 implies that, e.g., the SSCM with sample mean location is asymptotically normal for any distribution with finite second moments and bounded density.

We have two closing remarks.

(I) In all convergence results in this section, the population distribution F is completely arbitrary except for the moment condition on $1/|X-t|$, in particular no continuity is required. If we assume F to be continuous, condition (I) of Theorem 2 can be weakened to the boundedness in probability of $\sqrt{n}(t_n - t)$, and in Theorem 3 it can be dropped altogether. The connection is given by Lemma 4 in the Appendix. However, the continuity of F is not an essential assumption for these asymptotic statements to hold, and it can, e.g., be exchanged for uniform boundedness of the fourth moments of $\sqrt{n}(t_n - t)$, which is, while being far from strict, a condition often fulfilled and easy to verify.
(II) The results of this section should be compared to the analogous ones for Tyler's shape matrix [28, Section 4]. The conditions on the population distribution F and the location estimate t_n are similar.

3. Simulations

The simulation section has two parts: in Section 3.1, we want to get an impression how well, for several n and different population distributions, the distribution of $\sqrt{n}\{S_n(\mathbb{X}_n, t_n) - S(F, t)\}$ is approximated by its Gaussian limit. We study in particular the differences when using the mean and the spatial median as the location estimator. In Section 3.2, we consider distributions with unbounded densities to examine what happens near the limit case of the inverse moment condition of Theorem 1.
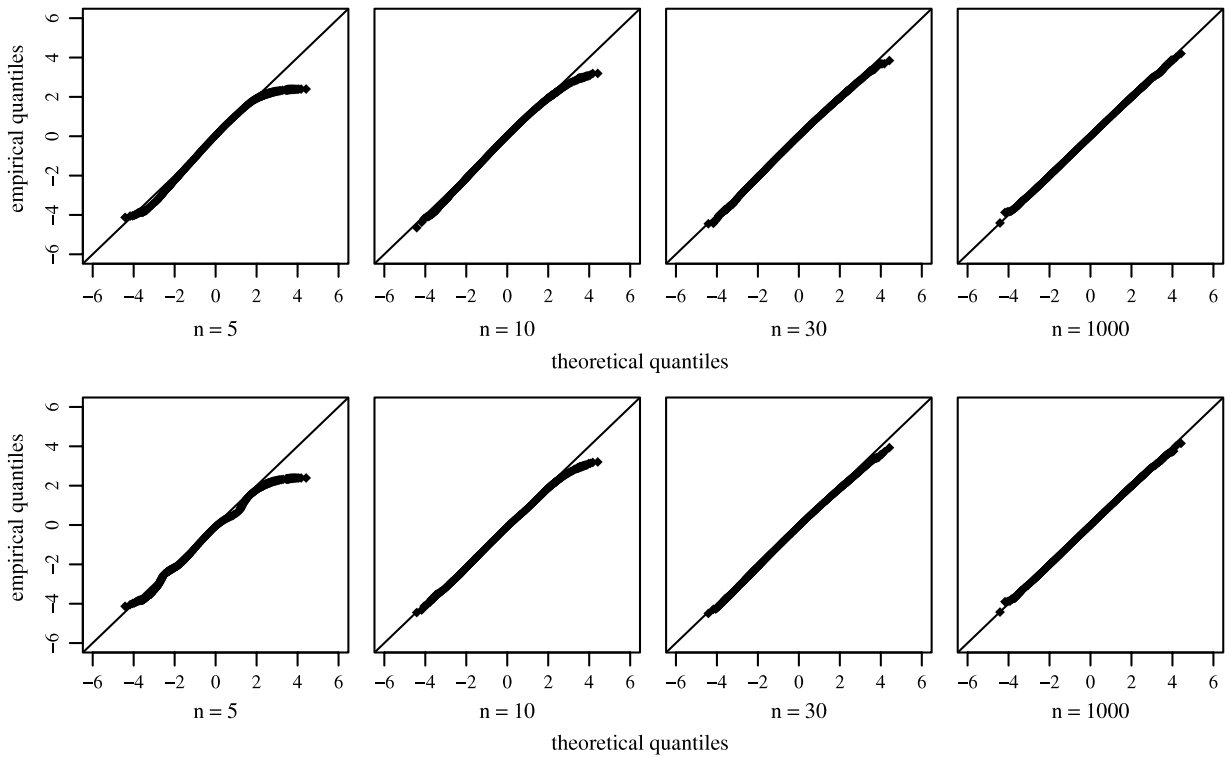
Fig. 1. QQ plots for the off-diagonal element $[\sqrt{n}\{S_n(\mathbb{X}_n, t_n) - S(F, t)\}]_{1,2}$ at the normal distribution with covariance (3), when using mean (top row) and spatial median (bottom row) as the location estimator.

3.1. Spatial median versus mean

We sample from a bivariate, centered normal distribution with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} \tag{3}$$

and from a bivariate, centered, elliptical t-distribution with two degrees of freedom and the same shape matrix Σ. The corresponding spatial sign covariance matrix is

$$S(F, 0) = \begin{pmatrix} 1/2 & 0.13397 \\ 0.13397 & 1/2 \end{pmatrix};$$

cf. [5]. With the variance of the t_2 distribution not being finite, we have that, when using the mean as the location estimator, the sequence $\sqrt{n}(t_n - t)$ does not converge. Hence we expect the SSCM not to be asymptotically normal in this case.

