Biometrika (2014), **101**, 3, *pp.* 673–688
Printed in Great Britain
doi: 10.1093/biomet/asu020
Advance Access publication 23 July 2014

The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions

BY ANDREW F. MAGYAR AND DAVID E. TYLER

Department of Statistics and Biostatistics, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, U.S.A.

afmagyar@gmail.com dtyler@rci.rutgers.edu

SUMMARY

The asymptotic efficiency of the spatial sign covariance matrix relative to affine equivariant estimators of scatter is studied. In particular, the spatial sign covariance matrix is shown to be asymptotically inadmissible, i.e., the asymptotic covariance matrix of the consistency-corrected spatial sign covariance matrix is uniformly larger than that of its affine equivariant counterpart, namely Tyler's scatter matrix. Although the spatial sign covariance matrix has often been recommended when one is interested in principal components analysis, its inefficiency is shown to be most severe in situations where principal components are of greatest interest. Simulation shows that the inefficiency of the spatial sign covariance matrix also holds for small sample sizes, and that the asymptotic relative efficiency is a good approximation to the finite-sample efficiency for relatively modest sample sizes.

Some key words: Affine equivariance; Eigenprojection; Principal components analysis; Relative efficiency; Robustness; Tyler's scatter matrix.

1. INTRODUCTION

Many robust alternatives to the sample covariance matrix have been proposed, such as M-estimators (Maronna, 1976; Huber, 1977), the minimum volume ellipsoid and minimum covariance determinant estimators (Rousseeuw, 1985), the Stahel–Donoho estimators (Hampel et al., 1986; Donoho & Gasko, 1992), S-estimators (Davies, 1987; Lopuhaä, 1989), P-estimators (Maronna et al., 1992; Tyler, 1994), CM-estimators (Kent & Tyler, 1996) and MM-estimators (Tatsuoka & Tyler, 2000; Tyler, 2002). All of these estimators of scatter are affine equivariant, and except for the M-estimators, they all have high breakdown points. On the other hand, M-estimators can be computed via a simple iterative reweighing algorithm, whereas the high-breakdown-point scatter estimators are computationally intensive, especially for large samples and/or high-dimensional datasets, and are usually obtained via approximate or probabilistic algorithms.

Due to the computational complexity of high-breakdown-point affine equivariant estimators of multivariate scatter, there has been recent interest in high-breakdown-point estimators of scatter which are not affine equivariant but are computationally easy. One such estimator is the spatial sign covariance matrix. The earliest and often-cited reference to this estimator is Locantore et al. (1999), although the name itself was first introduced by Visuri et al. (2000). The spatial sign covariance matrix is often recommended as a fast and easy high-breakdown-point estimator; see, e.g., Locantore et al. (1999) and Maronna et al. (2006). In particular, since it is orthogonally equivariant, it has been recommended in cases where only orthogonal equivariance and not full affine equivariance is needed, such as in principal component analysis.

The lack of affine equivariance of the spatial sign covariance matrix arises because the estimator is defined by downweighing observations based on their Euclidean distances from an estimated centre of the dataset. When the data are presumed to arise from a multivariate normal distribution or, more generally, an elliptically symmetric distribution, one might conjecture that this estimator would be relatively inefficient whenever the elliptical distribution is far from spherical, i.e., when the components of the multivariate vector are highly correlated. This can be problematic, since one usually implements multivariate procedures in situations where one suspects strong correlations between the variables, as is the case for principal components analysis. Our goal here is to study the efficiency of the spatial sign covariance matrix for general covariance structures. The results of this paper can be viewed as being complementary to those of Magyar & Tyler (2012), in which the inefficiency of the spatial median, a popular orthogonally equivariant estimator of location, is studied in detail under the elliptical model.

An important property of the spatial sign covariance matrix is that its asymptotic distribution does not depend on the particular elliptical family being sampled. This is also true of its finite-sample distribution when the centre of symmetry of the elliptical distribution is known. An affine equivariant estimator possessing these same properties is the distribution-free M-estimator of scatter proposed by Tyler (1987a), commonly referred to as Tyler's scatter matrix, which can be considered to be an affine equivariant version of the spatial sign covariance matrix. The asymptotic relative efficiency of the spatial sign covariance matrix to an affine equivariant estimator of scatter can be factored into two parts, with the first part being the asymptotic relative efficiency of the spatial sign covariance matrix to Tyler's scatter matrix and the second part being the asymptotic relative efficiency of Tyler's scatter matrix to the affine equivariant scatter matrix of interest. For an elliptically symmetric distribution, the first factor depends only on the underlying covariance structure and not on the particular elliptical family, whereas the second factor depends only on the particular elliptical family and not on the underlying covariance structure. Hence, the use of Tyler's scatter matrix serves as a convenient benchmark for understanding the effect that the lack of affine equivariance has on the efficiency of the spatial sign covariance matrix under different covariance structures.

2. Preliminaries

2·1. *Elliptical distributions and equivariance*

Elliptically symmetric distributions provide a simple generalization of the multivariate normal distribution and are often used to ascertain how multivariate statistical methods perform outside the normal family. An elliptically symmetric distribution in $\mathbb{R}^d$ is defined to be a distribution arising from an affine transformation of a spherically symmetric distribution, i.e., if $z \sim Qz$ for any $d \times d$ orthogonal matrix Q, then the distribution of $x = Az + \mu$ is said to have an elliptically symmetric distribution with centre $\mu \in \mathbb{R}^d$ and scatter matrix $\Gamma = AA^{\mathrm{T}}$; see, e.g., Bilodeau & Brenner (1999, § 13.2). The distribution of x is characterized by μ, Γ and the distribution of z. The distribution of z itself can be characterized as $z \sim R_G u_d$, with u_d having a uniform distribution on the unit d-dimensional sphere independent of its radial component R_G, a non-negative random variable with distribution function G. In particular, $(u_d, R_G) \sim (z/\|z\|, \|z\|)$, where the norm refers to the Euclidean norm on $\mathbb{R}^d$. Thus, we denote the distribution of $x \sim R_G \Gamma^{1/2} u_d + \mu$ by $\mathcal{E}_d(\mu, \Gamma; G)$, where $\Gamma^{1/2}$ refers to the unique positive-definite square root of Γ.

The scatter parameter Γ is well-defined only up to a scalar multiple, i.e., if Γ satisfies the definition of a scatter matrix for a given elliptically symmetric distribution, then so does $\lambda\Gamma$ for any $\lambda > 0$. If no restrictions are placed on the function G, then the parameter Γ is confounded with G. If x possesses first moments, then $\mu = E(x)$; and if it possesses second moments, then $\Gamma \propto \Sigma$, the covariance matrix, or, more specifically, $\Sigma = \lambda(G)\Gamma$ with $\lambda(G) = E(R_G^2)/d$.

The parameters of an elliptically symmetric distribution are affine equivariant; that is, if $x \sim \mathcal{E}_d(\mu, \Gamma; G)$, then $Bx + b \sim \mathcal{E}_d(B\mu + b, B\Gamma B^{\mathrm{T}}; G)$ for any nonsingular $d \times d$ matrix B and any vector $b \in \mathbb{R}^d$. Consequently, the maximum likelihood estimators for the parameters of an elliptically symmetric distribution based on a given function G are affine equivariant. Given a d-dimensional sample $x_1, \ldots, x_n$, estimators of μ and Γ are said to be affine equivariant if an affine transformation of the data $x_i \to Bx_i + b$ $(i = 1, \ldots, n)$ induces on the estimators the transformations $\hat{\mu} \to B\hat{\mu} + b$ and $\hat{\Gamma} \to B\hat{\Gamma}B^{\mathrm{T}}$. In addition to the maximum likelihood estimators associated with elliptically symmetric distributions, which include the sample mean vector and the sample covariance matrix, many proposed robust estimators of multivariate location and scatter, such as the multivariate M-estimators, are affine equivariant.

If it exists, the covariance matrix of an affine equivariant estimator of scatter under random sampling from an $\mathcal{E}_d(\mu, \Gamma; G)$ distribution has a relatively simple form, namely

$$\mathrm{var}\{\mathrm{vec}(\hat{\Gamma})\} = \sigma_{1,n,\hat{\Gamma},G}(I_{d^2} + K_{d,d})(\Gamma \otimes \Gamma) + \sigma_{2,n,\hat{\Gamma},G}\mathrm{vec}(\Gamma)\mathrm{vec}(\Gamma)^{\mathrm{T}}.$$

Here $\mathrm{vec}(A)$ denotes the pq-dimensional vector obtained from stacking the columns of the $p \times q$ matrix A, $K_{d,d}$ denotes the $d^2 \times d^2$ commutation matrix, and $\otimes$ is the Kronecker product; this notation is reviewed in Appendix A. The terms $\sigma_{1,n,\hat{\Gamma},G}$ and $\sigma_{2,n,\hat{\Gamma},G}$ are scalars that depend on the sample size, the particular estimator $\hat{\Gamma}$ and the underlying family of elliptical distributions, i.e., G, but not on μ or Γ; see Tyler (1983) for more details. When focusing on the shape of the scatter parameter, i.e., on Γ up to a scalar multiple, the second scalar becomes unimportant, at least asymptotically. That is, if $\hat{\Gamma}$ is consistent and asymptotically normal, and if the p-dimensional function H satisfies $H(\Gamma) = H(\beta\Gamma)$ for any $\beta > 0$, e.g., $H(\Gamma) = \mathrm{vec}(\Gamma)/\mathrm{tr}(\Gamma)$, then

$$n^{1/2}\{H(\hat{\Gamma}) - H(\Gamma)\} \to N_p\{0, \sigma_{1,\hat{\Gamma},G}M_H(\Gamma)\} \tag{1}$$

in distribution, where M_H is a function dependent on H with $M_H(\beta\Gamma) = M_H(\Gamma)$ for any $\beta > 0$, and $\sigma_{1,\hat{\Gamma},G}$ is again a scalar that depends only on the particular estimator and on G. The asymptotic relative efficiency at $\mathcal{E}_d(\mu, \Gamma; G)$ of any shape component based on one affine equivariant asymptotically normal estimator of scatter $\hat{\Gamma}$ versus another estimator $\tilde{\Gamma}$ then reduces to the scalar $\mathrm{ARE}\{H(\hat{\Gamma}), H(\tilde{\Gamma})\} = \sigma_{1,\tilde{\Gamma},G}/\sigma_{1,\hat{\Gamma},G}$, which depends neither on the shape function H nor on the parameters μ and Γ. We again refer the reader to Tyler (1983) for more details.

