SIGN AND RANK COVARIANCE MATRICES

Samuli Visuri[*†]
Signal Processing Laboratory
Tampere Univ. of Technology
P.O. Box 553,
FIN-33101 Tampere, Finland

Visa Koivunen
Signal Processing Laboratory
Tampere Univ. of Technology, Finland

Hannu Oja
Dept of Statistics
Univ. of Jyväskylä, Finland

Abstract

The robust estimation of multivariate location and shape is one of the most challenging problems in statistics and crucial in many application areas. The objective is to find highly efficient, robust, computable and affine equivariant location and covariance matrix estimates. In this paper three different concepts of multivariate sign and rank are considered and their ability to carry information about the geometry of the underlying distribution (or data cloud) are discussed. New techniques for robust covariance matrix estimation based on different sign and rank concepts are proposed and algorithms for computing them outlined. In addition, new tools for evaluating the qualitative and quantitative robustness of a covariance estimator are proposed. The use of these tools is demonstrated on two rank based covariance matrix estimates. Finally, to illustrate the practical importance of the problem, a signal processing example where robust covariance matrix estimates are needed is given.

AMS classification: primary 62G35 ; secondary 62H12

Keywords: covariance matrix, multivariate sign and rank, influence function

1 INTRODUCTION

The mean vector $\boldsymbol{\mu}$ and covariance matrix Σ are natural parameters in the general p-variate normal case. If the observations $\mathbf{x}_1, .., \mathbf{x}_n$ come from a $N_p(\boldsymbol{\mu}, \Sigma)$-distribution, the sample mean vector $\bar{\mathbf{x}}$ and the sample covariance matrix

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

*Corresponding author. E-mail: svisuri@cs.tut.fi
†Supported by the Academy of Finland.

are minimum variance unbiased estimators of $\boldsymbol{\mu}$ and Σ. Multivariate statistical methods are often based on the sample covariance or correlation matrix. As an example, the inference about the means (MANOVA) compares "between samples" and "within samples" sum of squares matrices B and W, respectively ($(n-1)S = B + W$). Other methods where covariance matrices are employed include principal component analysis (PCA), multivariate (multiple) regression, canonical correlation analysis, factor analysis and discriminant analysis. The usual sample covariance matrix S and consequently the standard multivariate techniques based on it are, however, highly sensitive to outlying observations.

Many robust techniques have been proposed for estimating multivariate location and scatter. In the following, main approaches in robust covariance matrix estimation are briefly listed. The affine equivariant multivariate location and scale M-estimators (maximum likelihood estimates as special cases) were proposed by Maronna. See Maronna (1976), Huber (1977), Campbell (1980) and Kent and Tyler (1991) for M-estimates and Davies (1987), Lopuhaä (1989) and Lopuhaä and Rousseeuw (1991) for so called S-estimates. The minimum volume ellipsoid (MVE) estimator (see Rousseeuw and Leroy (1987)) is given by the smallest regular ellipsoid covering at least $h = [n/2] + 1$ observations: The MVE location statistic is then the midpoint of that ellipsoid and the MVE scatter matrix is given by its shape. The related minimum covariance determinant (MCD) scatter statistic with higher asymptotic efficiency is given by the half sample covariance matrix with minimum determinant (Rousseeuw, 1984). The Stahel-Donoho statistics (Stahel, 1981; Donoho, 1982) are obtained by taking a weighted mean vector and covariance matrix which downweights outlying data points. They proposed a projection pursuit based method as well that uses robust univariate location and scale statistics. For still another projection pursuit based approach see Huber (1981), Li and Chen (1985) and Croux and Ruiz-Gazen (1997). These methods first compute robust estimate of the first eigenvector by finding the direction which yields the maximum univariate robust estimate of variance. The direction of the second eigenvector is orthogonal to the first one and again yields the maximum robust univariate variance estimate. The remaining eigenvalues and eigenvectors are found sequentially in similar manner.

Covariance and correlation matrices are of great importance in many signal processing application areas such as image processing, telecommunications and control systems. As a practical example, covariance or correlation matrix estimates are commonly needed in attenuating noise in multichannel signals such as EEG or color images (Koivunen and Kassam, 1996). These techniques are typically derived assuming that the underlying statistical model is multinormal (Gaussian). In practise, however, this assumption is often inaccurate. Nominally optimal estimators typically suffer from a drastic degradation in performance even for small deviations from nominal conditions. Therefore, there is a growing interest in robust covariance estimators in the signal processing research community. Hence, it is important to study how covariance estimators are influenced by outliers. Similar robustness study is of interest in the estimation of autocorrelation or autocovariance matrices in time series analysis.

This paper is organized as follows. Covariance matrix is a natural statistic in describing the scatter and dependence of an elliptically symmetric distribution. Therefore, in Section 2, basic concepts (location, scale, shape and orientation) needed in describing elliptically symmetric models and their geometry are considered. Further, tools for analysing the distinct properties of covariance matrix estimators are introduced. These tools include sensitivity plots that characterize qualitative robustness similarly to the influence function (Hampel et al., 1986). Three different concepts of sign and rank covariance matrices are discussed and their use in scatter matrix estimation is proposed in Section 3. As an illustration, qualitative and quantitative robustness as well as finite sample efficiencies of two proposed methods (spatial Kendall's tau and Oja rank correlation based methods) are studied in Section 4 using the tools proposed in section 2. A signal processing example where robust covariance matrix estimates are needed is described in Section 5. The paper ends with final comments in Section 6.

2 MODELS AND ROBUSTNESS CONSIDERATIONS

2.1 Elliptic models

The covariance matrix is a natural scatter statistic in the multivariate normal case. Is the (theoretical or sample) covariance matrix then a reasonable statistic also in non-normal models? In the case of the so called elliptic models the answer is affirmative. In order to see this, recall the concept of spherical symmetry. The distribution of $\mathbf{x} = (x_1, ..., x_p)^T$ is **spherical** or spherically symmetric with symmetry centre $\boldsymbol{\mu}$ if the p.d.f. of $\mathbf{x}$ can be represented in the form

$$f(\mathbf{x}) = f(x_1, ..., x_p) = f_0(|\mathbf{x} - \boldsymbol{\mu}|),$$

that is, it depends on $\mathbf{x}$ only through its Euclidean distance from $\boldsymbol{\mu}$. Write $\mathbf{x} = \boldsymbol{\mu} + r\mathbf{v}$, where $r = |\mathbf{x} - \boldsymbol{\mu}| = [(x_1 - \mu_1)^2 + ... + (x_p - \mu_p)^2]^{1/2}$ and $\mathbf{v} = |\mathbf{x} - \boldsymbol{\mu}|^{-1}(\mathbf{x} - \boldsymbol{\mu})$ are the length and the direction of $\mathbf{x} - \boldsymbol{\mu}$, r and $\mathbf{v}$ are independent and $\mathbf{v}$ is uniformly distributed on the periphery of the unit sphere $\mathcal{S}_p$. (For the p-variate spherical normal case, choose $r^2 \sim \chi^2_p$.) The components of $\mathbf{x}$ are uncorrelated and the covariance matrix of $\mathbf{x}$ is

$$E((\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T) = E(r^2)E(\mathbf{v}\mathbf{v}^T) = \frac{1}{p}E(r^2)I_p.$$