For both distributions and for various sample sizes, we compute the SSCM with mean and spatial median as the location estimator. For each setting, 100,000 samples are evaluated, which yields a fairly precise Monte-Carlo approximation of the actual distribution of the SSCM. In Fig. 1 we see QQ plots of the off-diagonal element $\left[\sqrt{n}\{S_n(\mathbb{X}_n, t_n) - S(F, t)\}\right]_{1,2}$ for the normal population distribution and $n = 5, 10, 30, 1000$. The simulated distribution is plotted against the asymptotic distribution, which is also given in [5]. The pictures in the first row (mean) and second row (spatial median) are very similar, and we note very little difference for the two location measures under normality. We suspect that the bumpiness of the QQ plot for the spatial median for $n = 5$ (lower left corner) is due to the spatial median's tendency to coincide with one of the data points. For both estimators, we find the distribution of $\left[\sqrt{n}\{S_n(\mathbb{X}_n, t_n) - S(F, t)\}\right]_{1,2}$ to have light tails for $n = 5$ (left column). The reason is the bounded range $[-0.5, 0.5]$ of the estimate $\{S_n(\mathbb{X}_n, t_n)\}_{1,2}$. If n is very small, this boundedness is also visible in the QQ plots of the centered and $\sqrt{n}$-scaled estimator. The asymmetry (lighter upper tail) is due to the true value 0.134 not lying in the center of the possible range. However, this departure from normality quickly vanishes as n increases, and the normal limit generally provides a very good approximation for fairly moderate sample sizes.

Fig. 2 shows analogous results for the t_2 distribution. Here we observe a qualitatively different behavior of the SSCM depending on the location estimate. In the case of the mean (first row), the off-diagonal element $[\sqrt{n}\{S_n(\mathbb{X}_n, t_n) - S(F, t)\}]_{1,2}$ is clearly non-normal with heavy tails, which persist also for large n. As for the spatial median, the results for normal and t_2 distributions are very similar: the second row of Fig. 2 is almost identical to the second row of Fig. 1.

Additional to the small bivariate example, we present some simulation results for higher dimensions. The results for various n and p are summarized in Tables 1 (normal distribution) and 2 (for the t_2 distribution). The set-up is slightly
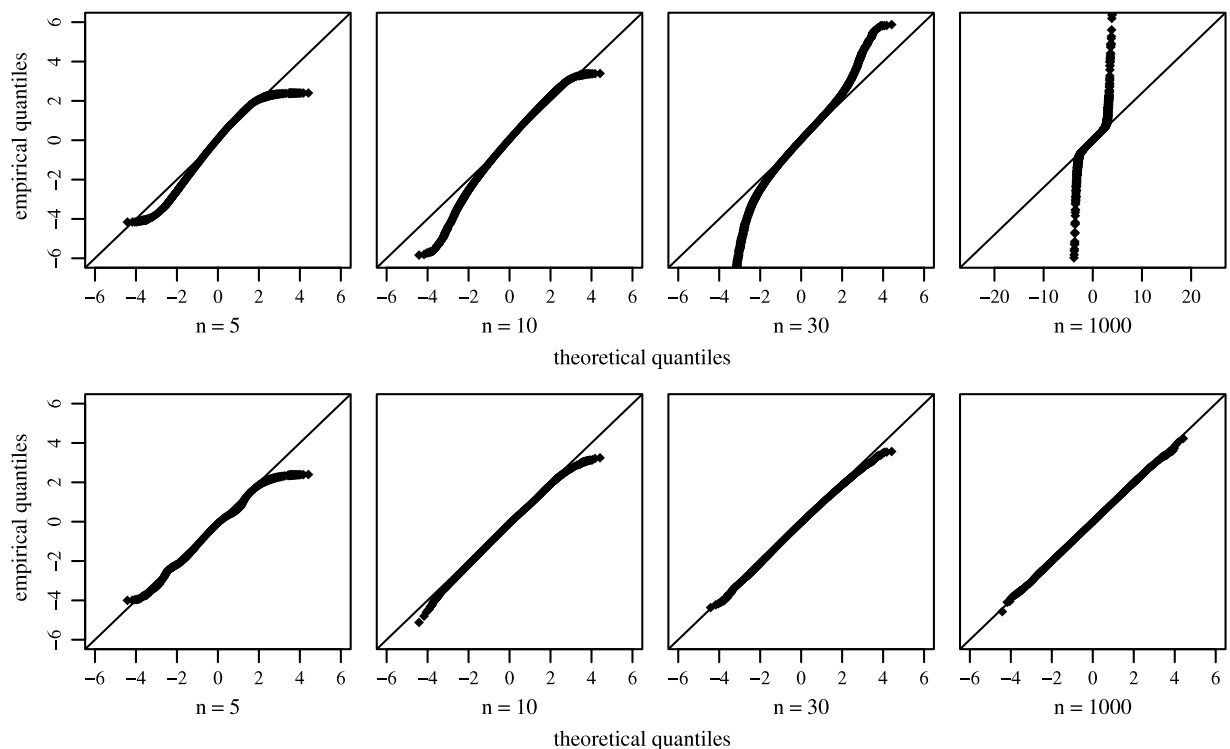
Fig. 2. QQ plots for $[\sqrt{n}\{S_n(\mathbb{X}_n, t_n) - S(F, t)\}]_{1,2}$ at t_2 distribution with shape matrix (3), when using mean (top row) and spatial median (bottom row) as the location estimator.

different: the true covariance is I_p, corresponding to $S(F, 0) = p^{-1}I_p$, and we do not concentrate on a single element, but consider the L_2 distance of the matrix estimate to the true $S(F, 0)$. The averages of $n\|S_n(\mathbb{X}_n, t_n) - p^{-1}I_p\|^2$ (based on 10,000 runs in each setting) are given. Note that the matrix distance does not blow up as p increases. But this is not surprising considering that $\|S_n(\mathbb{X}_n, t_n)\|^2$ is bounded by 1 in any dimension p. Furthermore, the SSCM with a known location is included in the comparison. For instance, at Table 1 we observe that the loss for not knowing the location is about 25% for $n = 5$ and about 3% for $n = 30$. In contrast, it appears to make practically no difference under normality which location estimator is used. This is in line with the observation that the mean and the spatial median have a similar efficiency at high-dimensional spherical normal distributions; cf. e.g. [16].