2·2. *Spatial sign covariance matrix*

Given a d-dimensional sample $x_1, \ldots, x_n$, the spatial sign covariance matrix about a point μ is defined to be

$$\hat{\mathcal{S}}(\mu) = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - \mu)(x_i - \mu)^{\mathrm{T}}}{(x_i - \mu)^{\mathrm{T}}(x_i - \mu)} = \frac{1}{n}\sum_{i=1}^{n}\theta_i\theta_i^{\mathrm{T}}, \tag{2}$$

where $\theta_i = (x_i - \mu)/\|x_i - \mu\|$. By definition, $\mathrm{tr}\{\hat{\mathcal{S}}(\mu)\} = 1$. For general random samples, the law of large numbers implies that $\hat{\mathcal{S}}(\mu)$ is consistent for $\Xi = E\{\hat{\mathcal{S}}(\mu)\} = E(\theta\theta^{\mathrm{T}})$, where $\theta = (x - \mu)/\|x - \mu\|$. Also, since $\|\theta\| = 1$, the central limit theorem immediately gives

$$n^{1/2}\,\mathrm{vec}\{\hat{\mathcal{S}}(\mu) - \Xi\} \to N_{d^2}(0, V_{\mathcal{S}}) \tag{3}$$

in distribution, where $V_{\mathcal{S}} = \mathrm{var}\{\mathrm{vec}(\theta\theta^{\mathrm{T}})\} = \mathrm{var}(\theta \otimes \theta) = \{E(\theta\theta^{\mathrm{T}} \otimes \theta\theta^{\mathrm{T}}) - \mathrm{vec}(\Xi)\mathrm{vec}(\Xi)^{\mathrm{T}}\}$. When the centre is estimated by $\hat{\mu}_n$, the asymptotic distribution of the spatial sign covariance matrix, denoted by $\hat{\mathcal{S}}_n = \hat{\mathcal{S}}(\hat{\mu}_n)$, has recently been shown to be the same as that of $\hat{\mathcal{S}}(\mu)$; see Dürre et al. (2014). Since $\hat{\mathcal{S}}_n$ depends on the directions of $x_i - \hat{\mu}_n$, it is not surprising to find that the distribution of x_i cannot be too concentrated about μ in order for this equivalence to hold.

LEMMA 1. *Suppose that $x \in \mathbb{R}^d$ is a continuous random vector symmetrically distributed about μ, i.e., $(x - \mu) \sim -(x - \mu)$, with $E(\|x - \mu\|^{-3/2}) < \infty$. For a random sample $x_1, \ldots, x_n$ of the random vector x, suppose that $n^{1/2}(\hat{\mu}_n - \mu) = O(1)$ in probability; then $n^{1/2}\{\hat{\mathcal{S}}_n - \hat{\mathcal{S}}(\mu)\} \to 0$ in probability.*

Under random sampling from $\mathcal{E}_d(\mu, \Gamma; G)$, the finite-sample distribution of $\hat{\mathcal{S}}(\mu)$, and hence its asymptotic distribution, as well as that of $\hat{\mathcal{S}}_n$, does not depend on the radial component G and depends only on Γ up to a scalar multiple. This is because if $x \sim \mathcal{E}_d(\mu, \Gamma; G)$, then $\theta \sim \text{ACG}_d(\Gamma)$, an angular central Gaussian distribution with parameter Γ, which does not depend on G, see, e.g., Watson (1983) and Tyler (1987b); furthermore, $\text{ACG}_d(\Gamma) \sim \text{ACG}_d(\beta\Gamma)$ for any $\beta > 0$. It is known that Ξ is not proportional to Γ and so, in contrast to affine equivariant estimators of scatter, the spatial sign covariance matrix is not consistent for any multiple of Γ; see, e.g., Boente & Fraiman (1999). Consequently, a shape component of the spatial sign covariance matrix is not necessarily consistent for the corresponding shape component of Γ. However, a transformation of $\hat{\mathcal{S}}_n$ can be made so that it is consistent for some multiple of Γ; this is discussed in more detail in § 3.

Although not affine equivariant, $\hat{\mathcal{S}}_n$ is orthogonally equivariant provided the estimator of location, $\hat{\mu}_n$, is orthogonally equivariant. In other words, if for any orthogonal matrix Q and any vector $b \in \mathbb{R}^d$ the transformation $x_i \to Qx_i + b$ $(i = 1, \ldots, n)$ induces the transformation $\hat{\mu}_n \to Q\hat{\mu}_n + b$, then it also induces the transformation $\hat{\mathcal{S}}_n \to Q\hat{\mathcal{S}}_n Q^{\mathrm{T}}$.

2·3. *Tyler's scatter matrix*

For the multivariate sample $x_1, \ldots, x_n$, Tyler (1987a) introduced the distribution-free M-estimator of multivariate scatter about a point μ as a solution to the implicit equation

$$\hat{\mathcal{T}}(\mu) = \frac{d}{n}\sum_{i=1}^{n} \frac{(x_i - \mu)(x_i - \mu)^{\mathrm{T}}}{(x_i - \mu)^{\mathrm{T}}\hat{\mathcal{T}}(\mu)^{-1}(x_i - \mu)} = \frac{d}{n}\sum_{i=1}^{n} \frac{\theta_i\theta_i^{\mathrm{T}}}{\theta_i^{\mathrm{T}}\hat{\mathcal{T}}(\mu)^{-1}\theta_i}, \tag{4}$$

where θ_i is defined as in (2). In the aforementioned paper, a solution $\hat{\mathcal{T}}(\mu)$ to (4) is shown to exist under general conditions and to be readily computed via an iterative reweighing algorithm. The solution is unique up to a scalar multiple, and so can be made unique by demanding, for example, that $\text{tr}\{\hat{\mathcal{T}}(\mu)\} = 1$ or $\det\{\hat{\mathcal{T}}(\mu)\} = 1$. Under general random sampling, $\hat{\mathcal{T}}(\mu)$ is asymptotically normal, and when μ is replaced by a consistent estimator $\hat{\mu}_n$, $\hat{\mathcal{T}}_n = \hat{\mathcal{T}}(\hat{\mu}_n)$ is asymptotically equivalent to $\hat{\mathcal{T}}(\mu)$ under the same conditions as in Lemma 1. If $\hat{\mu}_n$ is affine equivariant, then so is $\hat{\mathcal{T}}_n$ in the sense that an affine transformation of the data $x_i \to Bx_i + b$ $(i = 1, \ldots, n)$ induces the transformation $\hat{\mathcal{T}}_n \to \hat{\mathcal{T}}_n^* \propto B\hat{\mathcal{T}}_n B^{\mathrm{T}}$.

As with the spatial sign covariance matrix, under random sampling from $\mathcal{E}_d(\mu, \Gamma; G)$, the finite-sample distribution of $\hat{\mathcal{T}}(\mu)$ and the asymptotic distributions of $\hat{\mathcal{T}}(\mu)$ and $\hat{\mathcal{T}}_n$ depend neither on the radial component G nor on μ, and depend on Γ only up to a scalar multiple. This follows since $\hat{\mathcal{T}}(\mu)$ is a function of $\theta_1, \ldots, \theta_n$, a random sample from $\text{ACG}_d(\Gamma)$. The asymptotic distribution of a shape component of $\hat{\mathcal{T}}_n$ under random sampling from an elliptical distribution is given by (1) with $\sigma_{1,\hat{\mathcal{T}},G} = 1 + 2/d$, which does not depend on G.

3. Theoretical results

3·1. *The inadmissibility of the spatial sign covariance matrix*

Hereafter, it is assumed that $x_1, \ldots, x_n$ is a random sample from an $\mathcal{E}_d(\mu, \Gamma; G)$ distribution. As noted in § 2·2, the spatial sign covariance matrix is a consistent estimator of Ξ, but Ξ itself is not a multiple of Γ. However, $\hat{\mathcal{S}}_n$ can be transformed into a consistent estimator of a multiple of Γ in the following manner. Consider the spectral value decomposition for Γ, namely $\Gamma = Q \Lambda Q^{\mathrm{T}}$ where Q is an orthogonal matrix whose columns consist of an orthonormal set of eigenvectors of Γ, and Λ is a diagonal matrix whose diagonal elements $\lambda_1 \geqslant \cdots \geqslant \lambda_d$ are the ordered eigenvalues of Γ. The spectral value decomposition of the matrix Ξ, on the other hand, is known to have the form $\Xi = Q \Delta Q^{\mathrm{T}}$ where Q is the same as that for Γ, and Δ is a diagonal matrix whose elements $\phi_1 \geqslant \cdots \geqslant \phi_d$ are the ordered eigenvalues of Ξ. The relationship between Δ and Λ is given by

$$\phi_j = E\left(\frac{\lambda_j \, \chi^2_{1,j}}{\sum_{r=1}^{d} \lambda_r \, \chi^2_{1,r}} \right) \quad (j = 1, \ldots, d), \tag{5}$$

with $\chi^2_{1,1}, \ldots, \chi^2_{1,d}$ being mutually independent chi-squared variates on one degree of freedom; see, e.g., Boente & Fraiman (1999) or Taskinen et al. (2012). Hence the eigenvalues of the spatial sign covariance matrix, which consistently estimate the eigenvalues of Ξ, are not consistent estimators of the eigenvalues of Γ. However, (5) implies that Δ is a function of Λ, which we write as $\Delta = h(\Lambda)$. Without loss of generality, since Γ is confounded with the radial component G, we assume that Γ is normalized so that $\mathrm{tr}(\Gamma) = 1$; and since $\mathrm{tr}(\Xi) = 1$, the function h can be viewed as a function from $\mathbb{R}^{d-1}$ to $\mathbb{R}^{d-1}$. The function h can be shown to be one-to-one and continuously differentiable. This then implies that $h(\Gamma) = Q h(\Lambda) Q^{\mathrm{T}}$ is one-to-one and continuously differentiable. Hence, $\hat{\mathcal{W}}_n = h^{-1}(\hat{\mathcal{S}}_n)$ is asymptotically equivalent to $\hat{\mathcal{W}}(\mu) = h^{-1}\{\hat{\mathcal{S}}(\mu)\}$, which in turn is a consistent and asymptotically normal estimator of Γ, i.e.,

$$n^{1/2}\, \mathrm{vec}(\hat{\mathcal{W}}_n - \Gamma) \sim n^{1/2}\, \mathrm{vec}\{\hat{\mathcal{W}}(\mu) - \Gamma\} \to N_{d^2}\{0,\ V_{\mathcal{W}}(\Gamma)\} \tag{6}$$

in distribution. If we normalize Tyler's scatter matrix so that $\mathrm{tr}(\hat{\mathcal{T}}_n) = 1$, then

$$n^{1/2}\, \mathrm{vec}(\hat{\mathcal{T}}_n - \Gamma) \sim n^{1/2}\, \mathrm{vec}\{\hat{\mathcal{T}}(\mu) - \Gamma\} \to N_{d^2}\{0,\ V_{\mathcal{T}}(\Gamma)\} \tag{7}$$

in distribution. The forms of the asymptotic variances are discussed later. From the general theory of maximum likelihood estimation, it can be shown that the spatial sign covariance matrix is asymptotically inadmissible in the sense stated in the following theorem. Here, for two symmetric matrices of order d, the notation $A_1 > A_2$ means that $A_1 - A_2$ is positive definite, and $A_1 \geqslant A_2$ means that $A_1 - A_2$ is positive semidefinite; this is the usual partial ordering for symmetric matrices.