In the following, assume the covariance matrix determined by f_0 is the identity matrix I_p ($E(r^2) = p$). The distribution of $\mathbf{x}$ is **elliptic** or elliptically symmetric with symmetry center $\boldsymbol{\mu}$ and covariance matrix $\Sigma \in PDS(p)$ (the set of positive definite symmetric $p \times p$ matrices) if its p.d.f. is of the form

$$f(\mathbf{x}) = f(x_1, ..., x_p) = |det(L)| f_0(|L(x - \boldsymbol{\mu})|),$$

where L is lower triangular matrix with positive diagonal elements and $L^T L = \Sigma^{-1}$ (the Cholesky factorization).

Bensmail and Celeux (1996) presented eigenvalue decomposition for the (theoretical) covariance matrix in the form

$$\Sigma = \lambda\ U\ C\ U^T,$$

where U is the matrix of eigenvectors, C is a diagonal matrix with the normalized eigenvalues c_i on the diagonal ($\prod_{i=1}^p c_i = det(C) = 1$) and λ^p is the Wilks generalized variance. In passing, note that observations $\mathbf{x}$ from any elliptical distribution with symmetry centre $\boldsymbol{\mu}$ and covariance matrix $\Sigma = \lambda U C U^T$ can be generated in the following way:

1. Generate radius r from a 'standardized' distribution with $E(r^2) = p$.

2. Generate direction $\mathbf{v}$ from a uniform distribution on $\mathcal{S}_p$.

3. Set $\mathbf{x} = \boldsymbol{\mu} + U C^{1/2} \lambda^{1/2}\ r\ \mathbf{v}$.

Bensmail and Celeux (1996) use the terms **scale, shape** and **orientation** for items λ, C and U, for the successive steps in the above construction (stage 3). Diagonal matrix $\Lambda = \lambda C$ lists the usual eigenvalues. Wilks generalized variance, the determinant of the covariance matrix $Det(\Sigma) = Det(\Lambda)$, is just the geometrical mean of the eigenvalues to the power of p. Another 'global' measure of the multivariate scatter is the sum of eigenvalues $trace(\Sigma) = trace(\Lambda)$.

The symmetry centre $\boldsymbol{\mu}$ and covariance matrix Σ or the associated ellipsoid

$$\{x \in \Re^p \ : \ (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = d^2\} \quad = \quad \{x \in \Re^p \ : \ |\Lambda^{-1/2} U^T (\mathbf{x} - \boldsymbol{\mu})| = d\}$$

giving 'contours of equal depths' or 'tolerance contours' is thus a natural and sufficient way to illustrate the location, scale, shape and orientation in the elliptic case. Croux, Haesbroeck and Rousseeuw (1997) called UCU^T the 'shape matrix' because it determines the shape and orientation of the ellipsoid but not its magnitude.

Note that if in our construction,

$$x = UC^{1/2} \lambda^{1/2} r \mathbf{v}$$

the expectation $E(r^2)$ does not exist, then naturally the covariance matrix does not exist either. But also in this (elliptical) case orientation and shape are well defined concepts ($C^{-1/2} U^T$ spheres the distribution or the data) and can be estimated using sign and rank covariances. Note also that the standard distribution (f_0) may now be defined by $Med(r) = 1$ and for $\Sigma = \lambda UCU^T$ the associated ellipsoid

$$(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = 1$$

gives the 50 percent tolerance contour. In this construction the interpretation of the scale parameter λ has thus been changed.

2.2 Robustness in covariance estimation

Outliers may perturb covariance matrix estimates significantly. In particular, the variances may get inflated, the correlation structure may change significantly, insignificant dimensions may be added to the data and interesting structures in the data may remain unrevealed. Robust estimates of both the multivariate location parameter and covariance matrix are commonly needed in practical applications.

Many multivariate statistical techniques as well as tools for time series analysis rely heavily on covariance estimates. If unreliable covariance estimates are employed, the performance of the methods deteriorates as well. In the following we will characterize these perturbations in terms of the eigensystem of the covariance matrix. Eigenvalue analysis of the estimated covariance matrix may be used for analyzing the robustness. If Σ is the $p \times p$ covariance matrix, its eigenvector $\mathbf{u}$ and eigenvalue λ satisfy the following

$$\Sigma \mathbf{u} = \lambda \mathbf{u}$$

Suppose that Σ has p linearly independent eigenvectors. If these vectors are columns of U we have

$$U^T \Sigma U = \Lambda$$

and

$$\Sigma = U \Lambda U^T$$

where Λ is diagonal with the eigenvalues of Σ on the diagonal. Matrix U is orthogonal hence $U^T = U^{-1}$. Without a loss of generality, the eigenvalues and corresponding eigenvectors are often ordered such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. If eigenvectors correspond to different unequal eigenvalues then those eigenvectors are linearly independent. Eigenvalues are directly related to the variances and they reveal the significant orientations in the data.

The breakdown of the covariance matrix estimator $S = \lambda U C U^T = U \Lambda U^T$ may occur in several different ways. Here we list some possibilities.

1. The scale measured by Wilks generalized variance $Det(S)$ or by trace $trace(S)$ may increase over all bounds.

2. The condition number (shape),
$$Cond(S) = \frac{\lambda_1}{\lambda_p}$$
i.e, the ratio of the largest and smallest eigenvalue, may grow very large in the face of outliers. (This occurs when either the largest eigenvalue becomes very large or the smallest eigenvalue becomes infinitesimally small.)

3. Influential observations may change the 'ordered' eigenvectors and consequently drastically change the subspace spanned by the s first or s last columns of U.

4. Also the coordinate system for independent coordinates (given by orientation U) may change.

2.3 Tools for robustness studies

The influence function is a standard tool for characterizing the qualitative robustness of an estimator. The nominal distribution is perturbed by an infinitesimally small proportion of outliers and the amount of change in the estimator is observed. A robust estimator has an influence function that is bounded and continuous. Boundedness prevents an outlier from having an arbitrarily large influence on the estimate whereas continuity guarantees that small changes in data cause only small changes in the estimate. In the case of covariance estimation, we may consider changes in scale, shape and orientation as we perturb the nominal distribution by an infinitesimal proportion of multivariate outliers. The empirical influence function is obtained by using an empirical distribution and contaminating it with an outlier and varying the location of an outlier in p-dimensional space.