We have also carried out simulations for a variety of other elliptical distributions, with different generators g and different shape matrices V. The general picture throughout is the same that is conveyed by the examples here: choosing mean or spatial median as the location estimator makes little difference in situations where both estimators are root-n-consistent. (The differences are more pronounced for "strongly shaped" matrices V, where the spatial median is relatively less efficient.) For distributions without finite second moments, taking the sample mean substantially impairs the sample SSCM.

3.2. Singularity distribution

The second goal of our small simulation study is to assess the sensitivity of the convergence of $S_n(\mathbb{X}_n, t_n)$ with respect to condition (I) of Theorem 1, i.e., the probability mass concentration around t. For that purpose, we consider the p-variate elliptical distribution $F_{\gamma,p}$, $\gamma > 0$, with density $f_{\gamma,p}$ given by $\mu = 0$, $V = I_p$ and

$$g_{\gamma,p}(z) = c_{\gamma,p}z^{\gamma-p/2}\mathbb{1}_{[0,1]}(z);$$

cf. (2). The factor $c_{\gamma,p} = \gamma\Gamma(p/2)/\pi^{p/2}$ scales the corresponding p-variate density $f_{\gamma,p}$ to 1. If $X \sim F_{\gamma,p}$, the norm $|X|$ has, for any p, the density

$$h_\gamma(z) = 2\gamma z^{2\gamma-1}\mathbb{1}_{[0,1]}(z).$$

The densities $f_{\gamma,p}$ and h_γ are well defined for any $\gamma > 0$. A smaller value of γ corresponds to a stronger singularity at the origin, and $\gamma = 0$ constitutes the limit case, in which the densities are not integrable. Furthermore, if $-\beta/2 < \gamma$, we have $E|X|^\beta < \infty$. Thus, using the mean as the location estimator, Theorem 1 grants strong consistency for $\gamma > 1/3$.

In the simulations we consider $\gamma = 0.05, 0.1, 0.15, \ldots, 0.45$ and several sample sizes n ranging from 10 to 20 000. Fig. 3 illustrates the L_1 consistency of the SSCM depending on γ for $p = 2$ and $p = 10$: the absolute error ε of an off-diagonal

Table 1
Normal distribution: average of $n\|S_n - p^{-1}I_p\|^2$ (10,000 runs) for different samples n, dimensions p when the location is known or estimated by the mean or the spatial median.

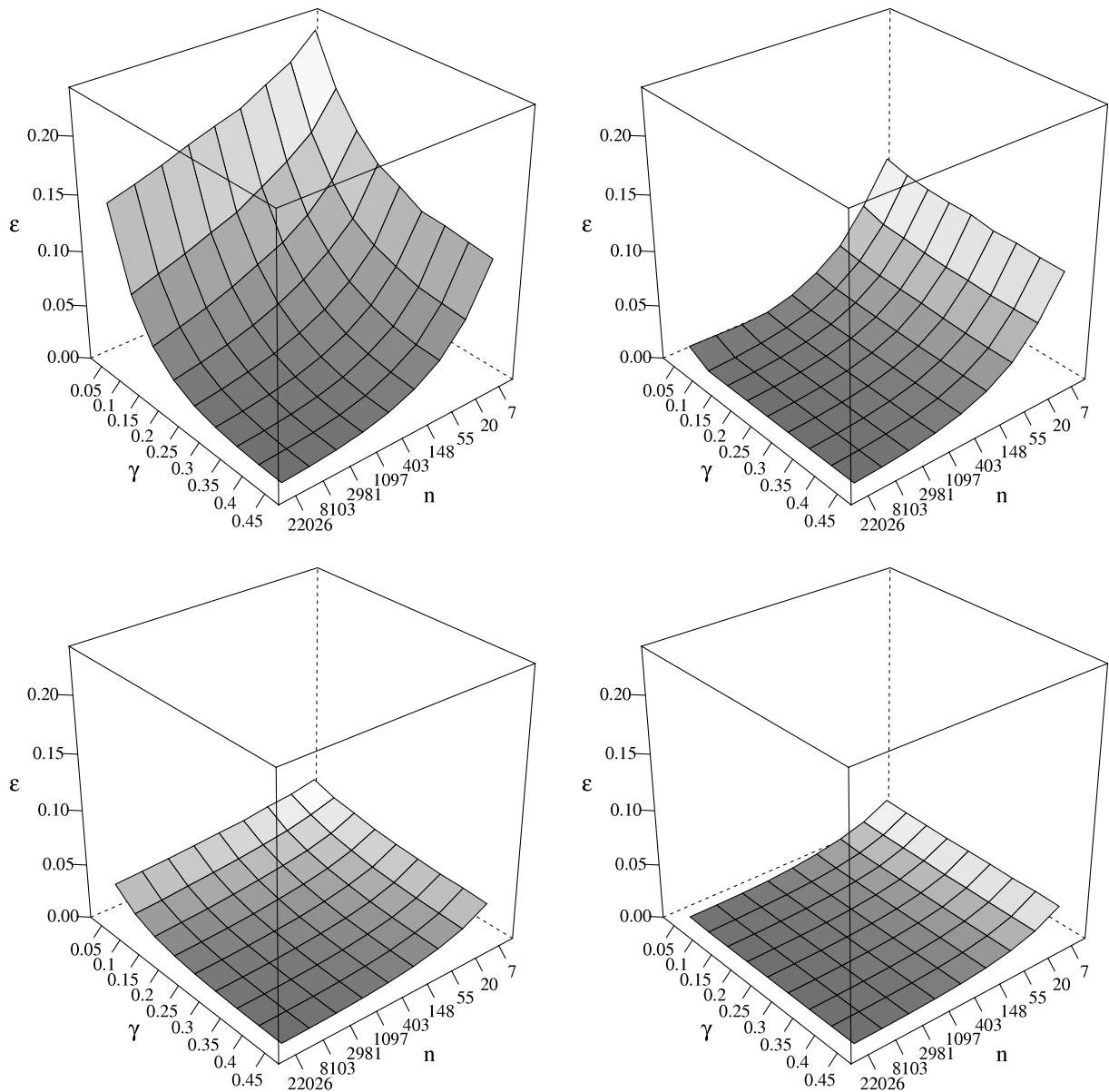| p	n											
	5			10			30			1000		
	known	mean	median	known	mean	median	known	mean	median	known	mean	median
10	0.901	1.090	1.074	0.900	0.983	0.979	0.899	0.924	0.923	0.897	0.897	0.897
50	0.981	1.217	1.213	0.980	1.085	1.084	0.979	1.012	1.012	0.980	0.981	0.981
200	0.995	1.241	1.241	0.995	1.104	1.104	0.995	1.029	1.029	0.995	0.996	0.996
1000	0.999	1.248	1.248	0.999	1.110	1.110	0.999	1.033	1.033	0.999	1.000	1.000