Theorem 1. *Let $x_1, \ldots, x_n$ represent a random sample from a $\mathcal{E}_d(\mu, \Gamma; G)$ distribution. Then, for $V_{\mathcal{W}}(\Gamma)$ and $V_{\mathcal{T}}(\Gamma)$ defined in (6) and (7), respectively, $V_{\mathcal{W}}(\Gamma) \geqslant V_{\mathcal{T}}(\Gamma)$ for all $\Gamma > 0$ and $V_{\mathcal{W}}(\Gamma) \neq V_{\mathcal{T}}(\Gamma)$ for some $\Gamma > 0$.*

Proof. It is sufficient to compare $\hat{\mathcal{W}}(\mu)$ with $\hat{\mathcal{T}}(\mu)$. As noted previously, both of these estimators are functions of $\theta_1, \ldots, \theta_n$, a random sample from $\mathrm{ACG}_d(\Gamma)$. As shown in Tyler (1987b), the maximum likelihood estimator of Γ is $\hat{\mathcal{T}}(\mu)$. The $\mathrm{ACG}_d(\Gamma)$ distribution satisfies sufficient regularity conditions for its maximum likelihood estimator to be asymptotically efficient, i.e., its asymptotic covariance matrix equals the inverse of its Fisher information matrix. Furthermore, it can be shown that the joint limiting distributions in (6) and (7) are jointly multivariate normal. Consequently, the theorem follows from the general theory of maximum likelihood estimators; see, e.g., Theorem 4.8 in Lehmann & Casella (1998). □

Remark 1. Consider the spherical case, where $\Gamma \propto I_d$. For this case, $\Xi = E(\theta\theta^{\mathrm{T}}) = d^{-1} I_d$ and $E(\theta\theta^{\mathrm{T}} \otimes \theta\theta^{\mathrm{T}}) = \{d(d+2)\}^{-1}\{I_{d^2} + K_{d,d} + \mathrm{vec}(I_d)\mathrm{vec}(I_d)^{\mathrm{T}}\}$; see, e.g., Tyler (1987a). Consequently, the asymptotic variance of the spatial sign covariance matrix given in (3) can be expressed as $V_{\mathcal{S}} = \{d(d+2)\}^{-1}M$, where $M = \{I_{d^2} + K_{d,d} - (2/d)\mathrm{vec}(I_d)\mathrm{vec}(I_d)^{\mathrm{T}}\}$. By comparison, $V_{\mathcal{T}} = \{(d+2)/d^3\}M$ for this case (Tyler, 1987a). Although $\{d(d+2)\}^{-1} < \{(d+2)/d^3\}$, this does not contradict the asymptotic optimality of the maximum likelihood estimator $\hat{\mathcal{T}}_n$, since $\hat{\mathcal{S}}_n$ itself is not consistent for $\Gamma/\mathrm{tr}(\Gamma)$ when $\Gamma \not\propto I_d$. Moreover, for the sample covariance matrix S_n, under multivariate normal sampling we have $n^{1/2}\,\mathrm{vec}\{S_n/\mathrm{tr}(S_n) - d^{-1}I_d\} \to N_{d^2}(0, d^{-2}M)$, but $d^{-2} > \{d(d+2)\}^{-1}$. So the spatial sign covariance matrix can be viewed as a super-efficient estimator of shape when $\Gamma \propto I_d$, but at a cost of inconsistency otherwise.

3·2. *Asymptotic calculations*

Theorem 1 states that the asymptotic variance of a consistency-adjusted spatial sign covariance matrix is greater than that of Tyler's scatter matrix. In this section, the degree of this inefficiency is studied. Even though the unadjusted spatial sign covariance matrix is an inconsistent estimator of Γ, its eigenvectors are consistent estimators of the eigenvectors of Γ. This consistency result, along with the orthogonal equivariance of the spatial sign covariance matrix, have led others, most notably Locantore et al. (1999), Marden (1999) and Maronna et al. (2006), to recommend using the eigenvectors of the spatial sign covariance matrix as robust estimators of the principal component directions. Therefore, here we focus on the eigenvectors. Ironically, principal components are primarily of interest when the eigenvalues of Γ are well separated; but, as shown here, this is the case in which the spatial sign covariance matrix is least efficient.

Before presenting our results, additional notation is needed. Recall the spectral value decompositions $\Gamma = Q\Lambda Q^{\mathrm{T}}$ and $\Xi = Q\Delta Q^{\mathrm{T}}$, with $\lambda_1 \geqslant \cdots \geqslant \lambda_d$ and $\phi_1 \geqslant \cdots \geqslant \phi_d$ being the eigenvalues of Γ and Ξ, respectively, and where the columns of Q are the corresponding eigenvectors of both Γ and Ξ. The asymptotic distribution of the estimated eigenvectors depends on the multiplicities of the eigenvalues, so we shall denote the distinct eigenvalues of Γ by $\lambda_{(1)} > \cdots > \lambda_{(m)}$ and those of Ξ by $\phi_{(1)} > \cdots > \phi_{(m)}$, with respective multiplicities $d_1, \ldots, d_m$. Equation (5) implies that both the order and the multiplicities of the eigenvalues of Γ and Ξ are the same. The matrix of eigenvectors $Q = (q_1, \ldots, q_d)$ is not uniquely defined, especially when multiple roots exist. However, the subspace spanned by the eigenvectors associated with $\lambda_{(j)}$, as well as the corresponding eigenprojection onto that subspace, is uniquely defined; that is, the spectral value decompositions have the unique representations $\Gamma = \sum_{j=1}^{m} \lambda_{(j)} P_j$ and $\Xi = \sum_{j=1}^{m} \phi_{(j)} P_j$, where $P_j = \sum_{k=m_j+1}^{m_j+d_j} q_k q_k^{\mathrm{T}}$, with $m_1 = 0$ and $m_j = d_1 + \cdots + d_{j-1}$ for $j = 2, \ldots, m$. Consequently, it is simpler to work with the asymptotic distribution of the eigenprojections rather than the asymptotic distribution of the eigenvectors themselves.

Let the spectral value decompositions for $\hat{\mathcal{T}}_n$ and $\hat{\mathcal{S}}_n$ be represented by $\hat{\mathcal{T}}_n = \hat{Q}_{\mathcal{T}}\hat{\Lambda}\hat{Q}_{\mathcal{T}}^{\mathrm{T}}$ and $\hat{\mathcal{S}}_n = \hat{Q}_{\mathcal{S}}\hat{\Delta}\hat{Q}_{\mathcal{S}}^{\mathrm{T}}$, respectively. For continuous multivariate distributions, when $n > d$, the eigenvalues of $\hat{\mathcal{T}}_n$ and the eigenvalues of $\hat{\mathcal{S}}_n$ are distinct with probability 1. Let $\hat{q}_{\mathcal{T},k}$ and $\hat{q}_{\mathcal{S},k}$ $(k = 1, \ldots, d)$ represent the corresponding normalized eigenvectors, i.e., the columns of $\hat{Q}_{\mathcal{T}}$ and $\hat{Q}_{\mathcal{S}}$, respectively. Since eigenprojections are analytic functions of their matrix argument, see, e.g., Kato (1966), consistent estimators for the eigenprojections P_j are

$$\hat{P}_{\mathcal{T},j} = \sum_{k=m_j+1}^{m_j+d_j} \hat{q}_{\mathcal{T},k}\hat{q}_{\mathcal{T},k}^{\mathrm{T}}, \quad \hat{P}_{\mathcal{S},j} = \sum_{k=m_j+1}^{m_j+d_j} \hat{q}_{\mathcal{S},k}\hat{q}_{\mathcal{S},k}^{\mathrm{T}} \quad (j = 1, \ldots, m).$$

The asymptotic distribution of the eigenprojection estimators can be obtained using the delta method, giving $n^{1/2}\,\mathrm{vec}(\hat{P}_{\mathcal{T},j} - P_j) \to N_{d^2}\{0, V_{\mathcal{T},j}(\Gamma)\}$ and $n^{1/2}\,\mathrm{vec}(\hat{P}_{\mathcal{S},j} - P_j) \to N_{d^2}\{0, V_{\mathcal{S},j}(\Gamma)\}$. The asymptotic covariance matrices, derived in Appendix B, are

$$V_{\mathcal{T},j}(\Gamma) = \sum_{k=1,k\neq j}^{m} \alpha_{\mathcal{T},j,k} M_{j,k}, \quad V_{\mathcal{S},j}(\Gamma) = \sum_{k=1,k\neq j}^{m} \alpha_{\mathcal{S},j,k} M_{j,k}, \tag{8}$$

where $M_{j,k} = (1/2)(I + K_{d,d})(P_j \otimes P_k + P_k \otimes P_j)$. The scalars $\alpha_{\mathcal{T},j,k}$ and $\alpha_{\mathcal{S},j,k}$ are functions of the eigenvalues of Γ and are given in (B4) of Appendix B.

