American Journal of Theoretical and Applied Statistics
2022; 11(4): 122-139
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.20221104.13
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Principal Component Analysis of Standard and Spherical Covariances from the Population and Random Samples to Real and Simulated Data

Inge Koch[1], **Lyron Winderbaum**[2], **Kanta Naito**[3]

[1]Department of Mathematics and Statistics, The University of Western Australia, Crawley, Australia
[2]Department of Mathematical Sciences, The University of South Australia, Adelaide, Australia
[3]Department of Mathematics, Chiba University, Chiba, Japan

Email address:

inge.koch@uwa.edu.au (Inge Koch), Lyron.Winderbaum@unisa.edu.au (Lyron Winderbaum), naito@math.s.chiba-u.ac.jp (Kanta Naito)
*Corresponding author

To cite this article:

Inge Koch, Lyron Winderbaum, Kanta Naito. Principal Component Analysis of Standard and Spherical Covariances from the Population and Random Samples to Real and Simulated Data. *American Journal of Theoretical and Applied Statistics*. Vol. 11, No. 4, 2022, pp. 122-139. doi: 10.11648/j.ajtas.20221104.13

Received: August 12, 2022; **Accepted:** September 5, 2022; **Published:** September 26, 2022

Abstract: Principal component analysis (PCA) is the tool of choice for summarising multivariate and high-dimensional data as features in a lower-dimensional space. PCA works well for Gaussian data, but may not do so well for high-dimensional, skewed or heavy-tailed data or data with outliers as encountered in practice. The availability of complex data has enhanced these shortcomings and increased the demand for PC approaches that perform well for such data. The purpose of this paper is to critically appraise a class of interpretable PC candidates which can respond to this demand and to compare their performance to that of standard PCA. Among the large variety of nonlinear PCA, we concentrate on the subclass that is based on spherical covariance matrices. This subclass includes the spatial sign, spatial rank, and Kendall's τ covariance matrix. We focus on three key aspects: population concepts and their properties; sample-based estimators; and actual practice based on the analysis of real and simulated data. At the population level we consider relationships between the standard covariance matrix and spherical covariance matrices. For the random sample we consider natural estimators of the population eigenvectors, look at appropriate distributional models, highlight relationships between different estimators and relate properties of estimators and their population analogues. We complement the theory we present with new analyses of multivariate and high-dimensional real data as well as simulated data from diverse distributions which elucidates behaviour patterns of spherical PCA for elliptic and non-elliptic distributions. The latter are not captured in the theoretical framework, and their inclusion therefore offers fresh insight into the performance of spherical PCA. The combination of the theory and the new analysis evidence that PCA of rank-based covariances severely outperforms that based on the potentially unstable spatial sign covariance matrix. Further, the overall good performance of rank-based PCA and its superior properties for data for which the sample covariance matrix has been known to perform poorly make rank-based PCA not only a desirable addition to standard PCA, but render it a serious competitor for dimension reduction and feature selection while retaining features valued in PCA.

Keywords: Multivariate Ranks, Multivariate Spatial Signs, Nonlinear Covariance Matrices, Performance of Nonlinear PCA, Spherical PCA

1. Introduction

Statistics deals with the population, the random sample and observed data, the interplay and interaction between them and their synthesis. At the population level, we consider models and distributional assumptions. For the random sample we construct estimators of population quantities and examine their performance. In the context of principal component analyiss (PCA), we investigate the performance of dimension reduction methods and their suitability for the analysis of complex high-

dimensional data, including data with fewer observations than variables. Distributional assumptions of the observed data may not agree with the theoretical models and this leads to new challenges that do not typically fit into the classical Gaussian framework of available theory. In this paper the emphasis is on finding lower-dimensional structure in multivariate and high-dimensional data and we explore how the population models and results compare with the reality of data from a variety of distributions which go beyond those of the theoretical models.

We present pertinent results without delving into proofs; the interested reader can find the proofs in the references. We go beyond standard PCA and the sample covariance matrix; we include a subset of nonlinear PC approaches based on spherical covariance matrices, as proposed in robust statistics a few decades ago. These PC approaches stand out among the nonlinear approaches by being intuitive, easy to construct and interpretable.

Principal component analysis has undoubtedly become one of the most commonly used methods for dimension reduction and feature selection and often represents a first step in a more in-depth analysis. Its success rests on a number of different properties including maximisation of the variance in the PCA directions—the eigenvectors of the covariance matrix—which are also called factor loadings. Variance calculations rely on the Euclidean norm which fits naturally with the Gaussian likelihood. This is one of the reasons why PCA is particularly successful at and suitable for finding high variance directions in Gaussian data.

With increasing dimension and complexity of data, dimension reduction and feature selection have gained more importance. New and exciting developments in PCA began during the last few decades of the 20th century in response to demands from these new data: high-dimensional and high-dimension low sample size (HDLSS) data led to new models and theory including sparse PCA and their sparse eigenvectors. In robust statistics the existence of outliers and skewed or heavy-tailed data and the recognition of the sensitivity of the sample covariance matrix to such observations required *PCA-like* approaches that can handle such data more appropriately. To address some of these issues, J. I. Marden [1] proposed two covariance-related concepts, based on multivariate signs and multivariate ranks which are now regarded as part of the collection of *robust covariance matrices* and *robust PCA*.

At first glance, Marden's results [1] for the population may seem surprising, but they are taken almost as *folklore* by researchers in the field. To give the reader some insight into the relationships between the different covariance matrices, we present pertinent results. As we will see, different patterns emerge for the eigenvectors and eigenvalues of the various covariance matrices. We explore these further in simulations.

In addition to the population sign and rank covariance matrices, which we define in (3) and (4), J. I. Marden [1] defined analogous sample quantities. We regard these sample matrices and their eigenvectors as estimators of the relevant population quantities; we describe consistency properties and highlight relationships between the sample covariance matrices. The properties and relationships reveal that the population-based intuition may not be reflected in the behaviour of the sample quantities. In particular, the spatial sign sample covariance matrix, which seems to be natural for observations from elliptic distributions, does not always possess the 'desirable' properties of its population counterpart. From the random sample we progress to data and the challenges data present as they deviate from the population assumptions or grow in complexity and dimension. Such challenges include the treatment of HDLSS data which abound and may pose computational challenges. We will consider high-dimensional data with outliers and simulated data from a range of distributions in the context of PCA based on different covariance matrices. Our simulations are informed by the theoretical results and the analyses of real data. Comparing different estimators and different distributions leads to new insight which enables us to make statements regarding pros and cons of different PC-based estimators.

This paper is organised as follows. Section 2 provides definitions and properties of the population quantities. Section 3 outlines relationships between sample quantities and their population equivalents, Section 4 looks at real data and highlights differences betweeen the various projections for HDLSS data and for data with a moderate number of variables. Section 5 focusses on simulated data; we consider a variety of models and families of elliptic and non-elliptic distributions which occur in practice. We vary the sample sizes and dimensions of the generated data and we relate the simulation results back to the theoretical models and theorems presented in Sections 2 and 3. A summary and some recommendations are given in Section 6. Our Appendix provides further detail of the data analyses presented in Sections 4 and 5 in the form of additional tables and figures.

We conclude this section with notation used in this paper.

Notation. $X \sim F$ or $X \sim F(\mu, \Sigma)$ d-dimensional random vector from a distribution F with mean μ and covariance matrix Σ;

$\mathbf{X} = [X_1, \ldots, X_n]$ random sample of size n from F with $X_i \in \mathbb{R}^d$;

$\Sigma_{ss} = \Gamma_{ss}\Lambda_{ss}\Gamma_{ss}^T$ population covariance matrix and spectral decomposition;

$(\lambda_{ss,j}, \eta_{ss.j})$ jth eigenvalue/eigenvector pair pertaining to Σ_{ss};

$S_{ss} = \widehat{\Gamma}_{ss}\widehat{\Lambda}_{ss}\widehat{\Gamma}_{ss}^T$ sample covariance matrix and spectral decomposition;

$(\widehat{\lambda}_{ss.j}, \widehat{\eta}_{ss,j})$ jth eigenvalue/eigenvector pair pertaining to S_{ss};

$S^{(d-1)}$ d-dimensional unit sphere;

$\mathcal{S}_d$ set of spherically symmetric random vectors in $\mathbb{R}^d$;

$\mathcal{E}_d(\mu, \Sigma)$ set of elliptically symmetric random vectors in $\mathbb{R}^d$ with location parameter μ and and covariance matrix Σ;

$\|X\|$ Euclidean norm of the vector X;

$\|A\|_F = \left[\mathrm{tr}(A^T A)\right]^{1/2}$ Frobenius norm of the matrix A;

$\|A\|_{sup} = \sup_{\|v\|=1} |v^T A v|$ the *sup*-norm or *spectral* norm of the matrix A;

$X \stackrel{\mathcal{D}}{=} Y$ equality *in distribution* of the random vectors X and Y.

The subscript $_{ss}$ in the notation above represents a place holder for the different covariance matrices.

For notational convenience we will refer to the different matrices as covariance matrices and only make distinctions if necessary. We refer to the canonical or natural covariance matrix as the *standard* covariance matrix, as this is the 'standard' we compare the others to. We use the subscript-free notation Σ and S for the population and sample standard covariance matrices. Here S contains the factor $(n-1)^{-1}$ where n is the sample size.

2. Population Covariance Matrices for the Direction Vector

2.1. Definitions for the Population

Let $X \sim F(\mu, \Sigma)$ be a d-dimensional random vector from a distribution F with expected value μ and covariance matrix Σ. As in J. I. Marden [1], the *multivariate* or *spatial sign* or the *direction (vector)* $\operatorname{dir}(X)$ of X is

$$\operatorname{dir}(X) = \begin{cases} X/\|X\| & \text{if } X \neq 0, \\ 0 & \text{if } X = 0, \end{cases} \tag{1}$$

and the *spatial rank* or *centred rank* of $Y \in \mathbb{R}^d$ with respect to F is

$$R_{\text{sp}}(Y, F) = \mathbb{E}\left[\operatorname{dir}(Y - X)\right] \quad \text{for } X \sim F. \tag{2}$$

The definition $\operatorname{dir}(X)$ is a multivariate generalisation of the *sgn*-function; the direction of each vector remains unchanged, since $X = \|X\| \operatorname{dir}(X)$ is a multiple of its direction. We may think of $\operatorname{dir}(X)$ as the random vector mapped onto the d-dimensional unit sphere. The spatial rank of Y is the expected value of $\operatorname{dir}(Y - X)$ with respect to F; it represents the expected deviation on the d-dimensional unit sphere of Y from a given distribution.

Centring random vectors prior to applying the transform of (1) is common. In the robust statistics literature a '*location parameter*' can mean different quantities including a centring parameter. Because of this ambiguity, in this paper the term '*centring parameter*' refers specifically to centring of random vectors. Typically the centring parameter refers to the expected value or the sample mean, but the choice is not restricted to these. For a centring parameter θ and $X \neq \theta$, we call $\operatorname{dir}(X-\theta) = (X-\theta)/\|X-\theta\|$ the *centred direction (vector)*.

We begin with the two covariance matrices which J. I. Marden [1] defines for direction vectors and spatial ranks.

The *(spatial) sign covariance matrix* Σ_{sptl} of X with respect to a centring parameter $\tilde{X} \in \mathbb{R}^d$, and the *(spatial) rank covariance matrix* Σ_{rank} of X are

$$\Sigma_{\text{sptl}}(\tilde{X}) = \mathbb{E}\left[\operatorname{dir}(X - \tilde{X}) \operatorname{dir}(X - \tilde{X})^T\right] \quad \text{and} \tag{3}$$

$$\Sigma_{\text{rank}} = \mathbb{E}\left[R_{\text{sp}}(X, F) R_{\text{sp}}(X, F)^T\right]. \tag{4}$$

If $\tilde{X}$ in (3) is the expected value of X, we write Σ_{sptl} instead of $\Sigma_{\text{sptl}}(\mu)$ and note that Σ_{sptl} is a natural candidate for the covariance matrix of centred direction vectors.

In the robust statistics literature, two further covariance matrices are considered: For $X_i \sim F$ and $i = 1, 2, 3$ *Kendall's τ covariance matrix* Σ_τ and *Spearman's rank covariance matrix* Σ_{spear} are

$$\Sigma_\tau = \mathbb{E}\left[\operatorname{dir}(X_1 - X_2) \operatorname{dir}(X_1 - X_2)^T\right] \quad \text{and} \tag{5}$$

$$\Sigma_{\text{spear}} = \mathbb{E}\left[\operatorname{dir}(X_1 - X_2) \operatorname{dir}(X_1 - X_3)^T\right]. \tag{6}$$

S. Taskinen, I. Koch, and H. Oja [2] called Kendall's τ covariance matrix the *symmetrised spatial sign covariance matrix*. They regard Kendall's τ covariance matrix as a substitute for the spatial covariance matrix when dealing with data. We will return to this point throughout this paper.

Remark. The construction of the covariance matrices in (3)–(6) include a nonlinear transformation, the mapping of the original vectors to direction vectors. For this reason, we may like to think of these covariance matrices and the resulting PCA as nonlinear covariance matrices and nonlinear PCA. In the literature nonlinear PCA typically includes kernel-based approaches, but we shall not be concerned with such approaches here. Some authors refer to PCA based on the covariance matrices in (3)–(6) as *spherical* PCA.