Fig. 3. Mean absolute error ε of the off-diagonal element of the SSCM at distribution $F_{\gamma,p}$ for $p = 2$ (top row) and $p = 10$ (bottom row) and different values of γ and n. Location estimated by mean (left) and spatial median (right).

element of the SSCM is plotted against γ and n with the mean (left) and the spatial median (right) as the location estimator. Each grid point is the average of 10,000 repetitions. The smaller errors for $p = 10$ (bottom row) are due to the fact that we examine one singly entry of the SSCM, the magnitude of which is of order p^{-2} as p increases.

Table 2
t_2 distribution: average of $n\|S_n - p^{-1}I_p\|^2$ (10,000 runs) for different n and p when the location is known or estimated by the mean or the spatial median.

p	n											
	5			10			30			1000		
	known	mean	median	known	mean	median	known	mean	median	known	mean	median
10	0.900	1.460	1.084	0.896	1.524	0.984	0.899	1.629	0.926	0.897	1.576	0.898
50	0.981	1.649	1.222	0.980	1.711	1.090	0.980	1.889	1.015	0.980	1.561	0.981
200	0.995	1.700	1.249	0.995	1.772	1.110	0.995	1.923	1.031	0.995	1.858	0.996
1000	0.999	1.705	1.256	0.999	1.753	1.116	0.999	1.930	1.036	0.999	1.662	1.000

For the mean, we observe a decline of the absolute error with increasing n for all γ considered, but the convergence is very slow for small values of γ. Also here, the SSCM with the spatial median behaves qualitatively different: the decay of the absolute error appears to be equally fast for all γ. This is a plausible observation. For the univariate median, the asymptotic efficiency is proportional to the squared density at the true median, and the boundedness of the density at the median is necessary to prove asymptotic normality. In the case of a singularity, the univariate sample median is known to converge faster than at the usual $\sqrt{n}$ rate. The situation is similar for the spatial median: the efficiency increases with higher probability mass concentration at the true spatial median μ, and the boundedness of the density is also a standard assumption for the asymptotic normality of the sample spatial median (see e.g. [21,16]). The sensitivity of the SSCM $S_n(\mathbb{X}_n, t_n)$ to a singularity at t, due to discontinuity of the spatial sign, and the increased efficiency of the spatial median in such a situation are opposing effects that seem to nullify each other. A thorough theoretical investigation of this situation seems to be an open problem.

In conclusion, we note that in all our simulations, the SSCM with the spatial median was non-inferior to the SSCM with the sample mean, with clear superiority in the case of the t_2 distribution and the singularity distribution F_γ. The use of the spatial median as the location estimator is advisable and clearly preferable to the mean. Besides the conceptual kinship of the SSCM and the spatial median and their common good robustness properties, the spatial median is also advantageous in the case of high mass concentration at the center.

All simulations were done in R 2.11.1 [24] using the packages mvtnorm [7] for generating elliptical distributions and pcaPP [6] for computing the spatial median.

Acknowledgments

Appendix. Proofs

In the following proofs, we let, without loss of generality, $t = 0$, and throughout write $S_n(t_n)$ short for $S_n(\mathbb{X}_n, t_n)$, S_n for $S_n(\mathbb{X}_n, 0)$ and S for $S(F, 0)$. The assumption $E|X|^{-1} < \infty$ is part of all three theorems, which implies $P(X = 0) = 0$. Thus, we restrict our attention to the case $X_i \neq 0$ for all $i \in \mathbb{N}$, but note that, with some notational effort, we can generalize the results to distributions having an atom at the origin. We use $\|\cdot\|$ to denote the Frobenius norm of a real matrix, i.e., $\|A\| = \{\text{tr}(A^T A)\}^{1/2}$ for any $A \in \mathbb{R}^{m \times n}$. Letting $\text{vec}\, A$ denote the mn-vector obtained by stacking the columns of $A \in \mathbb{R}^{m \times n}$ from left to right underneath each other, we have $\|A\| = |\text{vec}\, A|$.