As stated earlier, changes caused by outliers to the estimated covariance matrix are conveniently described using the eigenvalue decomposition. The eigendecomposition of the perturbed covariance matrix $\tilde{\Sigma}$ is denoted by $\tilde{\Sigma} = \tilde{\lambda} \tilde{U} \tilde{C} \tilde{U}^T$. The changes in scale and shape may be described considering the changes in the product or the sum of the eigenvalues. In practise one could consider the ratios of the determinants $det(\tilde{\Sigma})/det(\Sigma)$ or traces $trace(\tilde{\Sigma})/trace(\Sigma)$ of the covariance matrices. Perturbations in the shape may be captured using the whole spectrum of eigenvalues and the matching distance metric (see Stewart and Sun, 1990):

$$md(\Sigma, \tilde{\Sigma}) = \min_{\tau} \max_{i} (|\tilde{\lambda}_{\tau_i} - \lambda_i|)$$

where $\tau = (\tau_1, \ldots, \tau_p)$ is taken over all permutations of $(1, 2, ..., p)$ and $\tilde{\lambda}_i$ are the eigenvalues of the perturbed covariance matrix. If perturbation is small the matching distance will be small and matching pairs of eigenvalues are clearly found. The change in the condition number tells how the eigenvalue spread has changed and how ill-conditioned the covariance matrix has become due to contamination. It also reveals if the data falls into a lower dimensional space. In addition, the change in the ratio of the geometric and arithmetic mean of the eigenvalues could be considered.

Perturbation in the orientation may be described in terms of the direction of the eigenvectors. One may investigate the perturbations in all the eigenvectors or consider the subspaces spanned by a subset of eigenvectors. Typically we are interested in the eigenvectors corresponding to either the s largest or s smallest eigenvalues of Σ. Let U_s be a subset of eigenvectors from U. Then $|det(\tilde{U}_s^T U_s)|$ may be used to quantify the change in the subspace spanned by s columns of U. This quantity approaches unity when the subspaces come perfectly aligned. A more intuitive quantity is perhaps obtained by describing the change in the basis vectors of the subspace in terms of the eigenvalues of $\tilde{U}_s^T U_s$. The canonical angles between the eigenvectors are obtained by $cos^{-1}(\lambda_i)$ where λ_i are the singular values of $\tilde{U}_s^T U_s$. Subspace U_s and perturbed subspace $\tilde{U}_s$ are close if the largest canonical angle is small. The rank orders of the matching eigenvectors may not necessarily indicate that a change in order has occurred. A quantity related to the matching distance metric and measuring the distances between coordinate systems is given using individual eigenvectors $\mathbf{u}_i$ by:

$$md(\tilde{U}, U) = \min_{\tau} \max_{i}(|cos^{-1}(\tilde{\mathbf{u}}_{\tau_i}^T \mathbf{u}_i)|)$$

where again τ is taken over all permutations of $(1, 2, ..., p)$.

The concept of the empirical influence function may be illustrated by using the concept of shape (the eigenvalue spread) as an example. Let $Cond(\Sigma)$ be the condition number of a covariance matrix. The change in the condition number of the estimated covariance matrix computed from the perturbed distribution compared to the covariance matrix estimated from the empirical distribution F_{n-1} may be illustrated by plotting

$$\mathcal{IF}(\mathbf{x}; F_{n-1}, \Sigma) = \frac{Cond(\tilde{\Sigma})}{Cond(\Sigma)}$$

where $\tilde{\Sigma}$ is the covariance matrix of the perturbed empirical distribution $(1 - \frac{1}{n})F_{n-1} + \frac{1}{n}\Delta_{\mathbf{x}}$ where $\Delta_{\mathbf{x}}$ denotes point mass one at the point $\mathbf{x}$ and represents a multivariate outlier. Similarly, the changes in orientation and scale may be plotted using either difference between the true and perturbed values or their ratios depending on which is more appropriate for the quantity of interest. Examples of this type of empirical influence functions will be provided later in this paper.

3 SIGN AND RANK COVARIANCES

3.1 Marginal signs and ranks

In this section, techniques for estimating covariance matrices using three distinct sign and rank concepts, namely (i) vectors of marginal signs and ranks, (ii) spatial signs and ranks and (iii) Oja signs and ranks are proposed. We now define sign and rank covariance matrices of the different types and consider their usefulness in the estimation of the shape matrix.

First, recall the univariate concepts. Let $\mathbf{x} = \{x_1, ..., x_n\}$ be a univariate data set. The univariate sign function $S(x) = sign(x)$ is $1, 0, -1$ as $x > 0, = 0, < 0$ and the centered rank function is $R(x) = \text{ave}\{S(x - x_i)\}$. Note that $R(x)$ is the derivative function of the criterion function for the median $D(x) = \text{ave}\{|x - x_i|\}$. For p-variate data set X, construct the vectors of marginal signs $\mathbf{S}_1(\mathbf{x}_i - \mathbf{M}_1(X))$ (w.r.t. the vector of marginal medians $\mathbf{M}_1(X)$) and marginal centered ranks $\mathbf{R}_1(\mathbf{x}_i)$ and construct the corresponding sign covariance matrix (SCM)

$$SCM_1 = \text{ave}\{\mathbf{S}_1(\mathbf{x}_i - \mathbf{M}_1(X))\mathbf{S}_1^T(\mathbf{x}_i - \mathbf{M}_1(X))\}$$

and Spearman's rank covariance matrix (RCM)

$$RCM_1 = \text{ave}\{\mathbf{R}_1(\mathbf{x}_i)\mathbf{R}_1^T(\mathbf{x}_i)\} = \text{ave}\{\mathbf{S}_1(\mathbf{x}_i - \mathbf{x_j})\mathbf{S}_1^T(\mathbf{x}_i - \mathbf{x}_k)\}.$$

(The marginal rank vector $\mathbf{R}_1(\mathbf{x})$ is the gradient of the criterion function for the vector of the marginal medians $D_1(\mathbf{x}) = \text{ave}\{|x_1 - x_{i1}| + ... + |x_p - x_{ip}|\}$.) The Kendall's tau covariance matrix is defined as

$$TCM_1 = \text{ave}\{\mathbf{S}_1(\mathbf{x}_i - \mathbf{x}_j)\mathbf{S}_1^T(\mathbf{x}_i - \mathbf{x}_j)\}.$$