The forms of the asymptotic covariance matrices given in (8) correspond to their spectral value decompositions. The matrices $M_{j,k}$ are symmetric idempotent matrices, i.e., they are the eigenprojections associated with both $V_{\mathcal{T},j}(\Gamma)$ and $V_{\mathcal{S},j}(\Gamma)$. The corresponding eigenvalues are $\alpha_{\mathcal{T},j,k}$ and $\alpha_{\mathcal{S},j,k}$, with multiplicities $d_j d_k$ for $j \neq k$, $k = 1, \ldots, m$. The ranks of the $d^2 \times d^2$ matrices $V_{\mathcal{T},j}(\Gamma)$ and $V_{\mathcal{S},j}(\Gamma)$ are therefore $d_j(d - d_j)$. From Theorem 1 it follows that $V_{\mathcal{T},j}(\Gamma) \leqslant V_{\mathcal{S},j}(\Gamma)$, and consequently $\alpha_{\mathcal{T},j,k} \leqslant \alpha_{\mathcal{S},j,k}$. We believe that the last inequality is always strict. It is difficult, however, to derive this inequality directly from (B4).

To gain insight into the asymptotic inefficiency of the spatial sign covariance matrix as a function of Γ, we hereafter consider the special case where Γ has only two distinct eigenvalues. That is, suppose $m = 2$, with $\lambda_{(1)}$ and $\lambda_{(2)}$ having multiplicities d_1 and $d_2 = d - d_1$, respectively. The asymptotic efficiencies depend on the values of $\lambda_{(1)}$ and $\lambda_{(2)}$ only through their ratio $\rho^2 = \lambda_{(2)}/\lambda_{(1)}$. It is only necessary to consider the eigenprojection P_1, since $P_2 = I_d - P_1$. The corresponding asymptotic covariance matrices reduce to $V_{P_{\mathcal{T},1}}(\Gamma) = \alpha_{\mathcal{T}}(\rho)M_{1,2}$ and $V_{P_{\mathcal{S},1}}(\Gamma) = \alpha_{\mathcal{S}}(\rho)M_{1,2}$, where $\alpha_{\mathcal{T}}(\rho) = \alpha_{\mathcal{T},1,2}$ and $\alpha_{\mathcal{S}}(\rho) = \alpha_{\mathcal{S},1,2}$. Hence $V_{P_{\mathcal{T},1}}(\Gamma) \propto V_{P_{\mathcal{S},1}}(\Gamma)$, and so the asymptotic efficiency of $\hat{P}_{\mathcal{S},1}$ relative to $\hat{P}_{\mathcal{T},1}$, or equivalently the asymptotic efficiency of $\hat{P}_{\mathcal{S},2}$ relative to $\hat{P}_{\mathcal{T},2}$, reduces to the scalar comparison $\mathrm{ARE}_{d,d_1}(\hat{P}_{\mathcal{S},1}, \hat{P}_{\mathcal{T},1}; \rho) = \alpha_{\mathcal{T}}(\rho)/\alpha_{\mathcal{S}}(\rho)$. As shown in Appendix C, the asymptotic relative efficiency can then be expressed as

$$\mathrm{ARE}_{d,d_1}(\hat{P}_{\mathcal{S},1}, \hat{P}_{\mathcal{T},1}; \rho) = \frac{{}_2F_1^2\left(1, \frac{d_2+2}{2}; \frac{d+4}{2}; 1-\rho^2\right)}{{}_2F_1\left(2, \frac{d_2+2}{2}; \frac{d+4}{2}; 1-\rho^2\right)}, \tag{9}$$

where ${}_2F_1(a, b, c; x)$ denotes a Gauss hypergeometric function.

The asymptotic relative efficiency (9) simplifies in the two-dimensional case. For $d = 2$, $d_1 = 1$ and $d_2 = 1$, it is shown in Appendix C that

$$\mathrm{ARE}_{2,1}(\hat{P}_{\mathcal{S},1}, \hat{P}_{\mathcal{T},1}; \rho) = \frac{4\rho}{(1+\rho)^2}, \tag{10}$$

which goes to 0 as ρ approaches 0. Small values of ρ, though, correspond to situations where one would be most interested in principal components analysis.

Plots of the asymptotic relative efficiencies (9) as a function of ρ in the three- and five-dimensional cases are given in Fig. 1. For $d = 3$, the special case we are considering has two subcases, namely $\lambda_2 = \lambda_3$ and $\lambda_1 = \lambda_2$, or $(d_1, d_2) = (1, 2)$ and $(2, 1)$, respectively. As in two dimensions, the asymptotic relative efficiency is low when ρ is close to 0. Of the two scatter

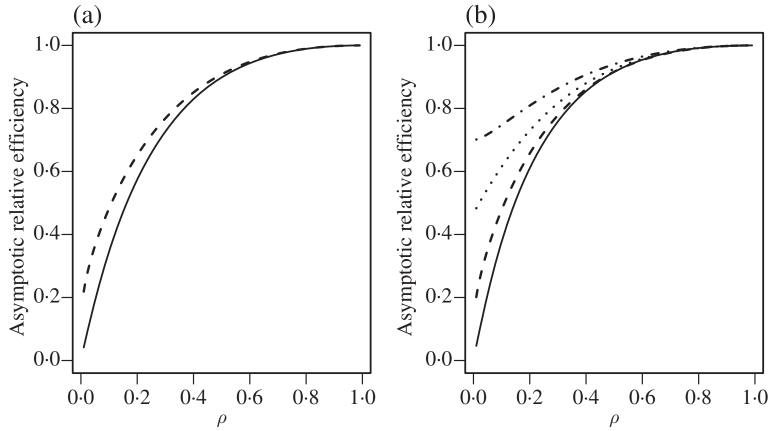

Fig. 1. Asymptotic efficiencies (9) of the eigenprojection of the spatial sign covariance matrix relative to Tyler's scatter matrix: (a) for $d = 3$, the curves correspond to $d_1 = 1$ (solid) and $d_1 = 2$ (dashed); (b) for $d = 5$, the curves correspond to $d_1 = 1$ (solid), $d_1 = 2$ (dashed), $d_1 = 3$ (dotted) and $d_1 = 4$ (dot-dash).

structures considered, the case of $(d_1, d_2) = (1, 2)$ is less favourable for the spatial sign covariance matrix. For $d = 5$, the subcases are $(d_1, d_2) = (1, 4)$, $(2, 3)$, $(3, 2)$ and $(4, 1)$. From the plots it can be observed that the larger the dimension of the main principal component space, i.e., the space associated with the larger root, the higher the asymptotic relative efficiency, with the asymptotic relative efficiency in the last subcase not being too low even as ρ approaches 0. This indicates that not much is lost by using the spatial sign covariance matrix in this last subcase, even under a nearly singular scatter structure. However, as the dimension of the main principal component space decreases, the asymptotic relative efficiency steadily decreases, and can be quite poor for nearly singular scatter structures.

In general, since ${}_2F_1(a, b; c; 0) = 1$, the asymptotic relative efficiency (9) goes to 1 as $\rho \to 1$. Also, as shown in Appendix C,

$$\lim_{\rho \to 0} \mathrm{ARE}_{d,d_1}(\hat{P}_{\mathcal{S},1}, \hat{P}_{\mathcal{T},1}; \rho) = \begin{cases} \left(1 + \frac{2}{d}\right)\left(1 - \frac{2}{d_1}\right), & d_1 > 2, \\ 0, & d_1 = 1, 2. \end{cases} \tag{11}$$

A small value of ρ implies that the main principal component space is well separated from the principal component space associated with the smaller root. For this case, in high dimensions, i.e., as $d \to \infty$, for (11) to be greater than 0·80 or 0·90, the dimension of the main principal component space needs to be approximately $d_1 \geqslant 10$ or $d_1 \geqslant 20$, respectively. Consequently, the spatial sign covariance matrix is not very efficient in cases where principal components analysis is used to reduce the dimensionality of the data to lower than 10.

4. FINITE-SAMPLE PERFORMANCE

The results of the previous section showed that the asymptotic inefficiency of the spatial sign covariance matrix relative to Tyler's scatter matrix can be quite severe under an elliptical model which is far from spherical. In the finite-sample setting, the behaviour of either scatter estimator depends on the choice of the estimator of location. Also, when estimating location, the finite-sample distributions are dependent on the particular elliptical family $\mathcal{E}_d(\mu, \Gamma; G)$, i.e., on

G. Here we focus on the estimator based on known μ. For this case, the finite-sample distributions of $\hat{\mathcal{S}}(\mu)$ and $\hat{\mathcal{T}}(\mu)$ do not depend on G or on μ. Consequently, for simulation purposes, we only need to consider random sampling from the multivariate normal distributions $N_d(0, \Gamma)$, even if the true underlying elliptical distribution possesses no moments.

As in § 3·2, we consider only the eigenprojections of the estimators, and we do so in the simple, yet informative, case where there are only two distinct eigenvalues with multiplicities d_1 and $d_2 = d - d_1$. For this case, the asymptotic variances of the eigenprojections of $\hat{\mathcal{S}}(\mu)$ and $\hat{\mathcal{T}}(\mu)$ under an elliptical distribution are proportional to each other. The finite-sample covariance matrices of these eigenprojection estimators do not, however, possess the same form as their asymptotic counterparts and are not necessarily proportional to each other. A natural way to compare the eigenprojection estimators to the theoretical eigenprojections is to use the concept of principal, or canonical, angles between subspaces.

The notation used in the next paragraphs follows that of Miao & Ben-Israel (1992). Principal angles can be used to describe how far apart one linear subspace is from another. Let L and M be linear subspaces of $\mathbb{R}^d$ with $\dim(L) = l \leqslant \dim(M) = m$. The principal angles between the subspaces L and M, $0 \leqslant \vartheta_1 \leqslant \cdots \leqslant \vartheta_l \leqslant \pi/2$, are the maximal invariants under orthogonal transformations of the subspaces, and are given in Afriat (1957) by

$$\cos\vartheta_i = \frac{\langle x_i, y_i\rangle}{\|x_i\|\,\|y_i\|} = \max\left\{\frac{\langle x, y\rangle}{\|x\|\,\|y\|} : x \in L,\ x \perp x_k,\ y \in M,\ y \perp y_k,\ k = 1, \ldots, i-1\right\}.$$

If $l = m = 1$, then the sole principal angle is simply the smallest angle between two lines. When $L = M$, the principal angles are all zero. In general, the number of nonzero principal angles is at most $\min(l, d-l)$, with the nonzero principal angles between L and M being the same as those between $L^\perp$ and $M^\perp$. The following lemma presents a result given in Björck & Golub (1973), which is useful for computing the principal angles between two linear subspaces.