Proof of Theorem 1. First we treat weak consistency. We show $\|S_n(t_n) - S_n\| \overset{p}{\longrightarrow} 0$ for $n \to \infty$. Associated with the random vector t_n is the following random partition of the space $\mathbb{R}^p$,

$$\mathcal{A}_n = \left\{ x \in \mathbb{R}^p \,\middle|\, |x - t_n| \geq \frac{1}{2}|x| \right\}, \qquad \mathcal{A}_n^C = \left\{ x \in \mathbb{R}^p \,\middle|\, |x - t_n| < \frac{1}{2}|x| \right\},$$

and the random partition $\mathcal{I}_n = \{1 \leq i \leq n \mid X_i \in \mathcal{A}_n\}$, $\mathcal{I}_n^C = \{1 \leq i \leq n \mid X_i \in \mathcal{A}_n^C\}$ of the index set $\{1, \ldots, n\}$. Letting $\Gamma_i = s(X_i)s(X_i)^T$ and $\Gamma_i(t_n) = s(X_i - t_n)s(X_i - t_n)^T$, $i \in \mathbb{N}$, we have

$$\|S_n(t_n) - S_n\| \leq \frac{1}{n} \sum_{i=1}^n \|\Gamma_i(t_n) - \Gamma_i\| = \frac{1}{n} \sum_{i \in \mathcal{I}_n} \|\Gamma_i(t_n) - \Gamma_i\| + \frac{1}{n} \sum_{i \in \mathcal{I}_n^C} \|\Gamma_i(t_n) - \Gamma_i\|.$$

Call the first sum on the right-hand side C_n and the second sum D_n. We show convergence to zero in probability for the random variables C_n and D_n separately, starting with C_n. For $X_i \neq 0$ and $X_i \neq t_n$ we have $\|\Gamma_i\| = \|\Gamma_i(t_n)\| = 1$ and

$$\text{tr}\left\{ \Gamma_i^T \Gamma_i(t_n) \right\} = 2 \frac{|X_i|^2 |t_n|^2 - (X_i^T t_n)^2}{|X_i - t_n|^2 |X_i|^2}$$

and consequently

$$\|\Gamma_i(t_n) - \Gamma_i\| = \frac{\left\{2\left(|X_i|^2|t_n|^2 - (X_i^T t_n)^2\right)\right\}^{1/2}}{|X_i|\,|X_i - t_n|},$$

which, for $i \in \mathcal{I}_n$, is bounded from above by $\sqrt{2}|t_n|/|X_i|$. Hence

$$C_n \leq \frac{1}{n}\sum_{i \in \mathcal{I}_n}\frac{\sqrt{2}|t_n|}{|X_i|} \leq \frac{1}{n}\sqrt{2}|t_n|\sum_{i=1}^{n}|X_i|^{-1} = \sqrt{2}\left[n^\alpha|t_n|\right]\left[\frac{1}{n^{1+\alpha}}\sum_{i=1}^{n}|X_i|^{-1}\right].$$

If $\alpha = 0$, the first factor in brackets converges in probability to zero by assumption (II) of Theorem 1, and the second factor converges almost surely to $E|X|^{-1}$ by Kolmogorov's strong law of large numbers. If $\alpha > 0$, then the first factor is bounded in probability and the second factor converges to zero almost surely by Marcinkiewicz's strong law of large numbers (e.g. [15, p. 255]).

It remains to show $D_n \xrightarrow{p} 0$. Noting that always $\|\Gamma_i(t_n) - \Gamma_i\| \leq 2$, we find that

$$D_n/2 \leq \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\mathcal{A}_n^C}(X_i). \tag{A.1}$$

We prove that the right-hand side converges to zero in probability by showing its L_1 convergence. Let $\varepsilon > 0$. Since $E|X|^{-1} < \infty$, there is a δ-ball $B_\delta = \{x \in \mathbb{R}^p \mid |x| \leq \delta\}$ around 0 with $F(B_\delta) < \varepsilon/2$, where F denotes the distribution of X. Further, since $t_n \xrightarrow{p} 0$, and $|t_n| \leq \delta/2$ implies $\mathcal{A}_n^C \subset B_\delta$, there is $n_0 \in \mathbb{N}$ such that $P(\mathcal{A}_n^C \subset B_\delta) \geq 1 - \varepsilon/2$ for all $n \leq n_0$. Thus, for every fixed $i \in \mathbb{N}$ and all $n \leq n_0$, we have

$$P(X_i \in \mathcal{A}_n^C) = P(X_i \in \mathcal{A}_n^C, \mathcal{A}_n^C \subset B_\delta) + P(X_i \in \mathcal{A}_n^C, \mathcal{A}_n^C \not\subset B_\delta)$$
$$\leq P(X_i \in B_\delta) + P(\mathcal{A}_n^C \not\subset B_\delta) \leq \varepsilon,$$

and finally $E\{n^{-1}\sum_{i=1}^{n}\mathbb{1}_{\mathcal{A}_n^C}(X_i)\} \leq n^{-1}\sum_{i=1}^{n}P(X_i \in \mathcal{A}_n^C) \leq \varepsilon$ for all $n \geq n_0$. Thus, the right-hand side of (A.1) converges in L_1, and hence in probability, to zero, and so does D_n. The proof of weak consistency is complete.