The definitions (3)–(6) given here agree with those commonly used in the literature. Some authors, including S.Visuri, V. Koivunen, and H. Oja [3] and A. Dürre, D. Vogel, and R. Fried [4] restrict their definition of the covariance matrix (3) to the spatial median. Following D. Gervini [5] the *spatial median* μ_{sptl} of $X \sim F$ is

$$\mu_{\text{sptl}}(X) = \operatorname*{argmin}_Y \mathbb{E}\left(\|X - Y\| - \|X\|\right). \tag{7}$$

The spatial median is also called the *ℓ_2-median*. From D. Gervini [5], the spatial median satisfies $\mathbb{E}\left[\operatorname{dir}(X - \mu_{\text{sptl}})\right] = 0$. Because of this property, the spatial median may be regarded as a 'natural' centring parameter for direction vectors. Depending on the underlying distribution, the spatial median may not be given by an algebraic expression and, as we shall see, this renders the spatial median less amenable in real data analysis.

2.2. Properties and Relationships of Population Covariance Matrices

We begin with a property that applies to two covariance matrices and which does not depend on the distribution of X.

Theorem 2.1 Let X be a random vector from a distribution F with covariance matrix Σ. The rank covariance matrix Σ_{rank} and Spearman's rank covariance matrix Σ_{spear} satisfy

$$\Sigma_{\text{rank}} = \Sigma_{\text{spear}}. \tag{8}$$

The proof involves a straightforward calculation based on evaluating expectations. Although the population quantities

are the same, we will see in Section 3 that different sample estimators exist for these population quantities—quite likely this is due to the two ideas having different origins.

For X from an elliptic distribution, we summarise pertinent results relating to Σ and the covariance matrices Σ_{sptl} and Σ_{rank} and give references to proofs.

Theorem 2.2 Assume the random vector X has an elliptic distribution $\mathcal{E}_d(\mu, \Sigma)$, and write $\Sigma = \Gamma\Lambda\Gamma^T$ for the spectral decomposition of Σ. Let Σ_{sptl} be the sign covariance matrix based on the centring parameter μ, and let Σ_{rank} and Σ_τ be the rank and Kendall's τ covariance matrices of X respectively. The following hold.

1. The expected value μ equals the spatial median μ_{sptl}.
2. Σ_{sptl} and Σ_τ satisfy

$$\Sigma_\tau = \Sigma_{\text{sptl}}. \tag{9}$$

3. The eigenvectors of Σ, Σ_{sptl} and Σ_{rank} agree and give rise to the spectral decompositions

$$\Sigma_{\text{sptl}} = \Gamma\Lambda_{\text{sptl}}\Gamma^T \quad \text{and} \quad \Sigma_{\text{rank}} = \Gamma\Lambda_{\text{rank}}\Gamma^T. \tag{10}$$

Part 2 of the theorem tells us that under the elliptic model Σ_{sptl} and Σ_τ agree. Part 3 of the theorem tells us that the eigenvectors of Σ, Σ_{sptl} and Σ_{rank} are the same under the elliptic model. This equality, however, does not extend to the eigenvalues of the three matrices, as the subscript notation in part 3 shows. We will present some results concerning the eigenvalues in the next theorem.

In Section 3 we will study estimators of the population eigenvectors and will come across a number of different estimators for these population quantities. Their performance will inform which one to choose for different distributions in principal component-based dimension reduction, subspace or feature selection. We will return to this choice at the end of the paper.

Part 1 of Theorem 2.2 is taken from A. Dürre, D. Vogel and R. Fried [4] and uses the symmetry inherent in elliptic random vectors X which can be written as $X = EY + b$, where E is an orthogonal matrix, Y has a symmetric distribution and b is the centring parameter. Part 1 may not hold for non-elliptic distributions. Part 2 is shown in S. Taskinen, I. Koch, and H. Oja [2]. The proof of part 3 is given in J. I. Marden [1] and relies on the same decomposition $X = EY + b$ as used in part 1. The orthogonal matrix E transforms the spatial rank of X into another spatial rank, and similarly for spatial signs. This leads to the desired result. For more general matrix transforms and non-elliptic distributions this equality may not hold.

Theorems 2.1 and 2.2 provide information about identical pairs of covariance matrices and equality of eigenvectors, but they do not answer the question of the relationship between the eigenvalues of the spherical covariance matrices and Σ. Our next result summarises pertinent results relating to the eigenvalues of Σ_{sptl} and hence also of Σ_τ.

Theorem 2.3 Assume the random vector X has an elliptic distribution $\mathcal{E}_d(\mu, \Sigma)$. Write $\Sigma = \Gamma\Lambda\Gamma^T$ for the spectral decomposition of Σ and put $V = \Gamma^T(X - \mu)$. Let Σ_{sptl} be the sign covariance matrix based on the centring parameter μ, and let Σ_τ be Kendall's τ covariance matrix of X. Denote the common diagonal matrix of eigenvalues of Σ_{sptl} and Σ_τ by Λ_s. The following hold.

1. The matrix Λ_s is given by

$$\Lambda_s = \mathbb{E}\left[\frac{VV^T}{\|V\|^2}\right] = \Gamma^T\mathbb{E}\left[\frac{(X-\mu)(X-\mu)^T}{\|X-\mu\|^2}\right]\Gamma, \quad \text{and} \quad \lambda_{s,j} = \mathbb{E}\left[\frac{\lambda_j Y_j^2}{\sum_{\ell\le d}\lambda_\ell Y_\ell^2}\right], \tag{11}$$

   with $Y = (Y_1, \ldots, Y_d) \in \mathcal{N}(0, I_d)$ and I_d the $d \times d$ identity matrix.
2. Pairs of eigenvalues of Λ_s and Λ satisfy

$$\frac{\lambda_{s,\ell}}{\lambda_{s,j}} \le \frac{\lambda_\ell}{\lambda_j} \quad \text{for } \ell < j \text{ provided } \lambda_j > 0$$

   and, if $\lambda_\ell > \lambda_j$, then the inequality is strict.
3. For $j \le d$ the eigenvalues $\lambda_{s,j}$ and λ_j or Λ_s and Λ are related by

$$\lambda_{s,j} \ge \frac{\lambda_j}{\text{tr}(\Sigma) + 4\|\Sigma\|_F\sqrt{\log d} + 8\|\Sigma\|_{sup}\log d}\left[1 - \frac{\sqrt{3}}{d^2}\right]$$
$$\text{and, provided } \text{tr}(\Sigma) > 4\|\Sigma\|_F\sqrt{\log d},$$
$$\lambda_{s,j} \le \frac{\lambda_j}{\text{tr}(\Sigma) - 4\|\Sigma\|_F\sqrt{\log d}} + d^{-4}. \tag{12}$$

The theorem provides expressions for the eigenvalues of Σ_{sptl} and, by Theorem 2.2, also for Σ_τ. It might be helpful to express to write $V = r\Lambda^{1/2}U$ with $r = \|\Lambda^{-1/2}V\|$, $U = \Lambda^{-1/2}V/\|\Lambda^{-1/2}V\|$, and note that $U \in S^{(d-1)}$, the d-dimensional unit sphere, and that r and U are independent.

From part 2 we glean that the eigenpairs $(\lambda_{s,j}, \eta_j)$ of Σ_{sptl} are given in the same order as those of Σ under the elliptic model. The bounds shown in part 3 may lead to useful information on the relationship between the eigenvalues of Σ and Σ_τ. For example, if the eigenvalues of Σ decrease with

the index j as a^j with $0.98 \le a < 1$ and d sufficiently large, then $\operatorname{tr}(\Sigma) > 4\|\Sigma\|_F \sqrt{\log d}$. Such conditions will be difficult to check in practice unless one substitutes S for Σ.

Part 1 is shown in C. Croux, E. Ollila, and H. Oja [6], S. Taskinen, I. Koch, and H. Oja [2] and F. Han and H. Liu [7]. We have presented the expression given in F. Han and H. Liu [7]. S. Taskinen, I. Koch, and H. Oja [2] state their expression in terms of vectors on the unit sphere. A. Dürre, D. E. Tyler, and D. Vogel [8] show the inequality in part 2, and proofs for part 3 can be found in F. Han and H. Liu [7].

S. Taskinen, I. Koch, and H. Oja [2] propose to exploit the equality of the two population matrices Σ_{sptl} and Σ_τ for the sample and to employ an estimator of Σ_τ as an estimator of Σ_{sptl}. This substitution avoids having to find or calculate (iteratively) the sample spatial median for data. We will compare the 'closeness' of the natural sample-based estimators for Σ_τ and Σ_{sptl} in the simulations of Section 5.

We summaries the results of this section in diagram (13) which shows the conditions under which the relationships hold. In the diagram $\mathcal{E}_d$ refers to the elliptic distribution with mean and covariance matrix (μ, Σ). The use of a common symbol Γ for all matrices in (13) indicates equality of the eigenvectors, and the '$=$' sign next to downwards arrows refers to equality of the quantities above and below the '$=$'-sign.

$$\begin{array}{ccc} & \Sigma = \Gamma\Lambda\Gamma^T & \\ \swarrow^{\mathcal{E}_d} & & \searrow^{\mathcal{E}_d} \\ \Sigma_{\text{sptl}} = \Gamma\Lambda_{\text{sptl}}\Gamma^T & & \Sigma_{\text{rank}} = \Gamma\Lambda_{\text{rank}}\Gamma^T \\ \Big\downarrow{\scriptstyle \mathcal{E}_d} \Big| = & & = \Big| \Big\downarrow{\scriptstyle \text{any } F} \\ \Sigma_\tau = \Gamma\Lambda_{\text{sptl}}\Gamma^T & & \Sigma_{\text{spear}} = \Gamma\Lambda_{\text{rank}}\Gamma^T \end{array} \tag{13}$$

In this section we considered population covariance matrices of the raw random vector. Many PCA users scale their data prior to a PCA. Scaling clearly does not affect the validity of Theorems 2.2–2.2. As robust statistics typically deals with raw, rather than the scaled vectors, we will not consider scaled vectors or data here.

3. Sample Covariance Matrices for Direction Vectors

3.1. Definitions for the Sample

In this section we survey and appraise sample covariance matrices corresponding to the covariance matrices of Section 2. We regard relationships of the sample covariance matrices and their population counterparts and also present relationships between the spherical sample covariance matrices.

The existing definitions are not unique. We aim to present the now accepted definitions, but we will indicate variations that are used.

For the sample, the definition of a direction vector is the same as for the population, however, the definition of the sample spatial rank requires some modification.

Let $\mathbf{X}$ be a sample of n random vectors X_i from the distribution F with sample mean $\overline{X}$ and sample covariance matrix S. Following J. I. Marden [1], the *sample spatial rank* or *sample centred rank* of a non-zero random vector Y with respect to a random sample $\mathbf{X}$ is

$$\text{rank}_{\text{sp}}(Y, \mathbf{X}) = \frac{1}{n^\#} \sum_{i \le n} \text{dir}(Y - X_i), \tag{14}$$

where $n^\#$ refers to the number of non-zero terms. If there is no ambiguity regarding the random sample, we write $\text{rank}_{\text{sp}}(Y)$ for the spatial rank of Y. Equipped with the notion of the sample spatial rank we consider covariance matrices for these random vectors. Consider a random sample $\mathbf{X} = [X_1, \ldots, X_n]$. The sample *(spatial) sign covariance matrix* S_{sptl} of $\mathbf{X}$ with respect to a centring parameter $\tilde{X} \in \mathbb{R}^d$, and the sample *(spatial) rank covariance matrix* S_{rank} of $\mathbf{X}$ are

$$S_{\text{sptl}}(\mathbf{X}, \tilde{X}) = \frac{1}{n} \sum_{i \le n} \text{dir}(X_i - \tilde{X}) \, \text{dir}(X_i - \tilde{X})^T \quad \text{and} \tag{15}$$

$$S_{\text{rank}}(\mathbf{X}) = \frac{1}{n} \sum_{i \le n} \text{rank}_{\text{sp}}(X_i) \, \text{rank}_{\text{sp}}(X_i)^T = \frac{1}{n} \sum_{i \le n} \left[\frac{1}{n-1} \sum_{j \ne i} \text{dir}(X_i - X_j) \right] \left[\frac{1}{n-1} \sum_{j \ne i} \text{dir}(X_i - X_j)^T \right]. \tag{16}$$

For $\mathbf{X}$ the sample version S_τ of *Kendall's* τ and S_{spear} of *Spearman's rank* covariance matrix are

$$S_\tau(\mathbf{X}) = \frac{1}{n(n-1)} \sum_{i,j} \text{dir}(X_i - X_j) \, \text{dir}(X_i - X_j)^T \quad \text{and} \tag{17}$$

$$S_{\text{spear}}(\mathbf{X}) = \frac{1}{n(n-1)(n-2)} \sum_{i,j,k; j \ne k} \text{dir}(X_i - X_j) \, \text{dir}(X_i - X_k)^T. \tag{18}$$

J. I. Marden [1] uses the factor $[(n-1)n^2]^{-1}$ in his definition of the sample rank covariance matrix S_{rank}, while later authors including K. Yu, X. Dang, and Y. Chen [9] mostly use the number of non-zero entries $[(n-1)^2 n]$ in (16).

Some authors, including K. Yu, X. Dang, and Y. Chen [9] and A. Dürre, D. Vogel, and R. Fried [4], restrict their definitions of the sample sign covariance matrix (15) to the mean $\overline{X}$ and the spatial median $\overline{X}_{\text{sptl}}$ as centring parameters. As in the population case, we explicitly include the centring parameter unless it is the sample mean. We define the *sample spatial median* or the ℓ_2*-median* $\overline{X}_{\text{sptl}}$ of $\mathbf{X}$ as in S. Visuri, V. Koivunen, and H. Oja [3] and put

$$\overline{X}_{\text{sptl}} = \underset{\tilde{X}}{\operatorname{argmin}} \sum_{i \le n} \|\tilde{X} - X_i\|, \quad X_i \neq \tilde{X}, \tag{19}$$

see also Section 6.3 of P. J. Huber [10] and N. Locantore "et al." [11]. The latter authors show that $\sum_{i \le n} \operatorname{dir}(X_i - \overline{X}_{\text{sptl}}) = 0$. This equality also follows from D. Gervini [5]. A comparison with (14) shows that $\operatorname{rank}_{\text{sp}}(\overline{X}_{\text{sptl}}) = 0$, that is, the (sample) spatial median has (sample) spatial rank 0.