The marginal sign and rank covariance matrices are scale invariant (rescaling the coordinates does not change the values of the matrices) which means that the original geometry and shape information (eigenvalues and eigenvectors) have been lost. These sign and rank covariance matrices are not rotation equivariant either. Despite this some information remains. Consider a p-variate elliptic distribution: all the bivariate marginal distributions are also elliptic. Assume now that the bivariate random variable $(x, y)^T$ is elliptically symmetric (w.r.t. the origin) with correlation coefficient ρ and write $(x_1, y_1)^T, (x_2, y_2)^T, (x_3, y_3)^T$ for three independent copies of $(x, y)^T$. Then for the theoretical sign correlation and theoretical Kendall's tau it is true that

$$E(sign(x)sign(y)) = E(sign(x_1 - x_2)sign(y_1 - y_2)) = \frac{2}{\pi}sin^{-1}(\rho),$$

since the expectation $E(sign(rv_1)sign(rv_2))$ does not depend on the distribution of the radius r. Note that $(x_1 - x_2, y_1 - y_2)^T$ is elliptic with the same correlation structure as $(x, y)^T$. For the normal case see Kendall (1975). If $(x, y)^T$ is bivariate normal, the (theoretical) Spearman's rho is

$$E(sign(x_1 - x_2)sign(y_1 - y_3)) = \frac{6}{\pi}sin^{-1}(\rho/2).$$

See Kendall (1975) and Capéraà and Genest (1993). The transformed sign and rank correlation estimates then converge in probability to the right correlation value (under elliptic and multinormal cases, respectively).

We do not know, however, whether the resulting $p \times p$ correlation matrix (after the transformation) is positive definite or not. One can of course circumvent the problem and first estimate robustly the partial correlations. The correlation estimates may then be calculated from estimated partial correlations. Masarotto (1987) proposed this technique for robust autocorrelation estimation. Yet another possibility is to replace the marginal ranks by corresponding expected normal scores (or van der Waerden scores) and calculate the covariance matrix using these resulting multivariate pseudo observations. See Gibbons and Chakraborti (1992), for example.

Based on the above discussion the following procedure for the robust covariance matrix estimation seems worth of considering in the elliptic (multinormal) case:

A1. Estimate correlations (robustly) with quadrant correlation (matrix SCM_1) or Kendall's τ (matrix TCM_1) (or Spearman's ρ, matrix RCM_1). Transform the correlations and write R for this transformed correlation matrix.

A2. Estimate the standard deviations using any univariate robust estimate (MAD with consistency correction, etc.). Write $H = diag(h_1, ..., h_p)$ for the estimated standard deviations.

A3. The covariance matrix estimate is $S = HRH^T$.

Another (computationally much more intensive) possibility would be first to try to find a coordinate system with independent marginal signs or ranks:

B1. Construct eigenvector (or principal vector) estimates in the following way: Find orthogonal U such that signs (ranks, correspondingly) of

$$\mathbf{U}^T\mathbf{x}_1,\ U^T\mathbf{x}_2,\ ...,\ U^T\mathbf{x}_n$$

are uncorrelated.

B2. Estimate the marginal variances (eigenvalues, principal values) of

$$\mathbf{U}^T\mathbf{x}_1,\ U^T\mathbf{x}_2,\ ...,\ U^T\mathbf{x}_n$$

using any univariate robust scale estimate with consistency correction (MAD, etc.). Write $\Lambda = diag(\lambda_1, ..., \lambda_p)$ for the estimates.

B3. The covariance matrix estimate is

$$S = U\Lambda U^T.$$

3.2 Spatial signs and ranks

The so called spatial sign is the function $\mathbf{S}_2(\mathbf{x}) = |\mathbf{x}|^{-1}\mathbf{x}$, that is, the unit vector in the direction of $\mathbf{x}$. The spatial rank function is defined as

$$\mathbf{R}_2(\mathbf{x}) = \text{ave}\{\mathbf{S}_2(\mathbf{x} - \mathbf{x}_i)\}.$$

Note that $\mathbf{R}_2(\mathbf{x})$ is the gradient of the criterion function $D_2(\mathbf{x}) = \text{ave}\{|\mathbf{x} - \mathbf{x}_i|\}$. The so called spatial median $\mathbf{M}_2(X)$ minimizes $D_2(\mathbf{x})$ or is the solution of equation $\mathbf{R}_2(\mathbf{x}) = \mathbf{0}$. See Chaudhuri (1996), Marden (1999a) and Möttönen and Oja (1995). Consider the observed spatial signs (w.r.t. the spatial median) and spatial ranks and find the corresponding sign and rank covariance matrices (SCM and RCM) (Oja and Koivunen, 1998; Marden, 1999b)

$$SCM_2 = \text{ave}\{\mathbf{S}_2(\mathbf{x}_i - \mathbf{M}_2(X))\mathbf{S}_2^T(\mathbf{x}_i - \mathbf{M}_2(X))\}$$

and

$$RCM_2 = \text{ave}\{\mathbf{R}_2(\mathbf{x}_i)\mathbf{R}_2^T(\mathbf{x}_i)\} = \text{ave}\{\mathbf{S}_2(\mathbf{x}_i - \mathbf{x}_j)\mathbf{S}_2^T(\mathbf{x}_i - \mathbf{x}_k)\}.$$

The spatial Kendall's tau covariance matrix can be defined as

$$TCM_2 = \text{ave}\{\mathbf{S}_2(\mathbf{x}_i - \mathbf{x}_j)\mathbf{S}_2^T(\mathbf{x}_i - \mathbf{x}_j)\}.$$

For a different multivariate version of Kendall's tau, see Choi and Marden (1998). These three matrices are rotation but not scale equivariant.

Assume that $\mathbf{x}_1, ..., \mathbf{x}_n$ is a random sample from a multivariate distribution with c.d.f. F. The spatial rank vector $\mathbf{R}_2(\mathbf{x})$ then converges (uniformly in probability) to a theoretical rank vector $\mathbf{R}$ (Möttönen and Oja, 1995) which naturally depends on the distribution F. In fact, there is a one-to-one correspondence between F and $\mathbf{R}$. (Koltchinskii, 1997). The theoretical spatial sign and rank covariance matrices for a multivariate cdf F are again defined as

$$E(\mathbf{S}_2(\mathbf{x}_1 - \boldsymbol{\mu})\mathbf{S}_2^T(\mathbf{x}_1 - \boldsymbol{\mu})), \quad E(\mathbf{S}_2(\mathbf{x}_1 - \mathbf{x}_2)\mathbf{S}_2^T(\mathbf{x}_1 - \mathbf{x}_3)) \quad \text{and} \quad E(\mathbf{S}_2(\mathbf{x}_1 - \mathbf{x}_2)\mathbf{S}_2^T(\mathbf{x}_1 - \mathbf{x}_2))$$

where $\mathbf{x}_1, \mathbf{x}_2$ and $\mathbf{x}_3$ are independent copies from a distribution F with the spatial median $\boldsymbol{\mu}$. Note that if F is elliptically symmetric, then the spatial sign and Kendall's tau covariance matrices coincide.