As for strong consistency, we treat C_n completely analogously. For D_n, we show, as before, that the right-hand side of (A.1) converges to zero, but now almost surely. Let, again, $\varepsilon > 0$ be arbitrary. Further, let $Z_{n,i} = \mathbb{1}_{\mathcal{A}_n^C}(X_i)$ and $\tilde{Z}_{\delta,i} = \mathbb{1}_{B_\delta}(X_i)$, where, as before, B_δ denotes the δ-ball around 0. We chose $\delta > 0$ such that $F(B_\delta) < \varepsilon$. We use $(\cdot)_+$ to denote the non-negative part, i.e., $(x)_+ = x$ for $x > 0$ and $(x)_+ = 0$ for $x \leq 0$. Then

$$\frac{1}{n}\sum_{i=1}^{n}Z_{n,i} \leq \frac{1}{n}\sum_{i=1}^{n}\tilde{Z}_{\delta,i} + \frac{1}{n}\sum_{i=1}^{n}\left(Z_{n,i} - \tilde{Z}_{\delta,i}\right)_+.$$

The first summand converges to $F(B_\delta)$, which is smaller than ε. For the second summand, we note that $|t_n| \leq \delta/2$ implies $(Z_{n,i} - \tilde{Z}_{\delta,i})_+ = 0$ for all $i \in \mathbb{N}$. Thus, since $|t_n| \xrightarrow{a.s.} 0$, we have for almost all $\omega \in \Omega$ that there is $n_0 \in \mathbb{N}$ such that $n^{-1}\sum_{i=1}^{n}(Z_{n,i}(\omega) - \tilde{Z}_{\delta,i}(\omega))_+ = 0$ for all $n \geq n_0$. Hence D_n converges almost surely to zero, and we have proven strong consistency. ∎

When dealing with the spatial sign covariance matrix thoroughly, the possibility of multiple instances of $X_i = t_n$ causes some nuisance. In the proof of Theorem 1, this is covered implicitly, since $X_i = t_n$ implies $i \in \mathcal{I}_n^C$. For Theorems 2 and 3, we state the following lemma.

Lemma 4. *Under conditions* (I) *and* (II) *of Theorem 2,* $(n - n^*)/\sqrt{n} \xrightarrow{p} 0$, *where* n^* *is defined in Section* 1.

Proof. Recall that throughout the Appendix, we let $t = 0$. Let $A = \sup_n E(\sqrt{n}|t_n|)^4$, $B = E(|X|^{-3/2})$ and $1/8 < \alpha < 1/6$. With Markov's inequality we have with condition (I) of Theorem 2, $P(|t_n| > n^{\alpha-1/2}) = P(\sqrt{n}|t_n| > n^\alpha) \leq A/n^{4\alpha}$, and with condition (II), $P(|X| < n^{\alpha-1/2}) = P(|X|^{-1} > n^{1/2-\alpha}) \leq Bn^{3\alpha/2-3/4}$. Thus

$$E\left[\frac{n - n^*}{\sqrt{n}}\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\mathbb{1}_{\{X_i=t_n\}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left[\mathbb{1}_{\{X_i=t_n, |t_n|>n^{\alpha-1/2}\}}\right] + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left[\mathbb{1}_{\{X_i=t_n, |t_n|\leq n^{\alpha-1/2}\}}\right]$$

$$\leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left[\mathbb{1}_{\{|t_n|>n^{\alpha-1/2}\}}\right] + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left[\mathbb{1}_{\{|X_i|\leq n^{\alpha-1/2}\}}\right] \leq An^{-4\alpha+1/2} + Bn^{3\alpha/2-1/4},$$

which converges to zero since $1/6 < \alpha < 1/8$. ∎

Proof of Theorem 2. We show $\sqrt{n}\{S_n(t_n) - S_n\} \xrightarrow{p} 0$ for $n \to \infty$. Due to Lemma 4 we may assume without loss of generality that $X_i \neq t_n$ for all $n, i \in \mathbb{N}$. Let, as before, $\Gamma_i = s(X_i)s(X_i)^T$ and $\Gamma_i(t_n) = s(X_i - t_n)s(X_i - t_n)^T$. Using the identity

$$\Gamma_i(t_n) = \Gamma_i + |X_i|^{-2} \cdot \left(t_n t_n^T - X_i t_n^T - t_n X_i^T\right) + \frac{2X_i^T t_n - t_n^T t_n}{|X_i|^2} \cdot \Gamma_i(t_n) \tag{A.2}$$

we have

$$\sqrt{n}\{S_n(t_n) - S_n\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\Gamma_i(t_n) - \Gamma_i)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{t_n t_n^T}{|X_i|^2} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i t_n^T}{|X_i|^2} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{t_n X_i^T}{|X_i|^2} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{2X_i^T t_n}{|X_i|^2} \Gamma_i(t_n) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{t_n^T t_n}{|X_i|^2} \Gamma_i(t_n).$$

Call the summands A_n, B_n, C_n, D_n and E_n from left to right. We will show convergence in probability to zero for each summand. For B_n and C_n we have

$$\|B_n\| = \|C_n\| \leq \left|\sqrt{n}\, t_n\right| \cdot \left|\frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{|X_i|^2}\right|.$$

The first factor is bounded in probability by assumption (I), and the second tends to 0 almost surely because of Kolmogorov's strong law of large numbers and assumption (III). For A_n we have

$$\|A_n\| \leq \|n\, t_n t_n^T\| \cdot \left(\frac{1}{n^{3/2}} \sum_{i=1}^{n} \frac{1}{|X_i|^2}\right).$$