Note that subtle differences exist between the sample quantities S_τ and S_{spear}: the former is defined as products of the vectors $\operatorname{dir}(X_i - X_j)$, while S_{spear} uses products of $\operatorname{dir}(X_i - X_j)$ and $\operatorname{dir}(X_i - X_k)$, with $X_j \neq X_k$. In Sections 4 and 5 we explore differences between the first eigenvectors of these two quantities for real and simulated datasets.

3.2. Consistency of Sample Covariance Matrices

Section 2.2 dealt with results pertaining to relationships between different population covariance matrices. In this section we collect and investigate results relating to population/sample pairs in order to gain some insight into which sample covariance matrices yield good estimators for the population eigenvectors. As in Section 2.2 most of the results in the literature are for the elliptic model. In some cases, results hold for more general settings; we will indicate when this is the case.

We will primarily focus on the relationship between the sample and population eigenvectors. Some results contain explicit expression for the asymptotic mean and covariance of the eigenvectors. We will provide references to these results for the interested reader.

We write $\Sigma_{ss} = \Gamma \Lambda_{ss} \Gamma^T$ and $S_{ss} = \widehat{\Gamma}_{ss} \widehat{\Lambda}_{ss} \widehat{\Gamma}_{ss}^T$ for pairs of population and sample covariance matrices as in the Notation at the end of Section 1. In the next theorem 'ss' refers to the spatial sign (sptl) or rank.

Theorem 3.1 Let $\mathbf{X}$ be a random sample of size n from the elliptic distribution $\mathcal{E}_d(\mu, \Sigma)$. For the pair (Σ_{ss}, S_{ss}) of population and sample covariance matrices, consider their jth eigenvalue/eigenvector pairs $(\lambda_{ss,j}, \eta_j)$ and $(\widehat{\lambda}_{ss,j}, \widehat{\eta}_{ss,j})$ respectively.

1. For the pair $(\Sigma_{\text{sptl}}, S_{\text{sptl}})$, based on the expected value and sample mean respectively, the following hold

$$\widehat{\eta}_{sptl,j} \to \eta_j \quad \text{as} \quad n \to \infty,$$

and, '*under general assumptions*', $\sqrt{n}(\widehat{\eta}_{sptl,j} - \eta_j)$ has a multivariate normal distribution for sufficiently large n.

2. Assume that the eigenvalues of Σ_{rank} are distinct, then the following hold

$$\widehat{\eta}_{rank,j} \xrightarrow{P} \eta_j \quad \text{and} \quad \widehat{\lambda}_{rank,j} \xrightarrow{P} \lambda_{rank,j} \quad \text{as} \quad n \to \infty,$$

and the limiting distributions of $\sqrt{n}(\widehat{\Gamma}_{\text{rank}} - \Gamma)$ and $\sqrt{n}(\widehat{\Lambda}_{\text{rank}} - \Lambda_{\text{rank}})$ are multivariate normal with mean zero.

Part 1 is given in Proposition 4.1 of C. Croux, E. Ollila, and H. Oja [6]. The authors do not specify what the '*general assumptions*' in the theorem are beyond requiring ellipticity. Part 2 follows from Theorem 6 of S. Visuri, E. Ollila, V. Koivunen, J. Möttoönen, and H. Oja [12]. The authors assume a slightly more general multivariate location-scale model which includes the elliptic model.

So far we considered the sign and the rank covariance matrices and the behaviour of the sample eigenvectors as the sample size increases. Our next result, which is shown in F. Han and H. Liu [7], deals with Kendall's τ covariance matrix. Han and Liu look at the sine of the angle between vectors as a means of measuring convergence. We denote the angle of two vectors, say ξ_1 and ξ_2, by $\angle(\xi_1, \xi_2)$.

Theorem 3.2 Let $\mathbf{X}$ be a random sample of size n from the elliptic distribution $\mathcal{E}_d(\mu, \Sigma)$. Consider the pair of Kendall's τ covariance matrices (Σ_τ, S_τ) for the population and sample respectively. Write $\lambda_{\tau,j}$ for the eigenvalues of Σ_τ, given in decreasing order, and let $\eta_{\tau,j}$ and $\widehat{\eta}_{\tau,j}$ denote the jth eigenvector of Σ_τ and S_τ respectively. The sine of the angle $\angle(\eta_{\tau,1}, \widehat{\eta}_{\tau,1})$ satisfies

$$|\sin(\angle(\eta_{\tau,1}, \widehat{\eta}_{\tau,1}))| \le \frac{2}{\lambda_{\tau,1} - \lambda_{\tau,2}} \|S_\tau - \Sigma_\tau\|_{sup}, \tag{20}$$

and the difference between the first m eigenvectors Γ_m of Σ_τ and $\widehat{\Gamma}_{\tau,m}$ of S_τ is bounded by

$$\|\widehat{\Gamma}_{\tau,m} - \Gamma_{\tau,m}\|_F \le \frac{2\sqrt{2m}}{\lambda_{\tau,m} - \lambda_{\tau,m+1}} \|S_\tau - \Sigma_\tau\|_{sup}. \tag{21}$$

For $0 < \beta < 1$, put

$$\kappa = \frac{16}{3} \left[\frac{\operatorname{tr}(\Sigma_\tau)}{\lambda_{\tau,1}} + 1 \right] [\log d + \log(1/\beta)].$$

If n is large enough and satisfies $n \ge \kappa$, then, with probability exceeding $(1 - \beta)$,

$$\|S_\tau - \Sigma_\tau\|_{sup} \le \|\Sigma_\tau\|_{sup} \sqrt{\kappa / n}.$$

Theorems 3.1 and 3.2 tell us that, under the elliptic model, the eigenvectors of the sample covariance matrices S_{sptl}, S_{rank} and S_τ converge to the corresponding eigenvectors of Σ. This implies that the eigenvectors of any of these matrices may be regarded as estimators of the population eigenvectors.

The proof of (21) follows from the Davis-Kahan inequality,

see the reference in F. Han and H. Liu [7]. A variation of the David-Kahan inequality leads to an estimate for the error of the first m eigenvectors.

For fixed dimension d, the term κ/n will eventually become small. However when d increases, the rate of growth of d/n will drive the process. In this case, spiked models with one large eigenvalue are candidates for establishing consistency provided $(\log d)/n \to 0$ as d increases. F. Han and H. Liu [7] provide consistency results for high-dimensional data with sparse eigenvectors. These scenarios are beyond the scope of this paper, and we will therefore not present the results here.

In the next two sections we study the behaviour of the sample eigenvectors for data from elliptic and non-elliptic distributions and assess the performance of these eigenvectors as estimators of the population eigenvectors. Before we turn to data, we briefly return to the sign covariance matrices and their dependence on centring parameters.

From Theorem 2.2 the spatial median is the expected value for the elliptic distribution. However, for non-elliptic models, this may not hold and the particular choice of centring parameter may matter. The next result by A. Dürre, D. Vogel, and D. E. Tyler [13] establishes conditions under which the sample sign covariance matrix is a consistent estimator for the population quantity.

Theorem 3.3 Let $\mathbf{X}$ be a random sample of size n, and X a random vector, both from a distribution F. Consider the sign covariance matrices Σ_{sptl} of X and S_{sptl} of $\mathbf{X}$. For a random vector Y assume there is $\beta \geq 0$ such that

$$\mathbb{E}\left(\|X - Y\|^{-1/(1+\beta)}\right) < \infty.$$

Take Y as a centring parameter for Σ_{sptl} and assume that there is a sequence Y_n of random vectors which satisfies

$$n^{\beta}\|Y_n - Y\| \leq t_n \quad \text{a.s.},$$

with t_n a sequence of random variables that converges a.s. for $\beta > 0$, and which converge a.s. to 0 for $\beta = 0$, then

$$S_{\text{sptl}}(Y_n) \overset{a.s.}{\to} \Sigma_{\text{sptl}}(Y). \tag{22}$$

Theorem 3.3 taken from A. Dürre, D. Vogel, and D. E. Tyler [13], tells us that the eigenvectors of the sample sign covariance matrix converge to those of the population if a centring parameter for the population covariance matrix is chosen that is not too far away from the distribution F and if the sample centring parameters lie in a suitably small neighbourhood of this population centring parameter. Theorem 3.2 indicates that the eigenvectors of $\Sigma_{\text{sptl}}(Y)$ may vary with the centring parameter Y. Further, the eigenvectors are likely to depend on the underlying distribution of the data. We will not pursue this avenue here but will look at different centring parameters for S_{sptl} in Section 4.

In their paper A. Dürre, D. Vogel, and D. E. Tyler [13] establish conditions for *in probability* convergence of the sample sign covariance matrix and prove results on the asymptotic normality of the vectorised form of $S_{\text{sptl}}(Y_n) - \Sigma_{\text{sptl}}(Y)$, which A. Dürre, D. Vogel and R. Fried [4] strengthen further.

In this section we presented results for the three covariance matrices sign, rank and Kendall's τ which relate the sample eigenvectors and the population eigenvectors. We have not presented results pertaining to Spearman's rank covariance matrix, primarily because $\Sigma_{\text{spear}} = \Sigma_{\text{rank}}$. A quick glance back to (18) shows that the sample version of Spearman's rank covariance matrix differs from that of (16). In the next section we examine this difference more closely and then explore all four estimators for real data.

3.3. Relationships Between Sample Covariance Matrices

From diagram (13) in Section 2.2 we know that the sign and Kendall's τ covariance matrices agree under the elliptic model. Further, the rank and Spearman's rank covariance matrices agree for any distribution, that is,

$$\begin{aligned} \Sigma_{\text{sptl}} &= \Sigma_{\tau} \quad \text{for } X \in \mathcal{E}_d(\mu, \Sigma) \text{ and location parameter } \mu, \\ \Sigma_{\text{rank}} &= \Sigma_{\text{spear}} \quad \text{for any } F. \end{aligned} \tag{23}$$

The eigenvectors of all four population matrices agree, and it is therefore natural to examine whether the same relationships hold for the corresponding sample covariance matrices.

Relationships between S_{rank}, S_{τ} and S_{spear} do not seem to have been considered in the literature. However, using (16) – (18) one can show that for a random sample $\mathbf{X}$ from any distribution the three covariance matrices satisfy

$$S_{\text{rank}}(\mathbf{X}) = \left[1 - \frac{1}{n-1}\right] S_{\text{spear}}(\mathbf{X}) + \frac{1}{n-1} S_{\tau}(\mathbf{X}). \tag{24}$$

Equation (24) shows that S_{rank} and S_{spear} converge to the same quantity as the sample size increases. This is expected since their population covariance matrices are the same. For a finite sample, S_{τ} can be interpreted as a correction term whose effect decreases with increasing sample size.

If we had used the parametrisation of J. I. Marden [1] in the definition of S_{rank} including Marden's factor $[n^2(n-1)]^{-1}$ and corresponding factors for the other two covariances matrices, (24) would be replaced by a similar equation with the same asymptotic behaviour.

Returning to the first equality of (23), which is shown in S. Taskinen, I. Koch, and H. Oja [2], we note that the authors desired a population quantity that equals the sign covariance matrix, but does not depend on the spatial median. This suggests that the sample analogue of Σ_{τ} could be regarded as an estimator of Σ_{sptl}. The latter would avoid having to calculate the spatial median or another quantity for the sample such as the sequence of Y_n of Theorem 3.2. We explore some of these ideas in in Section 4.

In the remaining part of the paper we restrict attention to real and simulated data, and explore the degree of similarity of the eigenvectors of the covariance matrices in order to gain insight into their behaviour of data from elliptic and non-elliptic distributions.

4. Covariance Matrices and Their Eigen-analysis for Real Data

4.1. Behaviour of Projections Based on the Eigenvectors for Real Data

The first few eigenvectors of the sample covariance matrix S are standard tools for projecting data into lower-dimensional subspaces and for determining interesting features. PCA is driven by variance and based on the Euclidean distance. These properties make PCA projections particularly suitable for data from the Gaussian distribution. For data with outliers, heavy-tailed or skewed data, however, the eigenvectors of the sample covariance matrix can be 'distorted', for example into the direction of the outliers, and may not adequately reflect the performance of the main point clouds.

In this section we consider PC score plots arising from the first two eigenvectors of the standard and the other four sample covariance matrices for three datasets: the HDLSS breast tumour data of L. J. van't Veer "et al." [14], the 12-dimensional athletes data of R. D. Cook and S. Weisberg [15] and the 13-dimensional wine recognition data of S. Aeberhard, D. Coomans, and O. de Vel [16].

Unless otherwise stated, we only work with raw data and assume that these data are centred. We calculate the *jth PC scores, PC_j,* as the projections of the centred raw data onto the jth eigenvector of the covariance matrices we consider: the standard covariance matrix S and the covariance matrices $S_{\text{sptl}}, S_{\text{rank}}, S_\tau$ and S_{spear}. Typically we will state which covariance matrix is used to obtain the PC scores. The *PC_1/PC_2 score plots* refer to the scatterplots of the PC_1/PC_2 scores. Since we only consider the first two PCs, we often refer to their score plots simply as *PC score plots*.