LEMMA 2. *Let the columns of $Q_L \in \mathbb{R}^{n\times l}$ and $Q_M \in \mathbb{R}^{n\times m}$ be orthonormal bases for L and M, respectively, and let $\sigma_1 \geqslant \cdots \geqslant \sigma_l \geqslant 0$ be the singular values of $Q_M^{\mathrm{T}} Q_L$. Then $\cos\vartheta_i = \sigma_i$ for $i = 1, \ldots, l$. Also, $\sigma_{l-k} > \sigma_{l-k+1} = \cdots = \sigma_l = 0$ if and only if $\dim(L \cap M) = k$.*

Furthermore, if P_L and P_M represent the orthogonal projections onto the linear subspaces L and M, respectively, then the cosines of the principal angles between L and M correspond to the square roots of the eigenvalues of $P_L P_M$.

Now consider random samples of size n from a $N_d(0, \Gamma)$ distribution, where Γ has two distinct eigenvalues with multiplicities d_1 and $d_2 = d - d_1$. By orthogonal and scale equivariance, one can also assume without loss of generality that Γ is a diagonal matrix with the first d_1 diagonal elements equal to $\gamma > 1$ and the last d_2 elements equal to 1, i.e.,

$$\Gamma_o \propto \mathrm{diag}(\underbrace{\gamma, \ldots, \gamma}_{d_1}, \underbrace{1, \ldots, 1}_{d-d_1}) \tag{12}$$

with γ repeated d_1 times. The eigenprojection associated with the largest root of Γ_o, namely P_1, then has its upper left $d_1 \times d_1$ block equal to the identity matrix of order d_1, with all the other blocks equal to zero. The principal angles between the space spanned by an estimated eigenspace, say $\hat{P}_1$, and the space spanned by P_1 can then be computed by taking the arccosines of the square roots of the eigenvalues of the upper left $d_1 \times d_1$ block of $\hat{P}_1$. For the special case of $d_1 = 1$, the sole principal angle is then the arccosine of the absolute value of the first element of the normalized eigenvector estimate.

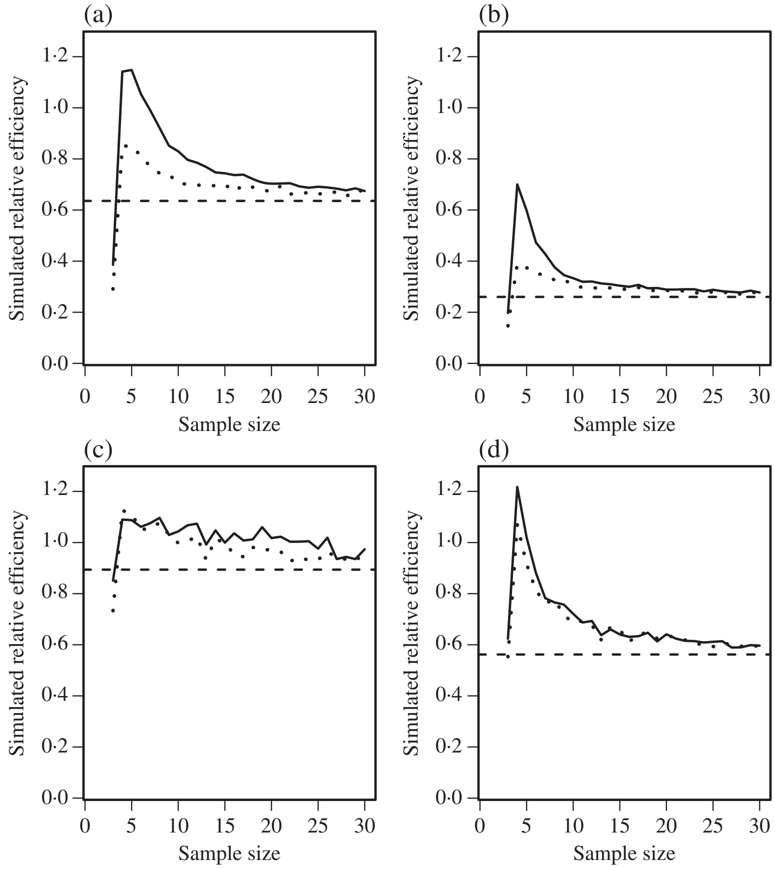

Fig. 2. Simulated finite-sample relative efficiencies $\text{RE}_{1,n}$ (solid) and $\text{RE}_{2,n}$ (dotted), given by (13), in dimension $d = 3$ for the scatter structure Γ_o, given by (12), of the eigenprojections of the spatial sign covariance matrix to those of Tyler's scatter matrix: (a) $d_1 = 1$, $\gamma = 18$; (b) $d_1 = 1$, $\gamma = 198$; (c) $d_1 = 2$, $\gamma = 4{\cdot}5$; (d) $d_1 = 2$, $\gamma = 49{\cdot}5$. The dashed horizontal lines correspond to the asymptotic relative efficiencies (9).

Simulations were undertaken for scatter structures of the form Γ_o in various dimensions. For Tyler's scatter matrix, due to its affine equivariance, it is only necessary to implement simulations for the case where $\Gamma = I_d$. For a fixed sample size and dimension, 10 000 datasets were generated from a $N_d(0, I_d)$ distribution, and $\hat{\mathcal{T}} = \hat{\mathcal{T}}(0)$ was calculated for each dataset. By affine equivariance, the simulated estimates based on other scatter structures Γ_o were obtained by applying the transformation $\hat{\mathcal{T}} \to \Gamma_o^{1/2} \hat{\mathcal{T}} \Gamma_o^{1/2}$. For the spatial sign covariance matrix, since it is only scale and orthogonally equivariant, it is necessary to implement a different simulation for each scatter structure considered. For a fixed sample size, dimension and scatter structure Γ_o, 10 000 datasets were generated from a $N_d(0, \Gamma_o)$ distribution, and $\hat{\mathcal{S}} = \hat{\mathcal{S}}(0)$ was calculated for each dataset. The principal angles between $\hat{P}_{\mathcal{T},1}$ and P_1, as well as the principal angles between $\hat{P}_{\mathcal{S},1}$ and P_1, were then calculated as previously described.

Denote the principal angles between $\hat{P}_{\mathcal{T},1}$ and P_1 by $\hat{\tau}_1 \geqslant \cdots \geqslant \hat{\tau}_{d_1} \geqslant 0$ and the principal angles between $\hat{P}_{\mathcal{S},1}$ and P_1 by $\hat{\omega}_1 \geqslant \cdots \geqslant \hat{\omega}_{d_1} \geqslant 0$. Since large deviations of the principal angles from zero indicate a poor estimator, the sum of the squared principal angles gives a measure of the loss for an eigenprojection estimator. A measure of the relative efficiency of $\hat{P}_{\mathcal{S},1}$ to $\hat{P}_{\mathcal{T},1}$ is then obtained by comparing the size of $T_n = \hat{\tau}_1^2 + \cdots + \hat{\tau}_{d_1}^2$ to that of $\Omega_n = \hat{\omega}_1^2 + \cdots + \hat{\omega}_{d_1}^2$ using some measure of central tendency. Here, we consider both the expected value and the median and thus define the following two measures of finite-sample relative efficiency:

$$\text{RE}_{1,n}(\hat{P}_{\mathcal{S},1}, \hat{P}_{\mathcal{T},1}; \Gamma) = \frac{E(T_n)}{E(\Omega_n)}, \quad \text{RE}_{2,n}(\hat{P}_{\mathcal{S},1}, \hat{P}_{\mathcal{T},1}; \Gamma) = \frac{\text{med}(T_n)}{\text{med}(\Omega_n)}. \tag{13}$$

We show in Appendix D that the limiting value of $\text{RE}_{2,n}$ is equal to the asymptotic relative efficiency given by (9). We conjecture that this also holds for $\text{RE}_{1,n}$.

Figure 2 shows simulated values of the relative efficiencies (13) for dimension $d = 3$ and sample sizes $n = d, \ldots, 30$. The two possible subcases here are $d_1 = 1$ and $d_1 = 2$, for each of which there is only one nonzero principal angle. Two different values of γ are considered in both situations, which correspond to the major principal component space explaining 90% and 99% of the total variance. For $d_1 = 1$ the corresponding values of γ are 18 and 198, and for $d_1 = 2$ they are 4·5 and 49·5. Extensive simulations for other values of γ, as well as for dimensions $d = 2$ and $d = 5$, have also been carried out. The results tell a similar story to the three-dimensional case, and so are not reported here. In all the panels of Fig. 2, it can be observed that the finite-sample relative efficiency quickly decreases to the asymptotic relative efficiency, indicated by the dashed horizontal line. For smaller sample sizes, the finite-sample relative efficiency using the median-based measure, $\text{RE}_{2,n}$, is considerably smaller than that using the expected value-based measure, $\text{RE}_{1,n}$. For relatively small sample sizes, the deficiency of the spatial sign covariance matrix is not as pronounced as for larger sample sizes. The one exception to this trend is when $n = d$. In this case, it is known that Tyler's matrix is proportional to the sample covariance matrix, and that the sample covariance matrix is proportional to the maximum likelihood estimator for the shape matrix Γ not only under the multivariate normal model but also under any elliptical model; see Tyler (2010).