Marcinkiewicz's strong law of large numbers guarantees convergence of the second factor to 0 almost surely as long as $E(|X|^{-2})^{2/3} < \infty$, which is fulfilled by assumption (II). Since $\|\Gamma_i(t_n)\| = 1$, the same argument applies to E_n. For D_n, we apply again identity (A.2) and obtain

$$D_n = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i^T t_n}{|X_i|^4} X_i X_i^T + \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i^T t_n}{|X_i|^4} t_n t_n^T - \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i^T t_n}{|X_i|^4} t_n X_i^T$$

$$- \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i^T t_n}{|X_i|^4} X_i t_n^T + \frac{4}{\sqrt{n}} \sum_{i=1}^{n} \frac{\left(X_i^T t_n\right)^2}{|X_i|^4} \Gamma_i(t_n) - \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \frac{\left(X_i^T t_n\right)\left(t_n^T t_n\right)}{|X_i|^4} \Gamma_i(t_n).$$

Thus D_n can be split up into six sums, which we denote by $D_{1,n}, \ldots, D_{6,n}$ from left to right in the formula above. The matrix D_{1n} is best shown to converge to zero in probability by treating it element-wise. The (l, m) element of $D_{1,n}$, $1 \leq l, m \leq p$, can be written as

$$\left(D_{1,n}\right)_{l,m} = 2 \sum_{k=1}^{p} \left(\sqrt{n}\, t_n^{(k)}\right) \cdot \left(\frac{1}{n} \sum_{i=1}^{n} \frac{X_i^{(l)} X_i^{(m)} X_i^{(k)}}{|X_i|^4}\right).$$

In each of the p summands, the first factor is bounded in probability by assumption (I), and the second converges to zero almost surely, due to assumption (III) and Kolmogorov's strong law of large numbers. The remaining terms $D_{2,n}, \ldots, D_{6,n}$ are shown to converge to zero in probability by Marcinkiewicz's strong law of large numbers. For the matrix norms of $D_{3,n}$, $D_{4,n}$ and $D_{5,n}$ we find the same upper bound as for A_n above. For $D_{2,n}$, we obtain the upper bound

$$\|D_{3,n}\| \leq 2n^{3/2} |t_n|^3 \cdot \left(\frac{1}{n^2} \sum_{i=1}^{n} \frac{1}{|X_i|^3}\right),$$

where the first factor is again bounded in probability and the second converges to zero by Marcinkiewicz's strong law of large numbers, since by assumption (III), $E(|X|^{-3})^{1/2} < \infty$. The same argument applies to $D_{6,n}$, and the proof is complete. ∎

Proof of Theorem 3. We use the same notation and the same decomposition of $\sqrt{n}\{S_n(t_n) - S_n\}$ as in the proof of Theorem 2, and also assume that $X_i \neq t_n$ for all $n, i \in \mathbb{N}$. The above expression has been split up into ten summands. Seven of these, namely $A_n, D_{2,n}, D_{3,n}, D_{4,n}, D_{5,n}, D_{6,n}, E_n$ (i.e., those terms that contain t_n in a multiplicity higher than 1) are shown to converge to zero by means of Marcinkiewicz's strong law of large numbers. We denote these seven terms by $R_n(\mathbb{X}_n, t_n)$ and focus on

the remaining three, B_n, C_n, and $D_{1,n}$. We have

$$\sqrt{n}\,\mathrm{vec}\{S_n(t_n) - S\} = \sqrt{n}\,\mathrm{vec}\{S_n - S\} + \mathrm{vec}\left\{\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i^T t_n}{|X_i|^4}X_i X_i^T - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_i t_n^T + t_n X_i^T}{|X_i|^2} + R_n(\mathbb{X}_n, t_n)\right\}$$

$$= \sqrt{n}\,\mathrm{vec}\{S_n - S\} + \left\{\frac{2}{n}\sum_{i=1}^{n}\frac{(X_i \otimes X_i)X_i^T}{|X_i|^4}\right\}\left(\sqrt{n}\,t_n\right) - \left(\sqrt{n}\,t_n\right)\otimes\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{|X_i|^2}\right\}$$

$$- \left\{\frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{|X_i|^2}\right\}\otimes\left(\sqrt{n}\,t_n\right) + \mathrm{vec}\,R_n(\mathbb{X}_n, t_n).$$

This has the same asymptotic distribution as

$$\sqrt{n}\,\mathrm{vec}\{S_n - S\} + 2E\left\{\frac{(X \otimes X)X^T}{|X|^4}\right\}\left(\sqrt{n}\,t_n\right) - \left(\sqrt{n}\,t_n\right)\otimes E\left\{\frac{X}{|X|^2}\right\} - E\left\{\frac{X}{|X|^2}\right\}\otimes\left(\sqrt{n}\,t_n\right),$$

which is a linear function of the random $p(p+1)$ vector $\xi_n = \sqrt{n}(t_n, \mathrm{vec}\{S_n - S\})$. This function is given by $A\xi_n$ with A being defined in Theorem 3. Hence, by the continuous mapping theorem, $\sqrt{n}\,\mathrm{vec}\{S_n(t_n) - S\}$ converges in distribution to $N_{p(p+1)}$ $(0, A\Xi A^T)$, and the proof is complete. ∎