The purpose of the PC score plots is to gain insight into differences that arise when we project data onto different sample eigenvectors and to attempt to gain some understanding of why they occur or what causes the differences. We complement the visual displays with tables showing angles between the eigenvectors of the different covariance matrices.

An integral part of any principal component analysis is the contribution to total variance of each component or direction. We consider eigenvalues of the standard and spherical covariance matrices. We extend the common definition of contribution to total variance—also known as the proportion of total variance— to the spherical covariance matrices, namely the proportion of each eigenvalue to the trace of the respective covariance matrix. We compare the ratios $\lambda_j / \operatorname{tr}(\Sigma)$ with the analogous parts of the different (sample) covariance matrices. For notational convenience we will often refer to these proportions as *eigenvalue ratios*.

4.2. High-dimensional Real Data

The breast tumour data of L. J. van't Veer "et al." [14] consist of 78 subjects with breast tumour and 4751 gene expressions given as $\log_{10}$-transformed expression levels of genes. Of the 78 patients 44 had tumours which did not metastasise within the critical five years. We use this information only in the presentation of Figure 1, which shows the two groups in different colours.

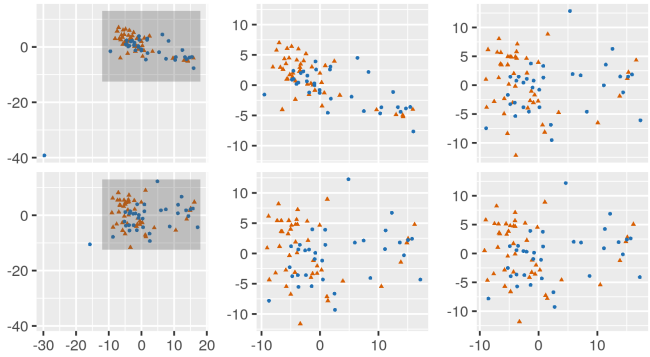

Figure 1. *PC score plots of the breast tumour data from S in the top row, and S_{rank} in the bottom row. Middle panels: zoomed-in grey parts of left panels; right panels:PC score plots from S and S_{rank} without the outlier at about (-30, -40) in the top left panel.*

We calculate PC_1/PC_2 scores using the standard covariance matrix and the rank covariance matrix, and we show the resulting PC score plots in Figure 1. The PC scores derived from the first two eigenvectors of the other three covariance matrices look similar to those derived from S_{rank} and are therefore not shown. Table 1 provides more quantitative details relating to the angles of pairs of vectors.

The top row in Figure 1 refers to PC_1/PC_2 score plots of the standard covariance matrix S and the bottom row shows the corresponding plots for the rank covariance matrix S_{rank}. The breast tumour data contain an outlier, observation 54, which is clearly visible in the PC score plot of S, but which is hardly noticeable as an outlier in the plot below.

Table 1. *Angles of pairs of first eigenvectors of the breast tumour data from different sample covariance matrices; numbers on the diagonal refer to angles of pairs of eigenvectors of the* 'same type of' *covariance matrices but calculated with and without observation 54. All other comparisons are based on all data.*

	$\boldsymbol{S}$	$\boldsymbol{S}_{\text{sptl}}$	$\boldsymbol{S}_{\text{rank}}$	$\boldsymbol{S}_{\boldsymbol{\tau}}$	$\boldsymbol{S}_{\text{spear}}$
S	20.3	23.2	18.4	18.4	18.4
S_{sptl}	23.2	0.5	6.3	6.4	6.3
S_{rank}	18.4	6.3	3.0	0.2	0
S_τ	18.4	6.4	0.2	3.0	0.2
S_{spear}	18.4	6.3	0	0.2	3.0

The two middle panels show the zoomed-in dark grey sections from the left panels respectively and the right panels show plots of PC_1/PC_2 scores obtained from the updated covariance matrices after removal of the outlying observation 54. The middle panels are clearly very different from each other, while the score plots on the right look much more similar.

The middle and right panels of the S_{rank}-plots are basically the same, while the S-related plots look very different indicating the change in the eigenvectors with and without the outlier, and the effect of the outlier on the standard covariance matrix and its eigenvectors. The score plots confirm that S_{rank} handles the outlier well while S is sensitive to outliers.

Table 1 complements the information provided in Figure 1, showing angles in degrees between first eigenvectors of the different covariance matrices. The numbers on the diagonal show the angles of the eigenvectors for the same two covariance matrices but calculated with and without the outlying observation 54. So for the pair $(S_{\text{rank}}, S_{\text{rank}}$ we obtain an angle of 3 degrees. These numbers tell us that for all covariance matrices other than S this angle, and hence the change in the eigenvectors, is very small.

The table highlights that the eigenvectors of the standard covariance matrix differ considerably form those of the other matrices. The eigenvectors of S_{rank} and S_{spear} are identical, and the eigenvectors of S_τ are also very similar to the other two spherical approaches. The eigenvectors of S_{sptl} are distinct from the other three spherical covariance matrices, and share the largest angle with S.

The left panel of Figure 2 shows the eigenvalue ratios—the contributions to total variance—of the breast tumour data. These eigenvalue plots, with the index on the x-axis, are based on S, shown in dark blue, and S_{rank} in light blue. The dashed lines show the corresponding plots of the data after removal of the outlying observation 54 again in dark and light blue. The light blue dashed line is almost invisible, as it agrees well with the solid line.

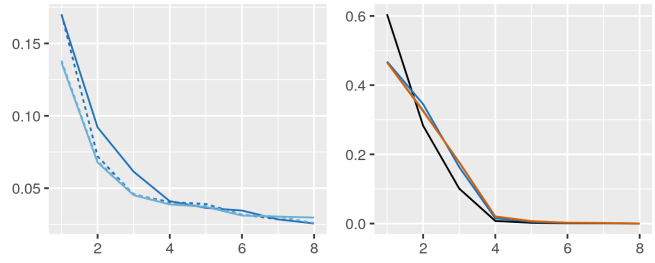

Figure 2. *Eigenvalue ratios of first 8 eigenvalues for the breast tumour data (left) and the athletes data (right). Left panel: all data with S dark blue, with S_{rank} light blue, without observation 54 shown as dashed –dark for S, light for S_{rank}; right panel: S black, S_{rank} light blue, S_{sptl} dark blue, S_τ red.*

Observe that the first three eigenvalue ratios of S are larger than those of S_{rank}. Of interest is the dark blue dashed line corresponding to S without observation 54: its first eigenvalue ratio is the same as that obtained from all data, but the next two eigenvalue ratios are more similar to those arising from S_{rank}. As we have noticed in the PC score plots in Figure 1, those arising from S of the reduced data are more similar to those of S_{rank} with or without the outlier.

The panel on the right corresponds to the athletes data which we will discuss in the next section.

The analysis of the breast tumour data highlights the big difference in the eigenvectors of the standard and the other sample covariance matrices in the presence of an outlier. It shows that by using S_{rank} in addition to S, we may gain valuable information about and insights in the data.

4.3. Classical Real Data

The *athletes* data were collected at the Australian Institute of Sport by Richard Telford and Ross Cunningham, see R. D. Cook and S. Weisberg [15]. For 100 female and 102 male athletes 12 variables consisting of physical as well as blood-related quantities are compared.

We calculate PC_1/PC_2 scores for a range of covariance matrices and show the resulting plots in Figure 3 with different colours for male and female athletes. From left to right the score plots are those obtained from S, S_{rank}, S_τ and S_{sptl}. The score plot obtained from S_{spear} is not shown here, as it looks identical to that of S_{rank}. The score plots of the first three panels from the left look almost identical, the score plot of S_{sptl} shows a slightly different distribution of some observations that are further away from the centre of the point cloud but the difference is not large.

The angles between pairs of first and pairs of second eigenvectors from different covariance matrices are shown in Table 2, with odd rows corresponding to first eigenvectors, and even rows to second eigenvectors. The mostly small angles shown in the table agree well with the similarity between the score plots. Table 2 includes three results for spatial sign covariance matrices S_{sptl}: they are based on the sample mean, the spatial or ℓ_2-median (called $S_{\text{sptl},2}$) and the ℓ_1-median (called $S_{\text{sptl},1}$) respectively as centring parameters. The ℓ_1-median is defined similarly to the ℓ_2-median, but uses the ℓ_1-norm.

It may be surprising to see that the eigenvectors of the athletes data corresponding to the three spatial covariance matrices appear to be more similar to each other, but, as a group, differ more from all others. The largest angles arise from the eigenvectors of the spatial sign covariance matrix $S_{\text{sptl},2}$, the 'natural' covariance matrix for direction vectors, when compared to S and the other 'non spatial sign' matrices. We will not examine reasons for this disprecency.

Table 2 also shows that the eigenvectors of S differ more from the spatial sign covariances matrices than from those of S_{rank}, S_{spear} and S_τ. The latter three covariance matrices have almost identical eigenvectors. A more detailed table, which extends the information in this table, is shown in Table A1 in the Appendix.

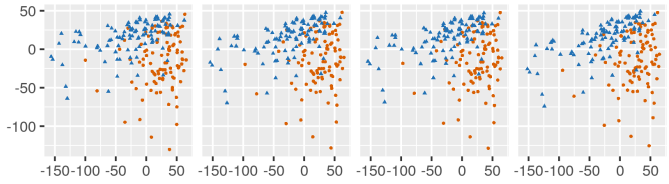

Figure 3. *PC score plots of the athletes data with S, S_{rank}, S_τ, and S_{sptl}.*

Table 2. *Athletes data: angles of first eigenvectors (rows 1,3,5) and second eigenvectors (rows 2,4,6) of sample covariance matrices.*

	S_{sptl}	S_{rank}	S_τ	S_{spear}	$S_{\text{sptl},1}$	$S_{\text{sptl},2}$
S	7.46	4.13	3.61	4.13	8.80	10.16
	8.63	2.99	2.34	3.00	8.42	10.26
S_{sptl}		3.35	3.86	3.34	1.52	3.00
		6.22	6.97	6.21	0.84	2.91
S_{rank}	3.35		0.55	0	4.74	6.06
	6.22		0.78	0	5.86	7.42

Regarding eigenvalue ratios pertaining to the athletes data, the right panel of Figure 2 shows these ratios of the first 8 eigenvalues corresponding to S in black, S_{rank} in light blue, S_{sptl} in dark blue, and S_τ in red. The last essentially coincides with the light blue one corresponding to S_{rank} and hence is barely visible. Note that the red S_τ graph differs from that of S_{sptl} mostly for the second and third ratio. This behaviour is of interest in light of the statement of Theorem 2.2 that the two population matrices agree under the elliptic model.

For all three spherical covariance matrices the first eigenvalue ratio is much smaller than that of S, while eigenvalue ratios 2 to 4 exceed those of S.

The aim of the analysis of the athletes data is to illustrate how the various covariance matrices, their eigenvalues and eigenvectors are related in practice.

Our second classical dataset is the *wine recognition* data of S. Aeberhard, D. Coomans, and O. de Vel [16] which are the results of a chemical analysis of three cultivars of wine grown in the same region in Italy. The data consist of measurements of 13 variables and 178 observations.

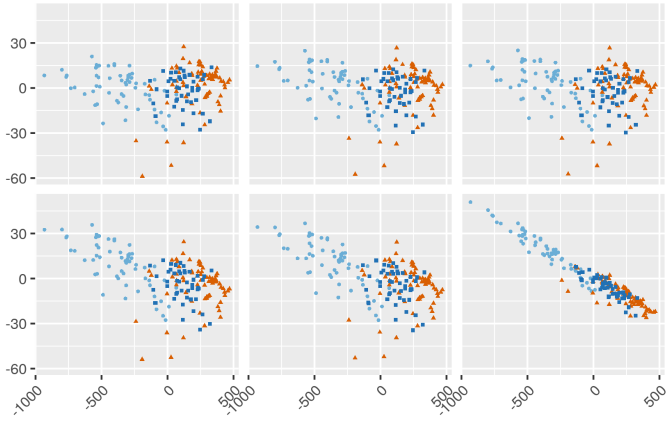

Figure 4. *PC_1/PC_2 score plots of the raw wine data from the standard (left), the rank (middle), and Kendall's τ (right) covariances in the top row and from spatial covariance matrices with centring parameters mean, ℓ_1-median and spatial median in the bottom row.*

Figure 4 shows the PC_1/PC_2 score plots relating to the standard, rank and Kendall's τ covariance matrices in the top row, and to sign covariances matrices in the bottom row which use, from left to right, the sample mean, the ℓ_1-median and the spatial median as centring parameter. The three cultivars are shown in contrasting colours but play no other role in these plots. We complement Figure 4 with Tables 3 and 4 in this section and with a Table A2 and Figure A1 in the Appendix. Table 3 shows angles between pairs of eigenvectors in the same format as in Table 2, and Table 4 lists the correlation coefficients of the first two PCs for each sample covariance matrix. Table A2 is a more comprehensive version of Table 3; and Figure A1 displays eigenvalue ratios for the wine data, it is similar in style to the panels of Figure 2.

The much larger angles arising from the second eigenvector pairs and shown in the even rows in Table 3 are worth noting. Indeed, angles with the second eigenvector of $S_{\text{sptl},2}$ exceed 70 degrees and the $S_{\text{sptl},2}$-PC scores, displayed in the bottom right panel of Figure 4, have the very high (negative) correlation of -0.97, which is listed in Table 4. Recall that population PC scores are uncorrelated.