All the correlation estimates are U-statistics and therefore under general conditions converge in probability to the corresponding theoretical value with limiting normal distribution. For elliptically symmetric distributions the eigenvectors of the theoretical SCM and RCM are the same as the eigenvectors of the ordinary covariance matrix (Marden, 1998). There is also a one-to-one (but quite complicated) correspondence between the standardized eigenvalues of the theoretical spatial sign and Kendall's tau covariance matrix and the eigenvalues of the usual covariance matrix. Detailed results will be reported in forthcoming papers. In the bivariate case, for example, the condition number of the ordinary covariance matrix is the squared condition number of the sign covariance matrix. This is utilized in the robustness and efficiency studies in the estimates considered later in Section 5.

The estimation strategy may be now as follows:

C1. Construct eigenvector estimates as follows: First calculate the SCM (SCM_2) or RCM (RCM_2, TCM_2) using spatial signs (w.r.t. the spatial median) or spatial ranks. Find the corresponding eigenvector estimates, that is, matrix U.

C2. Estimate the marginal variances (eigenvalues, principal values) of

$$U^T\mathbf{x}_1, \; U^T\mathbf{x}_2, \; ..., \; U^T\mathbf{x}_n$$

using any univariate robust scale estimate (MAD, etc.). Write $\Lambda = diag(\lambda_1, ..., \lambda_p)$ for the estimates.

C3. The covariance matrix estimate is

$$S = U\Lambda U^T.$$

As explained in this section, it is also possible to construct both eigenvector and standardized eigenvalue estimates ('shape and orientation'), say U and C, from matrices SCM_2 and TCM_2. To estimate the scale, use first the shape matrix UCU^T to construct robust Mahalanobis type distances from the spatial median $\mathbf{M}_2(X)$,

$$d_i \; = \; (\mathbf{x}_i - \mathbf{M}_2(X))^T \; UC^{-1}U^T \; (\mathbf{x}_i - \mathbf{M}_2(X)), \quad i = 1, ..., n.$$

Wilks generalized variance (λ^p) may then be estimated robustly by $h_p(Med(D))^p$ where the correction factor h_p can be selected to guarantee the convergence in the multinormal case.

3.3 Oja signs and ranks

Consider then the affine equivariant multivariate sign and rank concepts based on the Oja (1983) median. The volume of the **p-variate simplex** determined by $\mathbf{x}$ and p observations with indices $i_1 < i_2 < ... < i_p$ is

$$\frac{1}{p!} abs \left\{ det \begin{pmatrix} 1 & ... & 1 & 1 \\ \mathbf{x}_{i_1} & ... & \mathbf{x}_{i_p} & \mathbf{x} \end{pmatrix} \right\}.$$

Consider the objective functions

$$H_3(\mathbf{x}) = \text{ave} \left\{ abs \left\{ det \begin{pmatrix} 1 & 1 & ... & 1 & 1 \\ \mathbf{0} & \mathbf{x}_{i_1} & ... & \mathbf{x}_{i_{p-1}} & \mathbf{x} \end{pmatrix} \right\} \right\}$$

and

$$D_3(\mathbf{x}) = \text{ave} \left\{ abs \left\{ det \begin{pmatrix} 1 & ... & 1 & 1 \\ \mathbf{x}_{i_1} & ... & \mathbf{x}_{i_p} & \mathbf{x} \end{pmatrix} \right\} \right\}.$$

The Oja sign and rank functions, $\mathbf{S}_3(\mathbf{x})$ and $\mathbf{R}_3(\mathbf{x})$, are defined as the gradient functions of $H_3(\mathbf{x})$ and $D_3(\mathbf{x})$, that is,

$$\mathbf{S}_3(\mathbf{x}) = \nabla H_3(\mathbf{x}) \text{ and } \mathbf{R}_3(\mathbf{x}) = \nabla D_3(\mathbf{x}).$$

In the univariate case, usual univariate sign and rank functions are obtained. The affine equivariant Oja median (1983), $\mathbf{M}_3(X)$, minimizes $D_3(\mathbf{x})$ or is the solution of $\mathbf{R}_3(\mathbf{x}) = \mathbf{0}$. For affine equivariant sign and rank methods, see Oja (1998) and references therein.

To estimate the multivariate shape matrix, we construct again the sign covariance matrix SCM_3 (using residuals w.r.t. the Oja median) and the rank covariance matrix RCM_3. The sign and rank covariance matrices based on the Oja objective functions are **affine equivariant** in the sense that if $\mathbf{x}_i^* = A\mathbf{x}_i + \mathbf{b}$, $i = 1, \dots, n$ (A is of full rank) the sign and rank covariance matrices for the transformed data set (SCM_3^* and RCM_3^*) are

$$SCM_3^* = A^* SCM_3 {A^*}^T \text{ and } RCM_3^* = A^* RCM_3 {A^*}^T$$

where $A^* = abs(Det(A))(A^{-1})^T$. See Brown, Hettmansperger, Möttönen and Oja (1997) and Hettmansperger, Möttönen and Oja (1998). Note that if A is orthogonal, then $A^* = A$, and for diagonal $A = diag(a_1, ..., a_p)$ with positive elements, $A^* = diag(\frac{\prod_i a_i}{a_1}, ..., \frac{\prod_i a_i}{a_p})$. This means that the theoretical sign and rank covariance matrices carry the shape information (eigenvectors and standardized eigenvalues): If $c_1 \geq c_2 \geq ... \geq c_p$ and $\mathbf{u}_1, ..., \mathbf{u}_p$ are the standardized eigenvalues and the eigenvectors of the (genuine) covariance matrix, the standardized eigenvectors and eigenvalues from the sign and rank covariance matrices are $\frac{1}{c_1}, \frac{1}{c_2}, ..., \frac{1}{c_p}$ and $\mathbf{u}_1, ..., \mathbf{u}_p$. The standardized eigenvalues are just replaced by their inverses.

Assume that $\mathbf{x}_1, ..., \mathbf{x}_n$ is a random sample from a multivariate distribution with c.d.f. F. The affine equivariant sign and rank vectors $\mathbf{S}_3(\mathbf{x})$ and $\mathbf{R}_3(\mathbf{x})$ then converge (uniformly in probability) to the corresponding theoretical sign and rank vectors which depends on the distribution F, see Hettmansperger, Nyblom and Oja (1994) and Hettmansperger, Möttönen and Oja (1997).

A covariance matrix estimate S may then be obtained through steps C1-C3 in section 3.2 or as follows

D1. Calculate the SCM or RCM using Oja sign vectors (w.r.t. the Oja median) or Oja rank vectors. Find eigenvector and standardized eigenvalue estimates from the calculated SCM or RCM, that is, matrices U and C.