5. Concluding remarks

Tyler's scatter matrix serves as a convenient benchmark for understanding the effect that the lack of affine equivariance has on the performance of the spatial sign covariance matrix. When performing random sampling from $\mathcal{E}_d(\mu, \Gamma; G)$, the asymptotic efficiency of a shape component of the consistency-adjusted spatial sign covariance matrix relative to that of an affine equivariant estimator of scatter, say $\hat{\Gamma}$, is obtained by simply multiplying the asymptotic efficiency of the sign covariance matrix relative to Tyler's scatter matrix by the scalar

$$\text{ARE}\{H(\hat{\mathcal{T}}), H(\hat{\Gamma})\} = \frac{\sigma_{1,\hat{\Gamma},G}}{\sigma_{1,\hat{\mathcal{T}},G}} = \left(\frac{d}{d+2}\right)\sigma_{1,\hat{\Gamma},G}, \tag{14}$$

where $\sigma_{1,\hat{\Gamma},G}$ is defined as in (1). The factor (14) is not dependent on the shape function H or on Γ, but it is dependent on the underlying elliptical family, i.e., on the function G. On the other hand, the asymptotic efficiency of the spatial sign covariance matrix relative to Tyler's scatter matrix is not dependent on G but is dependent on H and Γ. By Theorem 1, it follows that $\hat{\Gamma}$ is asymptotically more efficient than the spatial sign covariance matrix over the class of all elliptical distributions for which (14) is at most 1. Expressions for $\sigma_{1,\hat{\Gamma},G}$ for the M-estimators of scatter are given in Tyler (1983). For the high-breakdown-point S-estimators and reweighted S-estimators, expressions for $\sigma_{1,\hat{\Gamma},G}$ are given in Lopuhaä (1989) and Lopuhaä (1999), respectively. For the sample covariance matrix, $\sigma_{1,S,G} = (1 + 2/d)E(R_G^4)/E(R_G^2)^2$, which equals unity under multivariate normality.

The main advantage of the spatial sign covariance matrix over a high-breakdown-point affine equivariant scatter estimator is its computational simplicity. The main advantage of the spatial sign covariance matrix over computationally simple affine equivariant scatter estimators, such as the M-estimators of scatter, is its higher breakdown point. The spatial sign covariance matrix has a breakdown point of 1 whenever the location is held fixed. When an estimator of location is used, the breakdown point of the spatial sign covariance matrix is the same as the breakdown point of the location statistic, which is at most $1/2$ for translation equivariant location statistics. On the other hand, the M-estimators of scatter have breakdown points of at best $1/d$, with Tyler's estimator of scatter achieving a breakdown point of $1/d$. Dümbgen & Tyler (2005) show, however, that the breakdown of the M-estimators of scatter occurs only under very specific types of contamination. It is also worth noting, as argued in Davies & Gather (2005), that the concept of the breakdown point of a scatter matrix is not as meaningful outside the affine equivariant setting. For example, the sample covariance matrix S_n can easily be adjusted to have breakdown point $1/2$ by expressing it in terms of its spectral value decomposition, $S_n = Q_n \Delta_n Q_n^{\mathrm{T}}$, and then defining $S_n^* = Q_n \Delta_n^* Q_n^{\mathrm{T}}$ where Δ_n^* is a diagonal matrix consisting of the squares of the marginal median absolute deviations of $Q_n^{\mathrm{T}} x_i$. Note that S_n and S_n^* have the same eigenvectors, but S_n^* is only orthogonally equivariant and has breakdown point $1/2$. Finally, we remark that for affine equivariant scatter estimators, the concept of the breakdown point and the breakdown point at the edge are equivalent; see Hampel et al. (1986) and the 1981 ETH Zurich PhD thesis by W. A. Stahel. This is not the case for scatter statistics that are not affine equivariant. The breakdown point at the edge of a scatter statistic is the multivariate equivalent of the exact fit property in regression. It roughly corresponds to the proportion of contamination one needs to add to a lower-dimensional dataset in order for the scatter statistics to become nonsingular. If the scatter statistic is not affine equivariant, then the breakdown at the edge is known to be zero. This again suggests that the spatial sign covariance matrix is not particularly robust when the underlying scatter structure is nearly singular.

APPENDIX

A. REVIEW OF SOME MATRIX ALGEBRA

The notation used in this paper is as follows. If A is a $p \times q$ matrix, then $\mathrm{vec}(A)$ is the pq-dimensional vector formed by stacking the columns of A. If A is a $p \times q$ matrix and B is a $r \times s$ matrix, then the Kronecker product of A and B is the $pr \times qs$ partitioned matrix $A \otimes B = [a_{jk} B]$. The commutation matrix $K_{a,b}$ is the $ab \times ab$ matrix $\sum_{i=1}^{a} \sum_{j=1}^{b} J_{ij} \otimes J_{ij}^{\mathrm{T}}$, where J_{ij} is an $a \times b$ matrix with a 1 in the (i, j) position and zeros elsewhere. Algebraic properties involving the vec transformation, the Kronecker product and the commutation matrix can be found in Bilodeau & Brenner (1999), for example. Some important properties, which we will use without further reference, are the following:

$$\begin{gathered} \mathrm{vec}(ABC) = (C^{\mathrm{T}} \otimes A)\mathrm{vec}(B), \quad (A \otimes B)(C \otimes D) = (AC \otimes BD), \\ K_{d,d}^2 = I_{d^2}, \quad K_{r,p}(A \otimes B) = (B \otimes A)K_{s,q}, \quad K_{p,q}\mathrm{vec}(A) = \mathrm{vec}(A^{\mathrm{T}}). \end{gathered} \tag{A1}$$

More details on these derivations can be found in the 2012 Rutgers University PhD thesis by A. Magyar.

B. ASYMPTOTIC DISTRIBUTION OF EIGENPROJECTIONS OF A SCATTER ESTIMATOR

We give the proof of equation (8) in this section. To begin with, results from Tyler (1981) for the asymptotic distribution of eigenprojections in general will be partially reviewed and extended. Let M be a $d \times d$ positive-definite symmetric matrix with eigenvalues $\eta_{(1)} > \cdots > \eta_{(m)}$ having multiplicities $d_1, \ldots, d_m$ and corresponding eigenprojections $P_1, \ldots, P_m$, respectively. Suppose that M_n is a sequence of random positive-definite symmetric matrices such that $n^{1/2}(M_n - M) \to N$ in distribution, with $\mathrm{vec}(N)$

having a $N_{d^2}(0, V_M)$ distribution. Similar to the definitions of $\hat{P}_{\mathcal{T},j}$ and $\hat{P}_{\mathcal{S},j}$, let $\hat{P}_j$ denote the consistent estimator of P_j. It then follows from Theorem 4.1 of Tyler (1981) that

$$n^{1/2}(\hat{P}_j - P_j) \to P_j N(M - \eta_{(j)} I_d)^+ + (M - \eta_{(j)} I_d)^+ N P_j$$

in distribution, where A^+ denotes the Moore–Penrose generalized inverse of A. Using the properties given in (A1), and noting that $\mathrm{vec}(N) = \mathrm{vec}(N^{\mathrm{T}}) = K_{d,d}\mathrm{vec}(N)$, one obtains

$$n^{1/2}\,\mathrm{vec}(\hat{P}_j - P_j) \to (I_{d^2} + K_{d,d})\{P_j \otimes (M - \eta_{(j)} I_d)^+ \mathrm{vec}(N)\}$$

in distribution. Hence $n^{1/2}\,\mathrm{vec}(\hat{P}_j - P_j) \to N_{d^2}(0, V_{P_j})$ in distribution, where

$$V_{P_j} = (I_{d^2} + K_{d,d})\{P_j \otimes (M - \eta_{(j)} I_d)^+\} V_M \{P_j \otimes (M - \eta_{(j)} I_d)^+\}(I_{d^2} + K_{d,d}). \tag{B1}$$

Suppose now that V_M is of the form

$$V_M = \sigma_1 (I_{d^2} + K_{d,d})(M \otimes M) + \sigma_2 \mathrm{vec}(M)\mathrm{vec}(M)^{\mathrm{T}}, \tag{B2}$$

as is the case when M_n is an affine equivariant scatter estimator based on a random sample from an elliptical distribution (Tyler, 1983); then (B1) becomes

$$V_{P_j} = \sigma_1 (I_{d^2} + K_{d,d}) \sum_{k=1,k\neq j}^{m} \frac{\eta_{(j)}\eta_{(k)}}{(\eta_{(k)} - \eta_{(j)})^2} (P_j \otimes P_k + P_k \otimes P_j). \tag{B3}$$

To verify this last result, observe that $M P_j = \eta_{(j)} P_j$, $P_j P_j = P_j$, and $P_j P_k = 0$ for $k \neq j$. So, for $k \neq j$,

$$(P_j \otimes P_k)\mathrm{vec}(M) = \mathrm{vec}(P_k M P_j) = 0, \quad (P_j \otimes P_k)(M \otimes M)(P_j \otimes P_k) = \eta_{(j)}\eta_{(k)}(P_j \otimes P_k),$$

and $(P_j \otimes P_k) K_{d,d} (M \otimes M)(P_j \otimes P_k) = K_{d,d}(P_k \otimes P_j)(M \otimes M)(P_j \otimes P_k) = 0$. Hence

$$(I_{d^2} + K_{d,d})(P_j \otimes P_k) V_M (P_j \otimes P_k)(I_{d^2} + K_{d,d}) = \sigma_1 \eta_{(j)}\eta_{(j)} (I_{d^2} + K_{d,d})(P_j \otimes P_k)(I_{d^2} + K_{d,d}).$$

In addition, $(I_{d^2} + K_{d,d})(P_j \otimes P_k)(I_{d^2} + K_{d,d}) = (I_{d^2} + K_{d,d})(P_j \otimes P_k + P_j \otimes P_k)$ and, for distinct j, k and l, $(P_j \otimes P_k) V_M (P_j \otimes P_l) = 0$. Using the spectral representation $(M - \eta_{(j)} I_d)^+ = \sum_{k\neq j} (\eta_{(k)} - \eta_{(j)})^{-1} P_k$ then yields expression (B3).

For Tyler's scatter estimator, the eigenprojections for $\hat{\mathcal{T}}_n$ and for $\hat{\mathcal{T}}_{o,n} = \hat{c}\hat{\mathcal{T}}_n$, where $\hat{c} = d/\mathrm{tr}(\Gamma^{-1}\hat{\mathcal{T}}_n)$, are the same, namely $\hat{P}_{\mathcal{T},j}$. Under any elliptical distribution, the asymptotic variance for $\hat{\mathcal{T}}_{o,n}$ is shown in Tyler (1987a) to have the form (B2) with $\sigma_1 = (d+2)/d$, $\sigma_2 = 2/d$ and $M = \Gamma$. Equation (B3) thus applies with $\eta_{(j)} = \lambda_{(j)}$, which gives the expression (8) for $V_{\mathcal{T},j}(\Gamma)$.