Table 3. *Wine data: angles of first eigenvectors (odd rows 1,3,5) and second eigenvectors (even rows 2,4,6) of the sample covariance matrices from a range of covariance matrices.*

	$\boldsymbol{S}_{\text{sptl}}$	$\boldsymbol{S}_{\text{rank}}$	$\boldsymbol{S}_{\boldsymbol{\tau}}$	$\boldsymbol{S}_{\text{spear}}$	$\boldsymbol{S}_{\text{sptl},1}$	$\boldsymbol{S}_{\text{sptl},2}$
S	1.49	0.39	0.42	0.39	1.65	3.14
	3.54	1.09	1.42	1.08	7.08	71.11
S_{sptl}		1.13	1.11	1.13	0.37	2.51
		3.67	3.93	3.66	7.32	71.60
S_{rank}	1.13		0.04	0	1.30	2.94
	3.67		0.43	0.01	6.60	71.02

Table 4. *Wine data: correlation coefficients of PC scores for each covariance matrix.*

$\boldsymbol{S}$	$\boldsymbol{S}_{\text{sptl}}$	$\boldsymbol{S}_{\text{rank}}$	$\boldsymbol{S}_{\boldsymbol{\tau}}$	$\boldsymbol{S}_{\text{spear}}$	$\boldsymbol{S}_{\text{sptl},1}$	$\boldsymbol{S}_{\text{sptl},2}$
0	-0.53	-0.16	-0.17	-0.16	-0.56	-0.97

Table 4 shows the correlation coefficients arising from the PC scores. The correlation coefficients of the S-based PC scores are, of course, zero, and those arising from $S_{\text{rank}}, S_{\text{spear}}$ and S_τ are small. At first glance it may be counterintuitive to see non-zero or even large correlations for some of the PC scores.

A moment's thought reveals that the PC scores are obtained as projections of the raw data onto the eigenvectors of the different covariance matrices, resulting in uncorrelated scores for the eigenvectors of S. It is worth noting that, as the angles in pairs of eigenvectors between S and those of a spherical covariance matrix increase, so does the strength of the correlation pertaining to that spherical covariance matrix. We may interpret the increase in angle between an eigenvector of S and the corresponding one of spherical covariance matrix as the latter 'rotating away from the uncorrelated position'. Tables 3 and 4 confirm this interpretation.

All three S_{sptl}-based PC scores are more correlated and their eigenvectors have much larger angles with those of S. Regarding $S_{\text{sptl},2}$, its second eigenvector is much closer to being orthogonal to all other second eigenvectors than to being aligned with them and this almost orthogonality is complemented by the almost maximal correlation of the PC scores of $S_{\text{sptl},2}$.

There could be a number of reasons for this 'behaviour' of $S_{\text{sptl},2}$; these could include the large discrepancy between the centring parameters which might affect the direction of the eigenvectors and the high correlation of the scores. Irrespective of the actual reason which we can only conjecture, the analysis shows clearly that S_{sptl} is sensitive to the choice of centring parameter.

Figure A1 in the Appendix shows that the overall behaviour of the eigenvalue ratios of the wine data is similar to that of the athletes data: the first eigenvalue ratios of the spherical covariance matrices are smaller than that of S, but the next three are larger.

Of interest here is that S_τ and S_{sptl} almost agree on the first few eigenvalue ratios, while those corresponding to $S_{\text{sptl},1}$ and $S_{\text{sptl},2}$ are lower for the first ratio and higher for the second to

forth ratios. In the athletes data we observe that the eigenvalue ratios of S_τ are closer to those of S_{rank} than to those of S_{sptl}. So again, the similarity of population quantities Σ_{sptl} and Σ_τ is not reflected in S_{sptl} and S_τ.

It is beyond the scope of this paper to examine potential reasons for the discrepancy of the eigenvectors and score plots arising from $S_{\text{sptl},2}$. Our aim is to draw attention to this surprising behaviour in order to showcase unexpected results that can occur for different centring parameters of the sign covariance matrix and to recommend that care be taken when employing a sign covariance matrix in dimension reduction or other PC-related analyses.

5. Simulations

This section presents new results showcasing the performance of the PCA methods we presented on simulated data. Our simulation results include the performance of standard and spherical PCA on non-elliptic scenarios which are not covered by the theoretical framework, but are common in applications, thereby enhancing and complementing the previous sections.

5.1. The Families of Distributions

We simulate data from elliptic and non-elliptic families of distributions and we relate the simulation results back to the theorems in Sections 2.2 and 3.2. Our families of distributions consist of two classes of elliptic distributions, followed by two classes of non-elliptic distributions:

1. MVN: multivariate normal distributions;
2. MVT-ν: multivariate t-distributions in ν degrees of freedom;
3. BMN: bimodal mixtures of two multivariate normal distributions, and;
4. MSN-ρ: multivariate skew normal distributions with skewness parameter ρ.

We start with the parameters for the multivariate normal distribution and then describe how we select the parameters in the other distributions so that the covariance matrices of the simulated data are comparable.

5.1.1. Multivariate Normals [MVN]

Let $\mathcal{N}(\mu, \Sigma)$ be the multivariate normal distribution. In the simulations we put

$$\begin{aligned} &\mu = 0, \text{ the zero vector with entries } \mu_j, j \le d, \text{ and} \\ &\Sigma = \text{diag}(\sigma_j) \quad \text{with } \sigma_j = 2^{1-j} \text{ for } j \le d. \end{aligned} \tag{25}$$

This results in a model with exponentially decreasing eigenvalues. We have picked this model as it reflects many datasets and results in distinct eigenvalues.

Other models, eg, spiked models with one or a small number of large eigenvalues and all other eigenvalues almost constant, could also be of interest especially for HDLSS data. The latter is not the main concern of this paper which primarily focusses on moderate dimensions and increasing sample sizes.

5.1.2. Multivariate t-distributions in ν Degrees of Freedom [MVT-ν]

Let $t_\nu(\mu, \Sigma_0)$ be the multivariate t-distributions and observe that for $\nu > 2$ the covariance matrix is the scaled matrix $\frac{\nu}{\nu-2}\Sigma_0$. In our case, $\Sigma = \frac{\nu}{\nu-2}\Sigma_0$, with μ and Σ as defined in (25).

5.1.3. Bimodal Mixture of Normals [BMN]

We use a mixture of two multivariate normal distributions $\mathcal{N}(\mu^{[k]}, \Sigma^{[k]})$ with $k = 1, 2$ and parameters

$$\begin{aligned} \mu^{[1]} =& 0,\ \Sigma^{[1]} \text{ diagonal with } \sigma_1^{[1]} = 2^{-1},\ \sigma_2^{[1]} = 1,\ \sigma_j^{[1]} = 2^{1-j} \text{ if } j > 2, \text{ and} \\ \mu_1^{[2]} =& \frac{20}{7},\ \mu_j^{[2]} = 0 \text{ if } j \ge 2,\ \Sigma^{[2]} \text{ diagonal with } \sigma_j^{[2]} = 2^{-j} \text{ for } j \le d. \end{aligned}$$

For $X^{[k]} \in \mathcal{N}(\mu^{[k]}, \Sigma^{[k]})$ let C be a random variable from the Bernoulli distribution with success probability $p = \frac{13}{20}$ and put

$$\begin{aligned} Z =& CX^{[1]} + (1-C)X^{[2]}, \text{ with components } Z = \{z_j : j \le d\} \quad \text{and} \\ X =& \{x_j\} \text{ with } x_1 = \sqrt{\frac{14}{33}}(z_j - 1),\ x_2 = \sqrt{\frac{40}{59}}z_2, \text{ and } x_j = \sqrt{\frac{40}{33}}z_j, \text{ if } j > 2. \end{aligned}$$

An easy calculation shows that X has the mean μ and covariance matrix Σ as in (25).

5.1.4. Multivariate Skew Mormal Distributions with Skewness Parameter ρ [MSN-ρ]

To generate random vectors from the multivariate skew normal distributions we follow A. Azzalini and A. Capitanio [17]. Their starting point for a skew distribution is a non-zero mean, which we take to be

$$\begin{aligned} &\mu_{sn} = [\mu_1, 0, \ldots, 0] \text{ with } \mu_1 \ne 0 \text{ and then put} \\ &\Omega = \Sigma + \mu_{sn}\mu_{sn}^T, \end{aligned} \tag{26}$$

with Σ as in (25).

A. Azzalini and A. Capitanio [17] phrase their proposal in terms of the pair (Ω, α), with Ω as in (26) and α a skewness parameter which is related to a vector δ by

$$\alpha = \left(\frac{1}{\sqrt{1 - \delta^T\Omega^{-1}\delta}}\right)\Omega^{-1}\delta, \tag{27}$$

and δ is further related to the population mean, which, in our case, leads to $\mu_{sn} = \sqrt{\frac{2}{\pi}}\delta$. From (26) it follows that only the first entry δ_1 of δ is non-zero.

The definition of α imposes the restriction $\delta^T\Omega^{-1}\delta < 1$. For Σ and Ω as in (25) and (26) respectively this implies that $\delta_1 < \sqrt{\frac{\pi}{\pi-2}}$. Instead of using δ_1, we reparametrise the multivariate skew normal using a parameter ρ defined by

$$\rho = \delta_1\sqrt{\frac{\pi-2}{\pi}} \times 100. \tag{28}$$

Our final step includes centring of the skew normal random vectors in order to make them comparable with our other models.

5.2. Evaluating Estimators of the Population Eigenvector as n Grows

For the families of distributions described in Section 5.1 we study the asymptotic behaviour of the first eigenvector of the different sample covariance matrices $S, S_{\text{sptl}}, S_{\text{rank}}, S_\tau$ and S_{spear} as estimators of the first population eigenvector as the sample size increases. For the calculation of S_{sptl} we use the sample mean as centring parameter unless otherwise specified. These results relate to Theorems 3.1 and 3.2.

We measure convergence of the estimators by the angle between the population eigenvector and each sample eigenvector, as done in Theorem 3.2. Algorithm 1 summarises the calculations relating to the eigenvectors and their contributions to total variance. In this and Algorithm 2 we consider the following six distributions, denoted by F_κ in the algorithms:

1. multivariate normal distribution, referred to as $\mathcal{N}$,
2. t-distributions MVT-4, MVT-10, referred to as t_4 and t_{10},
3. bimodal mixture distribution, referred to as B, and,
4. skew normal distributions MSN-95 and MSN-99, referred to as S_{95} and S_{99} in the tables and figures.

Algorithm 1. Estimators of population eigenvectors.

1. Fix the dimension d.
   For the population parameters (λ_j, η_j), q_j, Σ of (25), and $\text{tr}(\Sigma)$, calculate (λ_j, η_j) and $q_j = \lambda_j/\text{tr}(\Sigma)$ for $j \le d$.
2. Fix the distribution F_κ for $\kappa \in \{1, \ldots, 6\}$.
3. For sample parameters $(\widehat{\lambda}_j, \widehat{\eta}_j)$ and $q_{ss,.,j}$, fix $n = 10^\ell$, $\ell \in \{2,\ 2.1,\ 2.2, \ldots, 3.7\}$.
4. For $k \le 100$ generate $\mathbf{X}_k$, each of size n.
5. For $ss \le 5$ calculate $S_{ss,k}$, $\text{tr}(S_{ss,k})$, $(\widehat{\lambda}_{ss,k,j}\widehat{\eta}_{ss,k,j})$ and $q_{ss,k,j} = \widehat{\lambda}_{ss,k,j}/\text{tr}(S_{ss,k})$ for $j \le d$.
6. Compare population and estimators: for $j \le d$ do
   $\bar{q}_{ss,.,j} \leftarrow \sum_k q_{ss,k,j}/100$
   $\gamma_{ss,k,j} \leftarrow \angle(\eta_j, \widehat{\eta}_{ss,k,j})$ for $k \le 100$
   $\bar{\gamma}_{ss,j} \leftarrow \sum_k \gamma_{ss,k,j}/100$.
7. Return to 2 if $\kappa < 6$.

We carried out the calculations of Algorithm 1 for $d = 5, 10, 20$ and 50. For $d = 20$ we used all values ℓ in the Step 3 and a subset of values ℓ for the other dimensions. The results for different dimensions are similar. For this reason we only present results obtained for $d = 20$. In this section we look at the eigenvectors regarded as estimators of the population quantities, and Section 5.3 focusses on relationships between population and sample eigenvalues.

In the figures and discussion below we restrict attention to the first eigenvector, that is, the case $j = 1$ in Steps 5 and 6, and we show the results obtained from the three sample covariance matrices S, S_{sptl} and S_{rank}. As we shall see in Section 5.4, $S_{\text{rank}}, S_{\text{spear}}$ and S_τ are very similar and we therefore use S_{rank} as a representative for these three. We will comment on the results of these comparisons in relation to (24) in Section 5.5.

Figure 5 depicts graphs of the mean angles $\bar{\gamma}_{ss,j}$ of Step 6 in Algorithm 1 for the first eigenvectors as a function of ℓ and hence of the sample size. The graphs corresponding to the different covariance matrices are shown in blue for S, in red for S_{sptl} and in light blue for S_{rank}. In addition to the mean angles, we also show the angles between the population and sample eigenvector for each of the 100 datasets. The latter are shown in the form of boxplots – in the same colours as the three means and separated in the horizontal direction for easier visibility.