D2. Estimate λ (or Wilks generalized variance λ^p) using any robust estimate.

D3. The covariance matrix estimate is

$$S = \lambda U C U^T.$$

In passing we wish to note that, in all the above rank covariance matrix constructions, one can avoid the estimation of multivariate location. This is not true in the case of sign correlations.

4 ROBUSTNESS AND EFFICIENCY STUDIES

In this section we consider and compare robustness and efficiency properties of the spatial Kendall's tau covariance matrix (spatial TCM) and the Oja rank covariance matrix (Oja RCM). These two matrix estimates were selected for this illustration because of the similarity of the estimation procedures: They both avoid estimation of the location centre and they both provide shape and orientation estimates. Recall that eigenvectors of both the spatial TCM and the Oja RCM are convergent estimates of the eigenvectors of the theoretical covariance matrix. Moreover, the condition number of the Oja RCM and the squared condition number of the spatial TCM estimate the theoretical condition number.

To visualize the effects of outlying observations, the estimated tolerance ellipses based on the sample covariance matrix , the spatial TCM and the Oja RCM are plotted in Figure 1 for uncorrupted data as well as for data with different proportions of outliers. The data are normally distributed bivariate data ($n = 20$) and 1,2 or 7 observations are replaced by outliers in the examples. The tolerance ellipses are constructed to cover exactly 50 % of the data. The midpoints of the ellipses are located at natural companion estimates, namely the mean vector (sample covariance matrix), spatial median (spatial TCM) and Oja median (Oja RCM). For the uncorrupted data set, all three ellipses seem quite similar. Adding one outlier rotates strongly the tolerance ellipse based on the sample covariance matrix. The condition number (shape) of the Oja RCM has been changed also. No significant change can be seen in the spatial TCM. The effects of two outliers (in the same direction) are quite similar but naturally stronger for sample covariance matrix and Oja RCM whereas the ellipse associated with spatial TCM retains its shape and orientation. In the last case seven outliers have a drastic influence on the sample covariance matrix and the Oja RCM. Furthermore, the sample mean is not in the convex hull of the original uncontaminated data anymore. The spatial TCM tolerance ellipse has started to turn towards outlying observations.

In order to study the qualitative robustness of the spatial TCM and the Oja RCM, the empirical influence functions were plotted for these two rank covariance matrices and also for the sample covariance matrix. The changes between the original and perturbed matrices were quantified using two criteria, the proportional change in the condition numbers (shape) and the change in the direction of the first eigenvector (orientation). Independent samples of sizes 100 were drawn from the bivariate normal distribution with symmetry center $\boldsymbol{\mu} = (0, 0)^T$ and covariance matrix

$$\Sigma = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}.$$

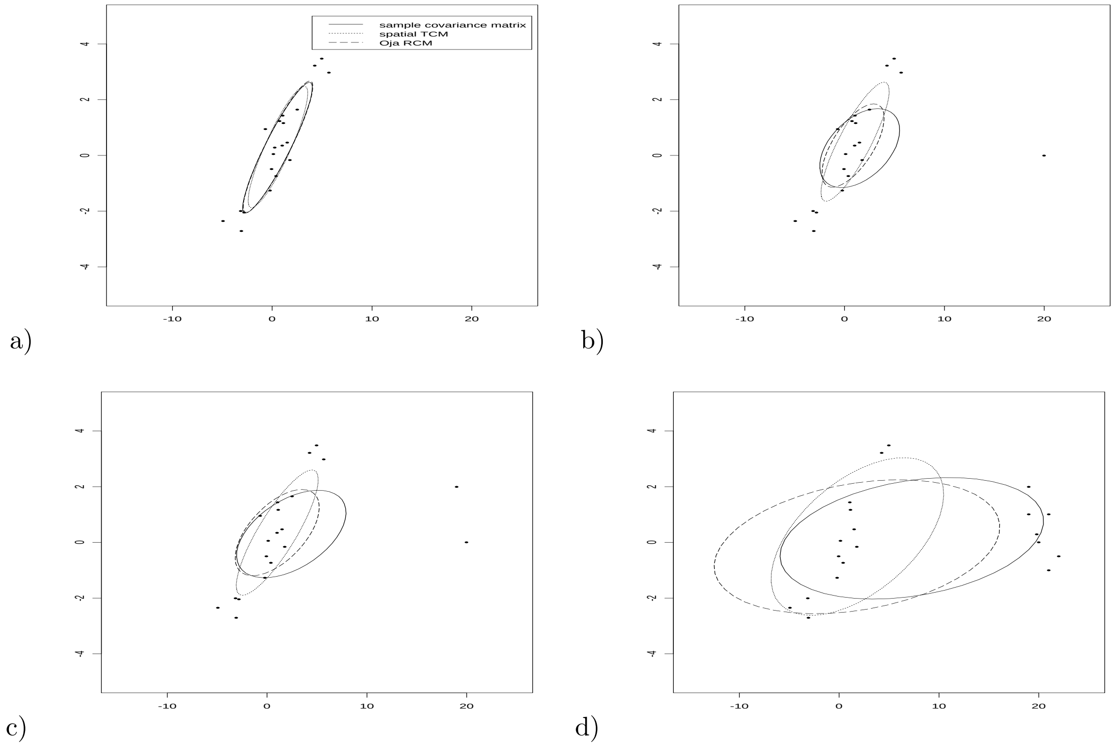

Figure 1: The 50% tolerance ellipses based on the sample covariance matrix, the spatial TCM and the Oja RCM in the case of a) original data ($n = 20$), b) one observation replaced by an outlier, c) two observations replaced by outliers and d) seven observations replaced by outliers.

Each of the samples were perturbed with the same multivariate outlier; the smoothed sensitivity plots (averages over 50 samples) are given in Figure 2. For the sample covariance matrix the influence of one additional observation is unbounded. The influence functions for the Oja RCM are also unbounded but the changes in criterion values are less dramatic there. The influence functions for the spatial TCM are bounded.

A simulation study was used to compare the small sample efficiencies of the spatial TCM and Oja RCM shape and orientation estimates. The sample size was $n = 50$ and the number of samples was 1000. The bivariate distributions used in the simulations were:

- Normal distribution and Laplace distribution with symmetry center $\boldsymbol{\mu} = (0, 0)^T$ and the same covariance matrix Σ as above.

- Contaminated normal distribution of type I (CN(I)) where distribution is $N(\mathbf{0}, \Sigma)$ with probability $1 - \epsilon$ and $N(\mathbf{0}, 10\Sigma)$ with probability ϵ. The two values used for ϵ were 0.05 and 0.1.