The result (8) for $V_{\mathcal{S},j}(\Gamma)$ is more complicated to establish, since $\hat{\mathcal{S}}_n$ is not affine equivariant. Using the spectral value decomposition $\Gamma = Q\Lambda Q^{\mathrm{T}}$, it follows that $V_{\mathcal{S},j}(\Gamma) = (Q \otimes Q) V_{\mathcal{S},j}(\Lambda)(Q \otimes Q)^{\mathrm{T}}$, and so it is sufficient to consider the special case where $\Gamma = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$. The asymptotic variance of $\hat{\mathcal{S}}_n$ is then given by $V_{\mathcal{S}} = \mathrm{var}\{\mathrm{vec}(\theta\theta^{\mathrm{T}})\} = E(\theta\theta^{\mathrm{T}} \otimes \theta\theta^{\mathrm{T}}) - \mathrm{vec}(\Delta)\mathrm{vec}(\Delta)^{\mathrm{T}}$, where $\theta \sim \mathrm{ACG}_d(\Lambda)$. We proceed by first finding an expression for $E(\theta\theta^{\mathrm{T}} \otimes \theta\theta^{\mathrm{T}})$.

Let $e_1, \ldots e_d$ represent the Euclidean basis elements in $\mathbb{R}^d$; then

$$E(\theta\theta^{\mathrm{T}} \otimes \theta\theta^{\mathrm{T}}) = \sum_{p=1}^{d}\sum_{q=1}^{d}\sum_{r=1}^{d}\sum_{s=1}^{d} a_{pqrs}(e_p e_q^{\mathrm{T}} \otimes e_r e_s^{\mathrm{T}}),$$

where $a_{pqrs} = E(\theta_p\theta_q\theta_r\theta_s)$. Since θ is symmetrically distributed in each coordinate, it follows that $a_{pqrs} = 0$ unless $p = q = r = s$, $p = q$ and $r = s$, $p = r$ and $q = s$, or $p = s$ and $q = r$. Let $\gamma_{pq} = E(\theta_p^2\theta_q^2)$; then, since $(e_p e_q^{\mathrm{T}} \otimes e_q e_p^{\mathrm{T}}) = K_{d,d}(e_p e_p^{\mathrm{T}} \otimes e_q e_q^{\mathrm{T}})$, the term $E(\theta\theta^{\mathrm{T}} \otimes \theta\theta^{\mathrm{T}})$ can be expressed as

$$(I_{d^2} + K_{d,d})\left\{\sum_{p=1}^{d}\sum_{q=1}^{d}\gamma_{pq}(e_p e_p^{\mathrm{T}} \otimes e_q e_q^{\mathrm{T}}) - \sum_{p=1}^{d}\gamma_{pp}(e_p e_p^{\mathrm{T}} \otimes e_p e_p^{\mathrm{T}})\right\} + \sum_{p=1}^{d}\sum_{q=1}^{d}\gamma_{pq}(e_p e_q^{\mathrm{T}} \otimes e_p e_q^{\mathrm{T}}).$$

Equation (B1), with $M = \Xi = \Delta = \mathrm{diag}(\phi_1, \ldots \phi_d)$, $P_j = E_j = \sum_{i=m_j+1}^{m_j+d_j} e_i e_i^{\mathrm{T}}$ and $\eta_{(j)} = \phi_{(j)}$, can now be applied to $\hat{P}_{\mathcal{S},j}$. We have $\{E_j \otimes (\Delta - \phi_{(j)})^{+}\}\mathrm{vec}(\Delta) = \mathrm{vec}\{(\Delta - \phi_{(j)})^{+}\Delta E_j\} = 0$ and, for $p \neq q$, $\{E_j \otimes (\Delta - \phi_{(j)} I_d)^{+}\}(e_p e_q^{\mathrm{T}} \otimes e_p e_q^{\mathrm{T}})\{E_j \otimes (\Delta - \phi_{(j)} I_d)^{+}\} = 0$. Furthermore, using the fourth property in (A1), it follows that $\{E_j \otimes (\Delta - \phi_{(j)} I_d)^{+}\}K_{d,d}(e_p e_p^{\mathrm{T}} \otimes e_q e_q^{\mathrm{T}})\{E_j \otimes (\Delta - \phi_{(j)} I_d)^{+}\} = 0$. Hence, $V_M = V_{\mathcal{S}}$ in (B1) can be replaced by $\sum_{p=1}^{d}\sum_{q=1}^{d}\gamma_{pq}(e_p e_p^{\mathrm{T}} \otimes e_q e_q^{\mathrm{T}})$.

Next, observe that

$$\{E_j \otimes (\Delta - \phi_{(j)} I_d)^{+}\}(e_p e_p^{\mathrm{T}} \otimes e_q e_q^{\mathrm{T}})\{E_j \otimes (\Delta - \phi_{(j)} I_d)^{+}\} = (\phi_q - \phi_{(j)})^{-2}(e_p e_p^{\mathrm{T}} \otimes e_q e_q^{\mathrm{T}})$$

for $p \in \{m_j + 1, \ldots, m_j + d_j\}$ and $q \notin \{m_j + 1, \ldots, m_j + d_j\}$, and is 0 otherwise. Also, the value of γ_{pq} is the same for all $p \in \{m_j + 1, \ldots, m_j + d_j\}$ and $q \in \{m_k + 1, \ldots, m_k + d_k\}$, say $\psi_{(j,k)}$. Hence

$$V_{\mathcal{S},j}(\Lambda) = (I_{d^2} + K_{d,d})\left\{\sum_{k=1,k\neq j}^{m}\frac{\psi_{(j,k)}}{(\phi_q - \phi_{(j)})^2}(E_j \otimes E_k)\right\}(I_{d^2} + K_{d,d}).$$

Using the identity $(I_{d^2} + K_{d,d})(E_j \otimes E_k)(I_{d^2} + K_{d,d}) = (I_{d^2} + K_{d,d})(E_j \otimes E_k + E_k \otimes E_j)$ then gives the expression for $V_{\mathcal{S},j}(\Gamma)$ in (8) for the special case of $\Gamma = \Lambda$.

Expressions for $\alpha_{\mathcal{T},j,k}$ and $\alpha_{\mathcal{S},j,k}$ in (8) are seen to be

$$\alpha_{\mathcal{T},j,k} = \frac{d+2}{d}\frac{2\lambda_{(j)}\lambda_{(k)}}{(\lambda_{(j)} - \lambda_{(k)})^2}, \quad \alpha_{\mathcal{S},j,k} = \frac{2\psi_{(j,k)}}{(\phi_{(j)} - \phi_{(k)})^2}. \tag{B4}$$

Expressions for the terms $\phi_{(j)}$ and $\psi_{(j,k)}$ can be obtained by noting that $\theta_p^2 \sim \lambda_p \chi_{1,p}^2 / \sum_{q=1}^{d}\lambda_q \chi_{1,q}^2$, where the $\chi_{1,i}^2$ $(i = 1, \ldots, d)$ have independent chi-squared distributions on one degree of freedom. Also, since $\phi_{(j)} = d_j^{-1}\sum_{p=m_j+1}^{m_j+d_j}\phi_p$ and $\psi_{(j,k)} = (d_j d_k)^{-1}\sum_{p=m_j+1}^{m_j+d_j}\sum_{q=m_k+1}^{m_k+d_k}\gamma_{pg}$, it follows that

$$\psi_{(j,k)} = \frac{1}{d_j d_k}E\left\{\frac{\lambda_{(j)}\lambda_{(k)}\ \chi_{(j)}^2\chi_{(k)}^2}{\left(\sum_{r=1}^{m}\lambda_{(r)}\ \chi_{(r)}^2\right)^2}\right\}, \qquad \phi_{(j)} = \frac{1}{d_j}E\left(\frac{\lambda_{(j)}\ \chi_{(j)}^2}{\sum_{r=1}^{m}\lambda_{(r)}\ \chi_{(r)}^2}\right), \tag{B5}$$

where the $\chi_{(r)}^2$ $(r = 1, \ldots, m)$ have mutually independent chi-squared distributions on d_r degrees of freedom. The form of the eigenvalue $\phi_{(j)}$ given above is equivalent to equation (5).

C. THE GAUSS HYPERGEOMETRIC FUNCTIONS

Proof of statement (9)

For $\mathrm{Re}(c) > \mathrm{Re}(b) > 0$, the Gauss hypergeometric functions admit the integral representation ${}_2F_1(a, b; c; \kappa) = B^{-1}(b, c - b)\int_0^1 x^{b-1}(1 - x)^{c-b-1}(1 - \kappa x)^{-a}\,\mathrm{d}x$, where $B(a, b)$ is the beta function; see, e.g., (15·2·1) in Abramowitz & Stegun (1992). Let $\kappa = 1 - \rho^2$ and, using the notation established in (B5), let $U = \chi_{(2)}^2/(\chi_{(1)}^2 + \chi_{(2)}^2) \sim \mathrm{Be}(d_2/2, d_1/2)$. It then follows from the integral representation that

$$E\{U^r(1 - U)^s(1 - \kappa U)^{-t}\} = \frac{B\{(d_2 + 2r)/2, (d_1 + 2s)/2\}}{B\{(d - d_1)/2, d_1/2\}}\ {}_2F_1\left(t, \frac{d - d_1 + 2r}{2}; \frac{d + 2s + 2r}{2}; \kappa\right).$$

This gives $\psi_{(1,2)} = \rho^2 E\{U(1 - U)(1 - \kappa U)^{-2}\}/(d_1 d_2) = \rho^2{}_2F_1\{2, (d_2 + 2)/2; (d + 4)/2; 1 - \rho^2\}/\{d(d + 2)\}$, with $\psi_{(2,1)} = \psi_{(1,2)}$. In addition, $\phi_{(1)} = E\{(1 - U)(1 - \kappa U)^{-1}\}/d_1 = {}_2F_1(1, d_2/2; (d + 2)/2; 1 - \rho^2)/d$ and $\phi_{(2)} = \rho^2 E\{U(1 - \kappa U)^{-1}\}/d_2 = \rho^2{}_2F_1\{1, (d_2 + 2)/2; (d + 2)/2; 1 - \rho^2\}/d$.