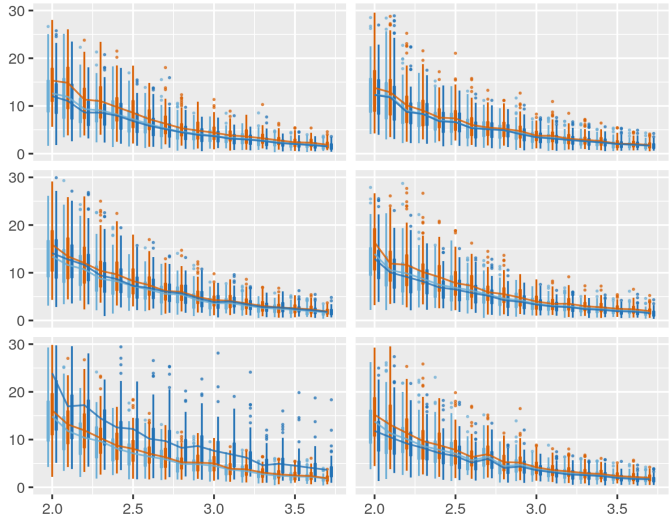

Figure 5. *Mean angles of first eigenvectors with $d = 20$ of Algorithm 1, based on Σ and S (in blue), S_{rank} (in light blue) and S_{sptl} (in red) as ℓ increases (and $n = 10^\ell$); with $\mathcal{N}(\mu, \Sigma)$ (top left), t_{10} and t_4 (middle and bottom left), bimodal (top right), and multivariate skew normal S_{95} and S_{99} (middle and bottom right).*

Table 5. *Mean angles corresponding to Figure 5.*

n	S_{ss}	$\mathcal{N}$	t_{10}	t_4	B	S_{95}	S_{99}
100	S	12.05	14.10	23.91	12.31	12.95	11.81
	S_{sptl}	15.31	15.89	16.12	13.79	16.36	15.22
	S_{rank}	12.63	13.39	14.87	12.33	13.94	13.80
5011	S	1.53	1.81	3.65	1.68	1.53	1.67
	S_{sptl}	1.88	1.87	1.89	1.82	1.97	2.00
	S_{rank}	1.57	1.66	1.78	1.69	1.64	1.80

Each figure panel focusses on a different underlying distribution for the data generating process; the left panels feature elliptic distributions: the multivariate normal and the

multivariate t-distribution in 10 and 4 degrees of freedom, and the right panels feature non-elliptic distributions: the bimodal mixture distribution and the multivariate skew normal distribution with $\rho = 0.95$ and 0.99.

For $n = 100$ and 5011, the values shown in the graphs for the corresponding $\ell = 2$ and $\ell = 3.7$ are listed in Table 5. Table A3 in the Appendix also contains results for $n = 501$ and 1000. From Figure 5 and Table 5 we find that the mean angles of all three S_{ss}, and all six distributions converge to zero. However, the decrease in the angle with increasing n depends on the distribution and the particular covariance matrix: the light blue curve arising from S_{rank} is on par with S for all but the heavy-tailed t_4-distribution shown in the bottom left panel, for which S performs considerably worse and appears to converge much more slowly than the other two. The graph of S_{sptl} is slowest to converge in all cases other than in the bottom left panel (t_4), where it is comparable to S_{rank}, as n becomes large. In Section 5.5 we link the results from this simulation to Theorems 2.2 and 3.1.

Overall the first eigenvector of S_{rank} appears to be an excellent estimator of the first population eigenvector irrespective of the underlying distribution. The mean angles decrease almost monotonically for each of the six distributions from just below 15 degrees at $n = 100$ (shown as the x-value 2) to below 2 degrees when $n = 10^{3.7} \sim 5011$.

5.3. Evaluations of Population and Sample Contributions to Variance

The trace of spherical covariances typically differs from that of the standard covariance matrix, and it therefore makes more sense to compare eigenvalue ratios. Here we normalise the eigenvalues with the trace of their matrix and then compare the individual contributions to total variance, as mentioned in Section 4.1.

In the Step 1 of Algorithm 1 we calculate the population ratios $q_j = \lambda_j / \operatorname{tr}(\Sigma)$. Steps 2 to 5 contain analogous calculations for the sample covariance matrices.

Using the same figure panel arrangement of distributions as in Figure 5 with the elliptic distributions on the left and the non-elliptic ones on the right, Figure 6 shows the first eigenvalue ratios of Σ in black, S in dark blue, S_{rank} in light blue and S_{sptl} in red. For the three sample cases, the boxplots summarise the spread of ratios over 100 simulations. As in the previous figure, ℓ is given on the x-axis, with $n = 10^{\ell}$ the sample size.

Figure 6 tells us that the blue S is centred around the black line arising from Σ, the latter being independent of n, as it represents the population value. As n increases the range of the boxplots decreases and the ratios converge. In the case of S this ratio converges to the eigenvalue ratio of Σ with the exception the non-elliptic multivariate skew normal distribution t_4. For S_{rank} and S_{sptl} the first eigenvalue ratios also converge with increasing n, but they do *not* converge to $\lambda_1 / \operatorname{tr}(\Sigma)$. Indeed, with the exception of S_{sptl} in the binomial case, shown in the top right panel, the first contribution of S_{rank} and S_{sptl} is considerably smaller than that of Σ.

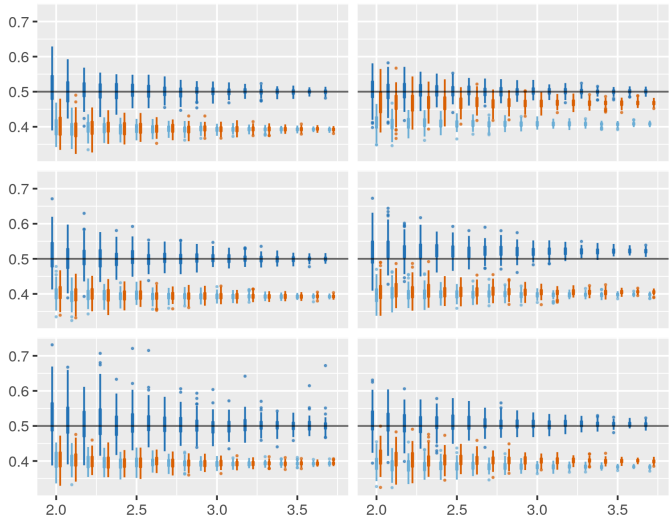

Figure 6. *First eigenvalue ratios as n increases: from Σ (in black), S (in blue), S_{rank} (in light blue), and S_{sptl} (in red); with $\mathcal{N}(\mu, \Sigma)$ (top left), t_{10} and t_4 (middle and bottom left), bimodal (top right), and multivariate skew normal S_{95} and S_{99} (middle and bottom right).*

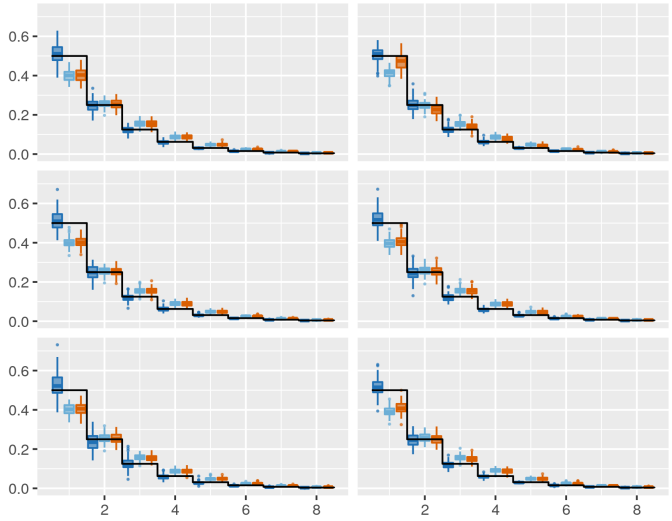

Figure 7. *Contribution of the first eight eigenvalues to total variance for $n = 100$ with the index of the eigenvalue on the x-axis: from Σ (in black), S (in blue), S_{rank} (in light blue), and S_{sptl} (in red); with $\mathcal{N}(\mu, \Sigma)$ (top left), t_{10} and t_4 (middle and bottom left), bimodal (top right), and multivariate skew normal S_{95} and S_{99} (middle and bottom right).*

Analogous graphs for the second and third eigenvalue ratios can be found in Figure A2 in the Appendix. These figure panels reveal that for all six distributions the mean ratios of the second and third contributions from the spherical covariances are larger than those of S and Σ. Recall that we noticed a similar pattern for the real data: the first spherical eigenvalue ratios as smaller than that of S, but the second and third ones are larger. As for the first eigenvalue ratios, the second and third ratios converge as n increases but not to the population value – with the exception of the values arising from S.

It may be of interest to see the pattern arising from smaller eigenvalues. Figure 7 shows these contributions with the eigenvalue index on the x-axis, here for $n = 100$. Similar graphs arise for larger values of n with the main difference

that the ranges of the boxplots decrease with increasing value of n.

Figure 7 provides a clear picture of the contributions to variance arising from the three sample covariances S, S_{rank} and S_{sptl}. For all six distributions the first contribution of S_{rank} and S_{sptl} are smaller than those of Σ, while the second and all later ones exceed those of Σ. Analogous results for S_{spear} and S_τ reveal that they are very similar to those of S_{rank} and are therefore not shown. We observe that S_{rank} and S_{sptl} show different eigenvalue ratios for the bimodal distribution. Because of the similarity of the eigenvalue patterns of S_τ and S_{rank} we may conclude that S_τ behaves more like S_{rank} and may not be a good approximation to S_{sptl} especially for non-elliptic models.

Our calculations and figures indicate that the normalised contributions to variance of S converge to those of Σ, while those of the spherical covariance matrices start smaller but then exceed those of Σ. Figure 2 and Figure A1 in the Appendix, tell us that the pattern for these three real datasets is similar to what we have seen for the simulated data: the first contribution to variance of spherical covariances is smaller than that of Σ.

5.4. Comparisons of Sample Eigenvectors as n Grows

For the real datasets of Section 4 we compared the first few eigenvectors and their PC scores calculated from different sample covariance matrices. In this section we look at similar comparisons for simulated datasets.

We focus on the six distributions listed before Algorithm 1 in Section 5.2 and, as n grows, compare mean angles of pairs of sample eigenvectors from two different covariance matrices in order to study their variability across different distributions and sample covariances. Algorithm 2 describes the calculations which share some of the steps of Algorithm 1, however, Algorithm 2 only focusses on eigenvectors, since results relating to eigenvalues are covered in Section 5.3.

Algorithm 2. Estimators of population eigenvectors.

1. Fix the dimension d.
2. Fix the distribution F_κ for $\kappa \in \{1, \ldots, 6\}$.
   For sample parameters $(\widehat{\lambda}_j, \widehat{\eta}_j)$ and $q_{ss,.,j}$, fix $n = 10^\ell$, $\ell \in \{2,\ 2.1,\ 2.2, \ldots, 3.7\}$
3. For $k \leq 100$ generate $\mathbf{X}_k$, each of size n.
4. For $ss \leq 5$ calculate $S_{ss,k}$ and $\widehat{\eta}_{ss,k,j}$ for $j \leq d$.
5. Compare pairs of estimators: for $j \leq d$ do
6. For pairs (s', s'') from $ss = 1, \ldots, 5$
   $\widehat{\gamma}_{(s',s'',k,j)} \leftarrow \angle(\widehat{\eta}_{s',k,j}, \widehat{\eta}_{s'',k,j})$ for $k \leq 100$
   $\bar{\widehat{\gamma}}_{(s',s'',j)} \leftarrow \sum_k \widehat{\gamma}_{(s',s'',k,j)}/100$.
7. Return to 2 if $\kappa < 6$.

We carry out the calculations of Algorithm 2 for $d = 5, 10, 20$ and 50, and ℓ with $n = 10^\ell$ as in Algorithm 1. The calculations show that the results are very similar across the different dimensions, and we therefore present and discuss the results pertaining to $d = 20$ only.

For the population scenario in the previous section we focussed on the sample covariance matrices S, S_{sptl} and S_{rank}. We now look at convergence of eigenvectors from pairs of these sample matrices. This will provide complementary insight to that of the population results.

For $n = 100$ and 5011, the values shown in the graphs are listed in Table 6. The more detailed Table A4, which also shows results for $n = 501$ and 1000, can be found in the Appendix.

Figure 8 shows the graphs of the pairwise comparisons. Angles pertaining to (S, S_{rank}) are shown in light blue, those of (S, S_{sptl}) in red, and those of $(S_{\text{rank}}, S_{\text{sptl}})$ in blue. The arrangement of the panels of distributions is the same as in Figure 5. Vertical boxplots show the variation of the 100 simulations in the same colours as the mean angles.

Note that for all values of n, the angles between eigenvectors of S_{sptl} and S, and between S_{rank} and S are smaller than the corresponding angles are between these two matrices and Σ, as shown in Figure 5. The largest angles between S and the spherical matrices S_{sptl} and S_{rank} occurred for the t_4 distribution. The main reason for this bigger discrepancy is that S is expected to perform poorly for t_4 data. Apart from the t_4 results, the eigenvectors of S and S_{rank} form smaller angle than either of these does with S_{sptl}. This suggests that S_{sptl} is not as good a substitute for S as S_{rank}.