Table 1: Efficiencies of the spatial TCM and Oja RCM relative to the sample covariance matrix. The criteria are the direction of the first eigenvector (orientation) and the logarithm of the condition number (orientation). The efficiencies are given in terms of ratios of MSE's. The error in the direction is defined as the difference in the direction of the first eigenvector in radians: $\sphericalangle(\hat{\mathbf{u}}_1, \mathbf{u}_1)$, where $\sphericalangle \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\hat{\mathbf{u}}_1$ is the estimated first eigenvector and $\mathbf{u}_1$ is the first eigenvector of the theoretical covariance matrix. The error in the shape is defined as the deviation of the logarithm of the estimated condition number from the theoretical condition number log(4).

	First Eigenvector		Log(Cond Numb)	
Distribution	Spatial	Oja	Spatial	Oja
Normal	0.89	0.99	0.87	0.99
CN(I)				
$\epsilon = 0.05$	1.96	1.51	1.59	1.42
$\epsilon = 0.1$	2.37	1.68	1.94	1.42
CN(II)				
$\epsilon = 0.05$	16.89	3.54	3.00	1.23
$\epsilon = 0.1$	19.32	2.12	2.21	0.99
Laplace	1.14	1.12	1.14	1.12

- Contaminated normal distribution of type II (CN(II)) where distribution is $N(\mathbf{0}, \Sigma)$ with probability $1 - \epsilon$ and $N(\mathbf{0}, \Sigma_2)$ with probability ϵ, where

$$\Sigma_2 = \begin{pmatrix} 10 & 0 \\ 0 & 40 \end{pmatrix}.$$

  The values for ϵ were same as above.

The results are reported in Table 1. In the bivariate normal case the efficiencies of the Oja RCM shape and orientation estimates are almost 1; also the spatial TCM estimates are doing well. For heavy tailed distributions, the rank based methods are highly efficient. In these cases the spatial TCM seems to be clearly better than the Oja RCM.

5 SIGNAL PROCESSING EXAMPLE

In multichannel signals, data vector component variances are often unequal and components are often correlated. In many applications there is a need to decorrelate the observed signal components and perform normalization by making the component variances equal. In signal processing jargon such operation is called the Whitening Transform. Whitening transformation is often assumed to be performed before applying any estimator to multichannel signals. The transform matrix V is given in terms of eigenvalues and eigenvectors of the sample covariance matrix as follows

$$V = \Lambda^{-1/2} U^T$$

where Λ is the matrix of eigenvalues and U is a matrix of eigenvectors. If Y are the original data with correlated components and unequal variances, a transformed data Z where components are uncorrelated and covariance matrix is an identity matrix I is obtained by

$$Z = VY.$$

If sample covariance matrix is used for whitening, it produces very unreliable results in the face of outliers. Therefore, robust covariance estimators should be considered.

As an example of the use of Whitening transform and application of robust covariance estimators, we consider the problem of Blind Source Separation (BSS). In BSS, one has a collection of sensors such as microphones that observe a linear combination of source signals such as speech of individual speakers. The task is then to separate the source signals (voices of individual speakers) from each other. The separation is achieved by finding statistically independent components from these linear mixtures, see Cardoso (1989), Oja, E. (1997). The term blind refers to the fact that we have no prior information about the structure of the mixing system and the source signals are unknown as well. In Blind Source Separation, the unobservable source signals and the observed mixtures are related by

$$\mathbf{y}_k = A\mathbf{s}_k + \mathbf{v}_k$$

where A is an $n \times m$ matrix of unknown constant mixing coefficients, $n \geq m$, $\mathbf{s}$ is a column vector of m source signals, $\mathbf{y}$ is a column vector of n mixtures and $\mathbf{v}$ is additive noise vector and k is time index. The matrix A is assumed to be of full rank and source signals are typically assumed to be zero mean, stationary and non-Gaussian. The separation task at hand is to estimate the elements of a separating matrix denoted by H so that the original sources are recovered from the noisy mixtures. As a preprocessing step the observed data $\mathbf{y}$ are decorrelated by Whitening transform. Whitening allows for solving the separation problem more easily because uncorrelated components with variance $\sigma^2 = 1$ are used as an input and if $n = m$, separating matrix will be orthogonal ($H^{-1} = H^T$). An estimate $\mathbf{x}$ of unknown sources $\mathbf{s}$ is given by

$$\hat{\mathbf{s}} = \mathbf{x} = \hat{H}^T\mathbf{y}.$$

The estimate can be obtained only up to a permutation of $\mathbf{s}$, i.e., the order of the sources may change. A solution may be obtained, for example, by using higher order cumulants (Cardoso 1989).

An example of the separation is depicted in Figure 3. In our simulation 2 source signals and 2 mixtures with randomly generated coefficient matrix A was used. Both observed mixture sequences of 500 observations are contaminated with zero mean additive Gaussian (Normal) noise with variance $\sigma^2 = 0.1$. Moreover, 5% of the observations are randomly replaced by outliers with large amplitude. The resulting contaminated distributions are asymmetric. The actual separation is performed using a least squares algorithm, see Koivunen et al. (1998) for details. It is preceded by robust whitening which allows for performing the separation reliably whereas whitening using sample covariance matrix followed by the same separation method produces incomprehensible results. Robust covariance matrix estimate used in this example is obtained using spatial Kendall's tau covariance matrix and the median of the robust Mahalanobis distances as described in Section 3.2.

If $n > m$, the estimated covariance matrix allows for determining the number of source signals present in the mixtures as well. If there are m sources, there is a clear pattern in the eigenvalues of the covariance matrix when the noise power is not too high. In such cases, the number of sources may be estimated using information theoretic criteria such as the Akaike Information Criterion or the Minimum Description Length (MDL) (Rissanen 1978). The ratio of the geometric to the arithmetic mean of the eigenvalues of the covariance matrix may be used to determine the number signals in Gaussian noise (Wax & Kailath 1985). If the noise distribution is contaminated Gaussian, sample covariance matrix gives incorrect results whereas a robust covariance estimator typically yields the correct result.

6 FINAL COMMENTS

Classical multivariate statistical techniques assume that the data were generated from a multivariate normal (Gaussian) distribution. The inference methods are then often based on the sample covariance matrix. The inference about means (MANOVA) compares "between-samples" and "within-samples" matrices, the multivariate (multiple) regression uses partitioned covariance matrices, and the starting point for the classical principal component analysis, factor analysis, and the canonical correlation analysis, for example, is the sample covariance or correlation matrix. Normal-theory based methods are, however, extremely sensitive to outlying observations. Recently many robust and nonparametric methods have been proposed for estimating a multivariate scatter.