To obtain the form of the asymptotic relative efficiency given by (9), insert the expressions for $\psi_{(1,2)}, \phi_{(1)}$ and $\phi_{(1)}$ into the formula for $\alpha_{\mathcal{S}}(\rho) = \alpha_{m\mathcal{S},1,2}$ in (B4). Also, using (B4), it follows that $\alpha_{\mathcal{T}}(\rho) = \alpha_{\mathcal{T},1,2} = 2(d+2)\rho^2/\{d(1-\rho^2)^2\}$. The ratio $\alpha_{\mathcal{T}}(\rho)/\alpha_{\mathcal{S}}(\rho)$ can then be simplified using the identity (15.2.20) in Abramowitz & Stegun (1992):

$$\frac{1-\kappa}{\kappa}\,{}_2F_1(a,b;c;\kappa) - \frac{1}{\kappa}\,{}_2F_1(a-1,b;c;\kappa) + \frac{c-b}{c}\,{}_2F_1(a,b;c+1;\kappa) = 0,$$

with $a = (d_2+2)/2$, $b = 1$ and $c = (d+2)/2$.

Proof of statement (10)

For the dimension $d = 2$ case, $\phi_{(1)} = 1/(1+\rho)$, $\phi_{(2)} = \rho/(1+\rho)$ and $\psi_{(1,2)} = \rho/\{2(1+\rho)^2\}$. The form for $\phi_{(2)}$ follows from the identity ${}_2F_1(a, 1/2+a; 2a; \kappa) = 2^{2a-1}\{1+(1-\kappa)^{1/2}\}^{1-2a}$ with $a = 1$; see (15.1.14) in Abramowitz & Stegun (1992). The form for $\phi_{(1)}$ then follows since $\phi_{(1)} + \phi_{(2)} = 1$. The form for $\psi_{(1,2)}$ follows from the identity $(1-\kappa)^{1/2}\,{}_2F_1(1+a, 1/2+a; 1+2a; \kappa) = 2^{2a}\{1+(1-\kappa)^{1/2}\}^{-2a}$ with $a = 1$; see (15.1.13) in Abramowitz & Stegun (1992). This then gives $\alpha_{\mathcal{S}}(\rho) = \rho/(1-\rho)^2$ and $\alpha_{\mathcal{T}}(\rho) = 4\rho^2/(1-\rho^2)^2$, which yields the form for the asymptotic relative efficiency (10).

Proof of statement (11)

Consider the identity ${}_2F_1(a, b; c; 1) = \{\Gamma(c)\Gamma(c-a-b)\}/\{\Gamma(c-a)\Gamma(c-b)\}$ which holds when $c > a+b$; see (15.1.20) in Abramowitz & Stegun (1992). Here $\Gamma(a)$ refers to the usual gamma function. Taking the limit as $\rho \to 0$ in (9) and applying this identity gives (11) for the case where $d_1 > 2$. The cases of $d_1 = 1, 2$ require special treatment, since for these cases $c - a - b = (d_1-2)/2 \leqslant 0$. Here the identity (15.3.3) in Abramowitz & Stegun (1992) is useful; this states that ${}_2F_1(a; b; c; \kappa) = (1-\kappa)^{c-a-b}\,{}_2F_1(c-a; c-b; c; \kappa)$. So the denominator in (9) goes to infinity as ρ goes to zero. Hence (11) holds.

D. Limiting value of the finite-sample relative efficiency

Write $P_1 = Q_1 Q_1^{\mathrm{T}}$ with $Q_1^{\mathrm{T}} Q_1 = I_{d_1}$. The ordered eigenvalues of $Q_1^{\mathrm{T}} \hat{P}_{\mathcal{T},1} Q_1$ then correspond to $\cos^2(\hat{\tau}_{1,n}), \ldots, \cos^2(\hat{\tau}_{d_1,n})$. Recall that $n^{1/2}(\hat{P}_{\mathcal{T},1} - P_1) \to \{\alpha_{\mathcal{T}}(\rho)\}^{1/2} Z$ in distribution, with $\mathrm{vec}(Z)$ being $N_{d^2}(0, M_{1,2})$. This implies that $n(I - Q_1^{\mathrm{T}} \hat{P}_{\mathcal{T},1} Q_1) = n Q_1^{\mathrm{T}} (\hat{P}_{\mathcal{T},1} - P_1)^2 Q_1 \to \alpha_{\mathcal{T}}(\rho) W$ in distribution, where $W = Q_1^{\mathrm{T}} Z^2 Q_1 = Q_1^{\mathrm{T}} Z^{\mathrm{T}} Z Q_1^{\mathrm{T}}$.

Since ordered eigenvalues are continuous functions of their symmetric matrix arguments, see, e.g., Kato (1966), it follows that for $i = 1, \ldots, d_1$, $n\{1 - \cos^2(\hat{\tau}_{i,n})\} = n\sin^2(\hat{\tau}_{i,n})$ converge jointly in distribution to $\alpha_{\mathcal{T}}(\rho)\sigma_i(W)$, with $\sigma_1(W) \geqslant \cdots \geqslant \sigma_{d_1}(W) \geqslant 0$ being the ordered eigenvalues of W. Using the delta method, one obtains that $n\{\sin^2(\hat{\tau}_{i,n}) - \hat{\tau}_{i,n}^2\} \to 0$ in probability, and so $nT_n \to \alpha_{\mathcal{T}}(\rho)\,\mathrm{tr}(W)$ in distribution. An analogous argument gives $n\Omega_n \to \alpha_{\mathcal{S}}(\rho)\,\mathrm{tr}(W)$ in distribution. Since $\mathrm{tr}(W)$ has a continuous distribution, the medians of nT_n and $n\Omega_n$ converge, respectively, to $\alpha_{\mathcal{T}}(\rho)\,\mathrm{med}\{\mathrm{tr}(W)\}$ and $\alpha_{\mathcal{S}}(\rho)\,\mathrm{med}\{\mathrm{tr}(W)\}$ in probability. Hence (13) converges to $\alpha_{\mathcal{T}}(\rho)/\alpha_{\mathcal{S}}(\rho)$ in probability when using the median. This limit would also hold when using the expected value if it could be established that both $\{nT_n \mid n = 1, \ldots, \infty\}$ and $\{n\Omega_n \mid n = 1, \ldots, \infty\}$ are uniformly integrable. We leave this problem for possible future research.

Acknowledgement

The research of both authors was supported in part by the U.S. National Science Foundation.

References

Abramowitz, M. & Stegun, I. A., eds. (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications. Reprint of the 1972 edition.

Afriat, S. N. (1957). Orthogonal and oblique projectors and the characteristics of pairs of vector spaces. *Proc. Camb. Phil. Soc.* **53**, 800–16.

Bilodeau, M. & Brenner, D. (1999). *Theory of Multivariate Statistics*. New York: Springer.
Björck, A. & Golub, G. (1973). Sketch of a proof that an odd perfect number relatively prime to 3 has at least eleven prime factors. *Math. Comp.* **27**, 579–694.
Boente, G. & Fraiman, R. (1999). Comment on 'Robust principal component analysis for functional data'. *Test* **8**, 28–35.
Davies, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15**, 1269–92.
Davies, P. L. & Gather, U. (2005). Breakdown and groups. *Ann. Statist.* **15**, 977–1035.
Donoho, D. L. & Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20**, 1803–27.
Dümbgen, L. & Tyler, D. E. (2005). On the breakdown properties of some multivariate M-functionals. *Scand. J. Statist.* **32**, 247–64.
Dürre, A., Vogel, D. & Tyler, D. E. (2014). The spatial sign covariance matrix with unknown location. *J. Mult. Anal.* **130**, 107–17.
Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley.
Huber, P. J. (1977). *Robust Statistical Procedures*. Philadelphia: Society for Industrial and Applied Mathematics.
Kato, T. (1966). *Perturbation Theory for Linear Operators*. Berlin: Springer.
Kent, J. T. & Tyler, D. E. (1996). Constrained M-estimation for multivariate location and scatter. *Ann. Statist.* **24**, 1346–70.
Lehmann, E. & Casella, G. (1998). *Theory of Point Estimation*. New York: Springer, 2nd ed.
Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. & Cohen, K. L. (1999). Robust principal component analysis for functional data. *Test* **8**, 1–73. With discussion and a rejoinder by the authors.
Lopuhaä, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann. Statist.* **17**, 1662–83.
Lopuhaä, H. P. (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *Ann. Statist.* **27**, 1638–65.
Magyar, A. F. & Tyler, D. E. (2012). The asymptotic efficiency of the spatial median for elliptically symmetric distributions. *Sankhyā* B **73**, 165–92.
Marden, J. I. (1999). Some robust estimates of principal components. *Statist. Prob. Lett.* **43**, 349–59.
Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4**, 51–67.
Maronna, R. A., Martin, R. D. & Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Chichester: John Wiley & Sons.
Maronna, R. A., Stahel, W. A. & Yohai, V. J. (1992). Bias-robust estimators of multivariate scatter based on projections. *J. Mult. Anal.* **42**, 141–61.
Miao, J. M. & Ben-Israel, A. (1992). On principal angles between subspaces in $\mathbb{R}^n$. *Lin. Algeb. Applic.* **171**, 81–98.
Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications, Volume B (Proc. 4th Pannonian Symp. Math. Statist., Bad Tatzmannsdorf, 1983)*. Dordrecht: Reidel, pp. 283–97.
Taskinen, S., Koch, I. & Oja, H. (2012). Robustifying principal component analysis with spatial sign vectors. *Statist. Prob. Lett.* **82**, 765–74.
Tatsuoka, K. S. & Tyler, D. E. (2000). On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *Ann. Statist.* **28**, 1219–43.
Tyler, D. E. (1981). Asymptotic inference for eigenvectors. *Ann. Statist.* **9**, 725–36.
Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika* **70**, 411–20.
Tyler, D. E. (1987a). A distribution-free M-estimator of multivariate scatter. *Ann. Statist.* **15**, 234–51.
Tyler, D. E. (1987b). Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika* **74**, 579–89.
Tyler, D. E. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.* **22**, 1024–44.
Tyler, D. E. (2002). High breakdown point multivariate M-estimation. *Estadística* **54**, 213–47.
Tyler, D. E. (2010). A note on multivariate location and scatter statistics for sparse data sets. *Statist. Prob. Lett.* **80**, 1409–13.
Visuri, S., Koivunen, V. & Oja, H. (2000). Sign and rank covariance matrices. *J. Statist. Plan. Infer.* **91**, 557–75.
Watson, G. S. (1983). *Statistics on Spheres*. New York: Wiley.

[*Received August* 2013. *Revised April* 2014]