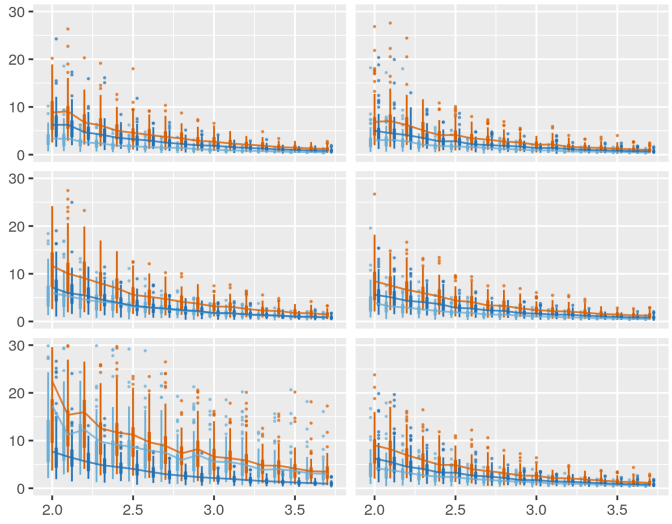

Figure 8. *Mean angles of Step 6, Algorithm 2, for (S, S_{rank}) (light blue), (S, S_{sptl}) (red) and (S_{rank}, S_{sptl}) (blue) as functions of n; with $\mathcal{N}(\mu, \Sigma)$ (top left), t_{10} and t_4 (middle and bottom left), bimodal (top right), and multivariate skew normal S_{95} and S_{99} (middle and bottom right).*

Table 6. *Mean angles corresponding to Figure 8.*

n	S_{ss1}, S_{ss2}	$\mathcal{N}$	t_{10}	t_4	B	S_{95}	S_{99}
100	S, S_{sptl}	8.85	11.64	22.46	6.87	8.42	8.99
	S, S_{rank}	3.36	5.98	17.30	2.98	3.86	4.23
	$S_{\text{rank}}, S_{\text{sptl}}$	6.22	7.13	7.70	5.00	5.59	6.24
5011	S, S_{sptl}	1.18	1.46	3.48	1.01	1.25	1.13
	S, S_{rank}	0.47	0.85	3.00	0.43	0.53	0.51
	$S_{\text{rank}}, S_{\text{sptl}}$	0.80	0.84	0.96	0.67	0.83	0.75

Our last comparison looks at mean angles of the eigenvectors of $S_{\text{rank}}, S_{\text{spear}}$ and S_τ. In Section 5.2 we only considered the convergence of S_{rank} to the population quantity.

If the eigenvectors of the three pairs of covariance are all very similar, then this step is justified, and S_{rank} can indeed be regarded as a good representative of all three.

Figure 9 shows graphs of mean angles of eigenvectors from pairs of covariance matrices as a function of n: $(S_{\text{rank}}, S_{\text{spear}})$ in light blue, $(S_{\text{rank}}, S_\tau)$ in red, and $(S_{\text{spear}}, S_\tau)$ in blue. Note that the red line is barely visible as it is directly below the blue line. The left panels feature the multivariate normal and the multivariate t-distributions in 10 and 4 degrees of freedom, and the right panels feature the bimodal mixture distribution and the multivariate skew normal distributions with $\rho = 0.95$ and 0.99.

For $n = 100$ and 5011, the values shown in the graphs are listed in Table 7. The more detailed Table A5, which also contains results for $n = 501$ and 1000, can be found in the Appendix.

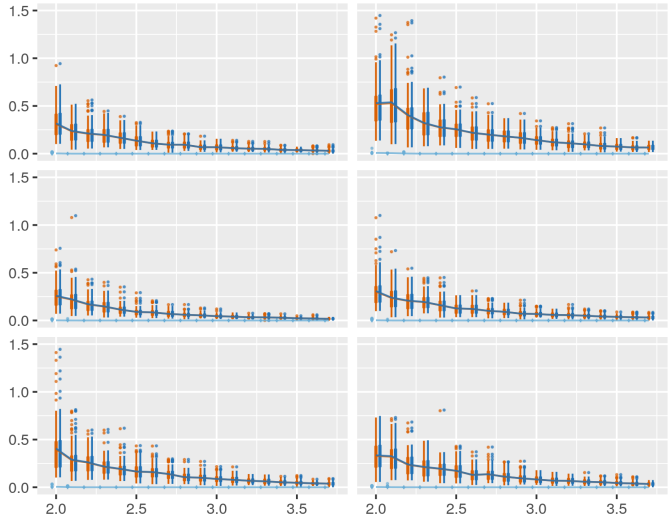

Figure 9. *Mean angles of Step 6, Algorithm 2, from (S_{rank}, S_{spear}) (light blue), (S_{rank}, S_τ) (red), and (S_{spear}, S_τ) (blue) as functions of n; with $\mathcal{N}(\mu, \Sigma)$ (top left), t_{10} and t_4 (middle and bottom left), bimodal (top right), and multivariate skew normal S_{95} and S_{99} (middle and bottom right).*

Table 7. *Mean angles corresponding to Figure 9.*

n	S_{ss}	$\mathcal{N}$	t_{10}	t_4	BMN	S_{95}	S_{99}
100	$S_{\text{rank}}, S_{\text{spear}}$	0.00	0.00	0.01	0.01	0.00	0.00
	S_{rank}, S_τ	0.31	0.25	0.40	0.52	0.30	0.33
	S_{spear}, S_τ	0.32	0.26	0.41	0.53	0.31	0.33
5011	$S_{\text{rank}}, S_{\text{spear}}$	0.00	0.00	0.00	0.00	0.00	0.00
	S_{rank}, S_τ	0.03	0.02	0.04	0.06	0.03	0.03
	S_{spear}, S_τ	0.03	0.02	0.04	0.06	0.03	0.03

Observe that the largest mean angle at $n = 100$, so $\ell = 2$ in the graphs, is 0.53 degrees and occurs between S_{spear} and S_τ. This angle is comparable to the smallest mean angle of S and S_{rank} which occurs for $n = 5011$, see Table 6. Note that this angle is about one third of the smallest angles between the eigenvectors of S_{rank} and the population eigenvectors as can be seen in Table 5. Further, as n increases, the angles between the spherical covariances considered in Table 7 decrease quickly from their values at $n = 100$.

For all six distributions the difference in the angles between S_{rank} and S_{spear} is noticeably smaller than that of S_τ with either of these two. Indeed, for $n \geq 500$, the mean angle of the eigenvectors of S_{rank} and S_{spear} is basically zero. We will return to this point in the next section.

5.5. Comparisons of Simulation Results and Theory

The simulations of Section 5.2 compare the first population eigenvector with the first eigenvector of the different S_{ss} as n increases. The results presented in Figure 5 and Table 5 show that for all six distributions and the three different covariances matrices S, S_{sptl}, S_{rank}, the mean angles of the population and sample eigenvectors converge to zero as n increases. The largest decrease in the angles occurs between $100 \leq n \leq 500$ across all six distributions, after which the angles decrease more slowly.

For the three elliptic distributions displayed in the left panels of Figure 5, Theorem 2.2 tells us that the population eigenvectors of the three covariance matrices Σ, Σ_{sptl} and Σ_{rank} agree and hence, by Theorem 3.1, their sample eigenvectors converge to those of Σ. Our simulations confirm this convergence result for all three S_{ss}, however, convergence of the first eigenvector of S is much slower for t_4. This behaviour of S agrees with knowledge that S is sensitive to outliers and heavy-tailed data. It may be encouraging to observe that, for S_{rank}, convergence of the sample eigenvectors to the eigenvector of Σ is almost as good for the non-elliptic distributions as for the elliptic ones. This suggests that the corresponding theoretical results may also hold for (at least some) non-elliptic distributions.

Based on the simulation results, the eigenvectors of S_{rank} are better estimators for those of Σ than the eigenvectors based on any of the other covariance matrices, including S. For the distributions we considered here, S_{rank}-based PCA seems to perform equally well for elliptic and non-elliptic distributions.

The results on eigenvalues are not as easily interpretable. Theorem 2.3 relates the eigenvalues of Σ_{sptl} and Σ_τ to those of Σ, and provides bounds for the population eigenvalues under the elliptic model. For the exponentially decreasing eigenvalues we used in the simulations, part 3 of Theorem 2.3 does not yield useful bounds, so is not directly relevant.

Our simulation results relating to the eigenvalues of S_{rank} strongly suggest that its eigenvalues converge as the sample size increases, and one would expect that they converge to the population value. To prove such a result is beyond the scope of this paper, but would be interesting to examine in more detail.

Returning to Theorem 2.2, we know that $\Sigma_{\text{sptl}} = \Sigma_\tau$ for elliptic distributions. The simulation results of Figure 6 and the comments following Figure 6 tell us that for the three elliptic models the sample eigenvalues are essentially the same, however, for the non-elliptic bimodal model, the eigenvalues of the two sample covariance matrices clearly differ. These examples indicate that care is required when one wants to use S_τ as an estimator of Σ_{sptl} as S. Taskinen, I. Koch, and H. Oja [2] suggest.

The theoretical developments of spherical covariance matrices and spherical PCA have focussed more on S_{sptl} than

on the other S_{ss}, most likely because S_{sptl} is the '*natural*' covariance matrix for spatial signed or directional data. Here we want to draw the reader's attention to one of the potential pitfalls in working with this natural covariance matrix S_{sptl}. By Theorem 2.2 and (19), the (sample) spatial median agrees with the (sample) mean under the elliptic model. For non-elliptic data the sample mean can differ from the sample spatial median. We have used the sample mean in all simulations with S_{sptl}. Our results show that this possibly 'incorrect' centring quantity does not have a serious adverse effect on the asymptotic behaviour of the eigenvectors of S_{sptl}: the first sample eigenvector of S_{sptl} converges across the elliptic and non-elliptic distributions to that of Σ at about the same rate. However, caution in the choice of the centring parameter may be required as the calculations of the wine data of Section 4.3 show. The graphs and angles presented in Figure 8 and Table 6 are consistent with the population results: For the elliptic distributions—with the exception of t_4—the sample estimators derived from S and S_{rank} converge quickly to each other and, indeed, more quickly than the individual estimators converge to the population eigenvectors. The eigenvector of S and S_{rank} are closer to each other than either of them to S_{sptl}. This suggests that S_{rank} is a better substitute for S than S_{sptl}, and a replacement in the presence of distributions or data for which S is known to behave poorly.

Surprisingly, for the three non-elliptic distributions the convergence of the eigenvector of S and S_{rank} to each other is even faster than for the elliptic distributions. We do not claim that this result holds in general; our results are based on the bimodal and multivariate skew normal distributions for up to 50 dimensions and sample sizes up to $n = 10^{3.7}$, and it is beyond the scope of this paper to make more general claims. However, the convergence results of these eigenvectors for elliptic and non-elliptic distributions lend support to regarding S_{rank} as a serious contender to S in PCA, both in terms of eigenvector directions and feature selection.

Our final comments relate to the sample covariances $S_{\text{rank}}, S_{\text{spear}}$ and S_τ and (24). This equation shows that, asymptotically, S_{rank} and S_{spear} are the same and their difference is of order $(1/n)$. A comparison with the results presented in Figure 9 and Table 7 confirms that the eigenvectors of these covariance matrices are basically the same for all six distributions. The results also confirm that the eigenvectors of S_τ converge to either of these quickly as n grows. Furthermore, the analysis of the real and simulated data shows that the contributions to variance of these three matrices are essentially the same. For practical consideration this therefore suggests that it suffices to work with S_{rank}.

6. Conclusion

We considered nonlinear covariance matrices, and more specifically spherical covariance matrices which arise from spatial sign or direction vectors, and we compared their eigenvalues, eigenvectors, and resulting PC scores to those of the standard covariance matrix. These spherical covariance matrices give rise to nonlinear principal component analysis based on natural concepts with interesting and interpretable properties. Seen through the different lenses of the population, the random sample and the data, the rank covariance matrix can be regarded as a serious contender to the standard covariance matrix: Under the elliptic model the eigenvectors of Σ_{rank} agree with those of Σ and for the random sample, the eigenvectors of S_{rank} are asymptotically consistent estimators.

As we deviate from the elliptic model, our findings, based on contrasting real data and simulated datasets from a variety of distributions, show that the rank covariance matrix leads to eigenvectors that are often similar to those of S. However, where S is known to perform poorly, the rank covariance matrix continues to perform well; this includes data with outliers, and data from heavy-tailed distributions. Regarded as estimators, the eigenvectors of the rank covariance matrix appear to converge at similar rates for elliptic and non-elliptic distributions and, as n increases, these eigenvectors show fast convergence to the population quantities.

We have focussed more on eigenvectors than eigenvalues of S_{rank} (or any of the other spherical S_{ss} we included in this paper), since the eigenvectors of S_{rank} appear to be good estimators of those of Σ. The same is not true for the eigenvalues of S_{rank}. Indeed, our results indicate that the first or first few percentage contributions to total variance of S_{rank} are smaller than those of S or Σ. To use S_{rank} as a substitute for S in PCA, it would be valuable to evaluate the contribution of each eigenvalue to total variance and to examine the asymptotic behaviour of the eigenvalues in more detail. This will have to be a the topic of future research.

Returning to eigenvectors, the results show that S_{spear} and S_{rank} are essentially the same, and Kendall's τ covariance matrix S_τ is similar to S_{rank}, but the latter is conceptually easier and seems to enjoy better performance as an estimator.

Theoretical results and calculations based on the 'natural' covariance matrix for direction vectors, the spatial sign covariance matrix, establish that its population eigenvectors agree with those of Σ under the elliptic model. This can be traced back to the fact that the sample mean agrees with the sample spatial median for these distributions. Our simulations, however, show that convergence of these sample eigenvectors to the population eigenvectors is slower than similar results based on the standard or the rank covariance matrices—even for elliptic models. Our results suggest that a good choice of the centring parameter for non-elliptic data may be difficult, and the sign covariance matrix may result in unstable or unpredictable eigenvectors with, at times, highly correlated principal component scores. These features detract from its use and usefulness for real data.

In summary, the rank covariance matrix S_{rank} has emerged as a valuable tool for estimating population quantities and for calculating principal components and features in data. Its performance is very similar to that of S when the latter works well, and it continues to lead to good results when S is not suitable. Its use in the analysis of HDLSS data is likely to highlight its superior performance over S further and could render the rank covariance matrix an essential tool for dimension reduction and feature selection.