References

[1] H. Bensmail, G. Celeux, Regularized Gaussian discriminant analysis through eigenvalue decomposition, J. Amer. Statist. Assoc. 91 (1996) 1743–1748.
[2] C. Croux, C. Dehon, A. Yadine, The k-step spatial sign covariance matrix, Adv. Data Anal. Classif. 4 (2010) 137–150.
[3] C. Croux, E. Ollila, H. Oja, Sign and rank covariance matrices: statistical properties and application to principal components analysis, in: Y. Dodge (Ed.), Statistical Data Analysis Based on the L_1-Norm and Related Methods (Papers of the 4th International Conference on Statistical Analysis on the L_1-norm and Related Methods, Neuchâtel, Switzerland, August 4–9, 2002), Birkhäuser, Basel, 2002, pp. 257–269.
[4] L. Dümbgen, On Tyler's M-functional of scatter in high dimension, Ann. Inst. Statist. Math. 50 (1998) 471–491.
[5] A. Dürre, D. Vogel, R. Fried, Spatial sign correlation, 2014. ArXiv e-prints 1403.7635.
[6] P. Filzmoser, H. Fritz, K. Kalcher, pcaPP: Robust PCA by Projection Pursuit. R Package Version 1.9-44, 2011.
[7] A. Genz, F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, T. Hothorn, mvtnorm: Multivariate normal and t distributions. R Package Version 0.9-9997, 2014.
[8] D. Gervini, Robust functional estimation using the median and spherical principal components, Biometrika 95 (2008) 587–600.
[9] J.C. Gower, The median centre, Appl. Stat. 23 (1974) 466–470.
[10] F.R. Hampel, The influence curve and its role in robust estimation, J. Amer. Statist. Assoc. 69 (1974) 383–393.
[11] P.J. Huber, E.M. Ronchetti, Robust Statistics, second ed., in: Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ, 2009.
[12] J.H.B. Kemperman, The median of a finite measure on a Banach space, in: Y. Dodge (Ed.), Statistical Data Analysis Based on the L_1-Norm and Related Methods, North-Holland, Amsterdam, 1987, pp. 217–230.
[13] V. Koltchinskii, R. Dudley, On spatial quantiles, in: V. Korolyuk, et al. (Eds.), Skorokhod's Ideas in Probability Theory, in: Proc. Inst. Math. Natl. Acad. Sci. Ukr., Math. Appl., vol. 32, Institute of Mathematics of NAS of Ukraine, Kiev, 2000, pp. 195–210.
[14] N. Locantore, J. Marron, D. Simpson, N. Tripoli, J. Zhang, K. Cohen, Robust principal component analysis for functional data, TEST 8 (1999) 1–73. (With comments).
[15] M. Loève, Probability Theory I, fourth ed., in: Graduate Texts in Mathematics, Springer-Verlag, New York, 1977.
[16] A. Magyar, D.E. Tyler, The asymptotic efficiency of the spatial median for elliptically symmetric distributions, Sankhyā B 73 (2011) 165–192.
[17] A.F. Magyar, D.E. Tyler, The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions, 2013. ArXiv e-prints 1309.1915.
[18] J.I. Marden, Some robust estimates of principal components, Statist. Probab. Lett. 43 (1999) 349–359.
[19] R.A. Maronna, D.R. Martin, V.J. Yohai, Robust Statistics: Theory and Methods, in: Wiley Series in Probability and Statistics, Wiley, Chichester, 2006.
[20] P. Milasevic, G.R. Ducharme, Uniqueness of the spatial median, Ann. Statist. 15 (1987) 1332–1333.
[21] J. Möttönen, K. Nordhausen, H. Oja, Asymptotic theory of the spatial median, in: J. Antoch, M. Huskova, P.K. Sen (Eds.), Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in Honor of Professor Jana Jurečková, in: IMS Collections, vol. 7, 2010, pp. 182–193.
[22] H. Oja, Multivariate Nonparametric Methods with R. An Approach based on Spatial Signs and Ranks, in: Lecture Notes in Statistics, vol. 199, Springer, New York, 2010.
[23] H. Oja, S. Sirkiä, J. Eriksson, Scatter matrices and independent component analysis, Austral. J. Statist. 35 (2006) 175–189.
[24] R Development Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0, 2010.
[25] P.J. Rousseeuw, C. Croux, Alternatives to the median absolute deviation, J. Amer. Statist. Assoc. 88 (1993) 1273–1283.
[26] S. Sirkiä, S. Taskinen, H. Oja, Symmetrised m-estimators of multivariate scatter, J. Multivariate Anal. 98 (2007) 1611–1629.
[27] S. Sirkiä, S. Taskinen, H. Oja, D.E. Tyler, Tests and estimates of shape based on spatial signs and ranks, J. Nonparametr. Stat. 21 (2009) 155–176.
[28] D.E. Tyler, A distribution-free M-estimator of multivariate scatter, Ann. Statist. 15 (1987) 234–251.
[29] D.E. Tyler, F. Critchley, L. Dümbgen, H. Oja, Invariant co-ordinate selection, J. R. Stat. Soc. Ser. B Stat. Methodol. 71 (2009) 549–592. (with discussion).
[30] Y. Vardi, C.H. Zhang, A modified weiszfeld algorithm for the Fermat–Weber location problem, Math. Program. A 90 (2001) 559–566.
[31] S. Visuri, V. Koivunen, H. Oja, Sign and rank covariance matrices, J. Statist. Plann. Inference 91 (2000) 557–575.
[32] S. Visuri, H. Oja, V. Koivunen, Subspace-based direction-of-arrival estimation using nonparamatric statistics, IEEE Trans. Signal Process. 49 (2001) 2060–2073.
[33] E. Weiszfeld, F. Plastria, On the point for which the sum of the distances to n given points is minimum, Ann. Oper. Res. 167 (2009) 7–41.
[34] Y. Zuo, Robust location and scatter estimators in multivariate analysis, in: J. Fan, H. Koul (Eds.), Frontiers in Statistics. Dedicated to Peter John Bickel in Honor of his 65th Birthday, Imperial College Press, London, 2006, pp. 467–490.