Appendix

Appendix 1. Additional Tables and Figure for the Athletes and Wine Data

Tables A1 and A2 contain and extend information provided in tables in Section 4.3. Figure A1 complements the information in Table A2.

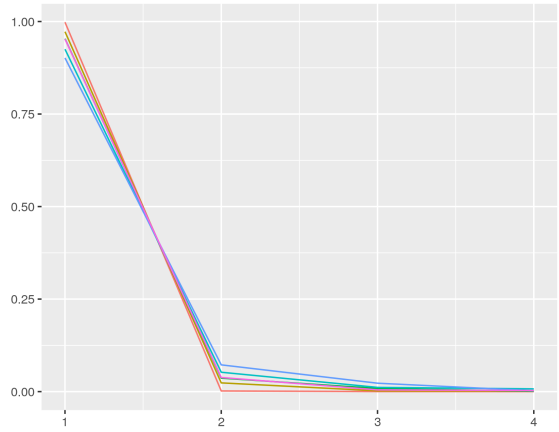

Figure A1. *Eigenvalue ratios of the first 4 eigenvalues for the wine data using different covariance matrices: S red, S_{rank} olive, S_τ violet, S_{sptl} green, $S_{sptl,1}$ blue-green and $S_{sptl,2}$ blue corresponding to Table 3 in Section 4.3.*

Table A1. *Athletes data: angles of first and second eigenvectors of sample covariance matrices in odd and even rows respectively–extending Table 2 in Section 4.3.*

	$\boldsymbol{S}_{\text{sptl}}$	$\boldsymbol{S}_{\text{rank}}$	$\boldsymbol{S}_{\boldsymbol{\tau}}$	$\boldsymbol{S}_{\text{spear}}$	$\boldsymbol{S}_{\text{sptl,1}}$	$\boldsymbol{S}_{\text{sptl,2}}$
S	7.46	4.13	3.61	4.13	8.80	10.16
	8.63	2.99	2.34	3.00	8.42	10.26
S_{sptl}		3.35	3.86	3.34	1.52	3.00
		6.22	6.97	6.21	0.84	2.91
S_{rank}	.35		0.55	0	4.74	6.06
	6.22		0.78	0	5.86	7.42
S_τ	3.86	0.55		0.55	5.21	6.60
	6.97	0.78		0.79	6.63	8.20
S_{spear}	3.34	0	0.55		4.73	6.06
	6.21	0	0.79		5.85	7.41
$S_{sptl,1}$	1.52	4.74	5.21	4.73		2.49
	0.84	5.86	6.63	5.85		8.42

Table A2. *Wine data: angles of first and second eigenvectors of sample covariance matrices in odd and even rows respectively–extending Table 3 in Section 4.3.*

	$\boldsymbol{S}_{\text{sptl}}$	$\boldsymbol{S}_{\text{rank}}$	$\boldsymbol{S}_{\boldsymbol{\tau}}$	$\boldsymbol{S}_{\text{spear}}$	$\boldsymbol{S}_{\text{sptl,1}}$	$\boldsymbol{S}_{\text{sptl,2}}$
S	1.49	0.39	0.42	0.39	1.65	3.14
	3.54	1.09	1.42	1.08	7.08	71.11
S_{sptl}		1.13	1.11	1.13	0.37	2.51
		3.67	3.93	3.66	7.32	71.60
S_{rank}	1.13		0.04	0	1.30	2.94
	3.67		0.43	0.01	6.60	71.02
S_τ	1.11	0.04		0.04	1.28	2.93
	3.93	0.43		0.44	6.56	70.92
S_{spear}!	1.13	0	0.04		1.30	2.94
	3.66	0.01	0.44		6.56	71.03
$S_{sptl,1}$	0.37	1.30	1.28	1.30		2.51
	7.32	6.60	6.60	6.56		70.75

Appendix 2. Additional Tables and Figures for the Simulations

Tables A3, A4 and A5 contain and extend information provided in tables in Sections 5. Figure A2 complements the first eigenvalue plots shown in Figure 6.

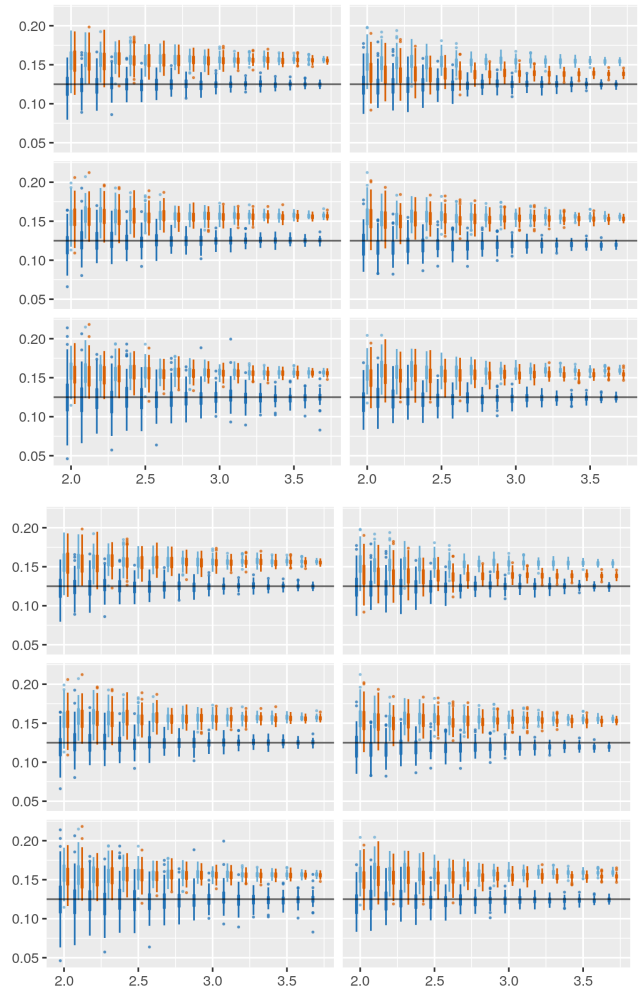

Figure A2. *Contribution of second eigenvalues (upper six) and third eigenvalues (lower six) to total variance as n increases: from Σ (in black), S (in blue), S_{rank} (in light blue), and S_{sptl} (in red); with $\mathcal{N}(\mu, \Sigma)$ (top left), t_{10} and t_4 (middle and bottom left), bimodal (top right), and multivariate skew normal S_{95} and S_{99} (middle and bottom right).*

Table A3. *Mean angles of first eigenvectors shown in Figure 5–extending Table 5 in Section 5.2.*

n	$\boldsymbol{S}_{ss}$	$\mathcal{N}$	t_{10}	t_4	BMN	MSN_{95}	MSN_{99}
100	S	12.05	14.10	23.91	12.31	12.95	11.81
	S_{sptl}	15.31	15.89	16.12	13.79	16.36	15.22
	S_{rank}	12.63	13.39	14.87	12.33	13.94	13.80
501	S	5.13	6.03	9.70	5.11	5.19	5.96
	S_{sptl}	6.28	6.16	6.29	5.48	5.94	6.94
	S_{rank}	5.25	5.58	5.87	5.23	5.42	6.40
1000	S	3.67	3.85	7.64	3.43	3.33	3.57
	S_{sptl}	4.39	4.28	5.02	3.77	3.97	4.18
	S_{rank}	3.74	3.62	4.51	3.46	3.52	3.82
5011	S	1.53	1.81	3.65	1.68	1.53	1.67
	S_{sptl}	1.88	1.87	1.89	1.82	1.97	2.00
	S_{rank}!	1.57	1.66	1.78	1.69	1.64	1.80

Table A4. *Mean angles of first eigenvectors shown in Figure 8–extending Table 6 in Section 5.4. In column 2, S_s and S_τ refer to S_{sptl} and S_{rank}.*

n	S_1, S_2	$\mathcal{N}$	t_{10}	t_4	BMN	MSN_{95}	MSN_{99}
100	S, S_s	8.85	11.64	22.46	6.87	8.42	8.99
	S, S_r	3.36	5.98	17.30	2.98	3.86	4.23
	S_r, S_s	6.22	7.13	7.70	5.00	5.59	6.24
501	S, S_s	3.71	4.73	8.91	3.14	3.36	3.61
	S, S_r	1.46	2.60	7.45	1.42	1.44	1.78
	S_r, S_s	2.53	2.73	2.95	2.02	2.33	2.47
1000	S, S_s	2.73	3.14	6.61	2.06	2.41	2.62
	S, S_r	1.03	1.65	5.57	0.89	1.08	1.26
	S_r, S_s	1.88	1.79	2.14	1.38	1.64	1.71
5011	S, S_s	1.18	1.46	3.48	1.01	1.25	1.13
	S, S_r	0.47	0.85	3.00	0.43	0.53	0.51
	S_r, S_s	0.80	0.84	0.96	0.67	0.83	0.75

Table A5. *Mean angles of first eigenvectors shown in Figure 9–extending Table 7 in Section 5.4. In column 2, S_r and S_s refer to S_{rank} and S_{spear}.*

n	S_1, S_2	$\mathcal{N}$	t_{10}	t_4	BMN	MSN_{95}	MSN_{99}
n	S_1, S_2	$\mathcal{N}$	t_{10}	t_4	BMN	MSN_{95}	MSN_{99}
100	S_r, S_s	0.00	0.00	0.01	0.01	0.00	0.00
	S_r, S_τ	0.31	0.25	0.40	0.52	0.30	0.33
	S_s, S_τ	0.32	0.26	0.41	0.53	0.31	0.33
501	S_r, S_s	0.00	0.00	0.00	0.00	0.00	0.00
	S_r, S_τ	0.09	0.07	0.14	0.20	0.10	0.14
	S_s, S_τ	0.09	0.07	0.14	0.20	0.10	0.14
1000	S_r, S_s	0.00	0.00	0.00	0.00	0.00	0.00
	S_r, S_τ	0.07	0.05	0.09	0.14	0.07	0.08
	S_s, S_τ	0.07	0.05	0.09	0.14	0.07	0.08
5011	S_{rank}, S_s	0.00	0.00	0.00	0.00	0.00	0.00
	S_r, S_τ	0.03	0.02	0.04	0.06	0.03	0.03
	S_s, S_τ	0.03	0.02	0.04	0.06	0.03	0.03

References

[1] J. I. Marden (1999). "Some robust estimates of principal components," Statistics & probability letters, vol. 43, no. 4, 349-359.

[2] S. Taskinen, I. Koch, and H. Oja (2012). "Robustifying principal component analysis with spatial sign vectors," Statistics and Probability Letters, vol. 82, 765-774.

[3] S. Visuri, V. Koivunen, and H. Oja (2000). "Sign and rank covariance matrices," Journal of Statistical Planning and Inference, vol. 91, no. 2, 557-575.

[4] A. Dürre, D. Vogel, and R. Fried (2015). "Spatial sign correlation," Journal of Multivariate Analysis, vol. 135, 89-105.

[5] D. Gervini (2008). "Robust functional estimation using the median and spherical principal components," Biometrika, vol. 95, no. 3, 587-600.

[6] C. Croux, E. Ollila, and H. Oja (2002). "Sign and rank covariance matrices: statistical properties and application to principal components analysis," in Statistical data analysis based on the L1-norm and related methods. Springer, 257-269.

[7] F. Han and H. Liu (2018). "Eca: High-dimensional elliptical component analysis in non-gaussian distributions," Journal of the American Statistical Association, vol. 113, no. 521, 252-268.

[8] A. Dürre, D. E. Tyler, and D. Vogel (2016). "On the eigenvalues of the spatial sign covariance matrix in more than two dimensions," Statistics & Probability Letters, vol. 111, 80-85.

[9] K. Yu, X. Dang, and Y. Chen (2015). "Robustness of the affine equivariant scatter estimator based on the spatial rank covariance matrix," Communications in Statistics-Theory and Methods, vol. 44, no. 5, 914-932.

[10] P. J. Huber (1981). Robust statistics. John Wiley & Sons.

[11] N. Locantore, J. Marron, D. Simpson, N. Tripoli, J. Zhang, K. Cohen, G. Boente, R. Fraiman, B. Brumback, C. Croux et al. (1999). "Robust principal component analysis for functional data," Test, vol. 8, no. 1, 1-73.

[12] S. Visuri, E. Ollila, V. Koivunen, J. Möttoönen, and H. Oja (2003). "Affine equivariant multivariate rank methods," Journal of Statistical Planning and Inference, vol. 114, no. 1-2, 161-185.

[13] A. Dürre, D. Vogel,and D. E. Tyler (2014). "The spatial sign covariance matrix with unknown location," Journal of Multivariate Analysis, vol. 130, 107-117.

[14] L. J. van't Veer, H. Dai, M. J. van de Vijver, Y. D. He, A. A. M. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards, and S. H. Friend (2002). "Gene expression profiling predicts clinical outcome of breast cancer," Nature, vol. 415, 530-536.

[15] R. D. Cook and S. Weisberg (1999). Applied Statistics Including Computing and Graphics. New York: Wiley.

[16] S. Aeberhard, D. Coomans, and O. de Vel (1992). "Comparison of classifiers in high dimensional settings," Tech. Rep. no. 92-02, Dept. of Computer Science and Dept of Mathematics and Statistics, James Cook University of North Queensland, data sets collected by Forina et al and available on http://www.kernel-machines.com/.

[17] A. Azzalini and A. Capitanio (1999). "Statistical applications of the multivariate skew normal distribution," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 61, no. 3, 579-602.