In this paper we investigated robust covariance matrix estimators. We proposed new robust and nonparametric techniques for estimating covariances and correlations. The new covariance matrix estimates are based on different multivariate rank and sign concepts and they may be used to construct efficient equivariant robust techniques for multiple regression analysis, principal component analysis, canonical correlation analysis, factor analysis, discrimination analysis, etc. New tools for studying the qualitative robustness in terms of the eigensystem of the covariance matrix were proposed and illustrated. The efficiency and robustness properties of two rank based estimators were studied in a simulation study. More detailed statistical properties of the estimates (convergence, efficiencies, equivariance, robustness) will be considered in subsequent papers.

Covariance and correlation estimation play an important role in many signal processing tasks, in telecommunication applications, in pattern recognition and in machine vision. In this paper, we showed how a robust covariance estimator provide reliable results in applications where the observed noisy signal is contaminated by outliers. The sample covariance matrix, on the other hand, had an unacceptable poor performance. The new robust covariance estimators developed in this work will be applied to different signal processing and image analysis problems. Therefore, the development of computationally efficient algorithms is very important. The work on robust covariance estimators can also be extended to estimation of autocorrelation/autocovariance matrices.

References

Bensmail, H. and Celeux, G. 1996. Regularized gaussian discriminant analysis through eigenvalue decomposition. *Journal of the American Statistical Association*, 91, 1743-1749.

Brown, B.M., Hettmansperger, T.P., Möttönen, J. and Oja, H. 1997. Rank plots in the affine invariant case. In *L1-Statistical Analysis and Related Methods.* Ed. by Y. Dodge, 351-362.

Campbell, N.A. 1980. Robust procedures in multivariate analysis I: Robust covariance estimation. *Applied Statistics*, 29, 231-237.

Capéraà, P. and Genest, C. 1993. Spearman's ρ is larger than Kendall's τ for positively dependent random variables. *Nonparametric Statistics*, 2, 183-194.

Cardoso, J-F. 1989. Source separation using higher order moments. *Proc. of the IEEE Intl Conf. on Acoustics, Speech and Signal Processing (ICASSP-89)*, 2109-2112.

Chaudhuri, P. 1996. On a geometrical notion of quantiles for multivariate data. *Journal of the American Statistical Association,* 91, 862-872.

Choi, K. and Marden, J. 1998. A multivariate version of Kendall's τ. *Nonparametric Statistics*, 9, 261-293.

Croux, C., Haesbroeck, G. and Rousseeuw, P.J. 1997. Location adjustment for the minimum volume ellipsoid estimator. Submitted.

Croux, C. and Ruiz-Gazen, A. 1997. A fast algorithm for robust principal components based on projection pursuit. Submitted.

Davies, P.L. 1987. Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, 15, 1269-1292.

Donoho, D.L. 1982. *Breakdown properties of multivariate location estimators.* Ph.D. qualifying paper, Harvard University.

Gibbons, J.D. and Chakraborti, S. 1992. *Nonparametric Statistical Inference.* Marcel Dekker, New York.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel W.A. 1986. *Robust Statistics.* Wiley, New York.

Hettmansperger, T.P., Nyblom, J. and Oja, H. 1994. Affine invariant multivariate one-sample sign tests. *Journal of the Royal Statistical Society, Ser. B,* 56, 221-234.

Hettmansperger, T.P., Möttönen, J. and Oja, H. 1997. Multivariate affine invariant one-sample signed-rank tests. *Journal of the American Statistical Society,* 92, 1591-1600.

Hettmansperger, T.P., Möttönen, J. and Oja, H. 1998. Affine invariant multivariate rank tests for several samples. *Statistica Sinica*, 8, 785-800.

Huber, P.J. 1977. *Statistical Decision Theory and Related Topics II,* S.S. Gupta and D.S. Moore, Eds., Academic Press, New York.

Huber, P.J. 1981. *Robust Statistics.* Wiley, New York.

Kendall, M.G. 1975. *Rank Correlation Methods.* Griffin, London.

Kent, J.T. and Tyler, D.E. 1991. Redescending M-estimates of multivariate location and scatter. *The Annals of Statistics*, 19, 2102-2119.

Koivunen, V. and Kassam, S. A. 1996. Covariance estimation in multivariate OS-filtering. *IEEE ICIP-96*, Vol. II, 981-984.

Koivunen, V., Pajunen, P., Karhunen, J. and Oja, E. 1998. Blind separation from ϵ-contaminated mixtures. *EUSIPCO'98*, Vol. III, 1817-1820.

Koltchinskii, V. 1997. M-estimation, convexity and quantiles. *The Annals of Statistics*, 25, 435-477.

Li, G. and Chen, Z. 1985. Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte-Carlo. *Journal of the American Statistical Association,* 80, 759-766.

Lopuhaä, H.P. 1989. On the relation between S-estimators and M-estimators of multivariate location and covariance. *The Annals of Statistics,* 17, 1662-1683.

Lopuhaä, H.P. and Rousseeuw, P.J. 1991. Breakdown points of affine-equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics,* 19, 229-248.

Marden, J. I. 1999a. Multivariate rank tests. To appear in *Multivariate design, and Sampling,* ed. by Ghosh.

Marden, J. I. 1999b. Some robust estimates of principal components. *Statistics & Probability Letters,* in print.

Maronna, R.A. 1976. Robust M-estimators of multivariate location and scatter. *The Annals of Statistics* 20, 51-67.

Masarotto, G. 1987. Robust identification of autoregressive moving average models. *Applied Statistics*, 36, 214-220.

Möttönen, J. and Oja, H. 1995. Multivariate spatial sign and rank methods *Journal of Nonparametric Statistics*, 5, 201-213.

Oja, E. 1997. The Nonlinear PCA Learning Rule in Independent Component Analysis. *Neurocomputing*, 17, 25-45.

Oja, H. 1983. Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, 1, 327-332.

Oja, H. 1997. Affine invariant multivariate sign and rank tests and corresponding estimates: A review. *Scandinavian Journal of Statistics* (invited paper), to appear.

Oja, H. and Koivunen, V. 1998. Robust covariance matrix estimation with signal processing applications. In *The Yerbook of the Finnish Statistical Society 1997.* Helsinki, 87-109.

Rissanen, J. 1978. Modeling by shortest data description. *Automatica,* 14, 465-471.

Rousseeuw, P.J. 1984. Least median of squares regression. *Journal of American Statistical Association*, 79, 871-880.

Rousseeuw, P.J. and Leroy, A. 1987. *Robust regression and outlier detection.* Wiley, New York.

Stahel, W.A. 1981. Breakdown of covariance estimators. Research Report 31, Fachgruppe für Statistik, E.T.H. Zürich.

Stewart, G.W. and Sun J. 1990. *Matrix perturbation theory.* Academic Press, New York.

Wax, M. and Kailath, T. 1985 Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech, and Signal Processing (T-ASSP),* 33, 387-392.

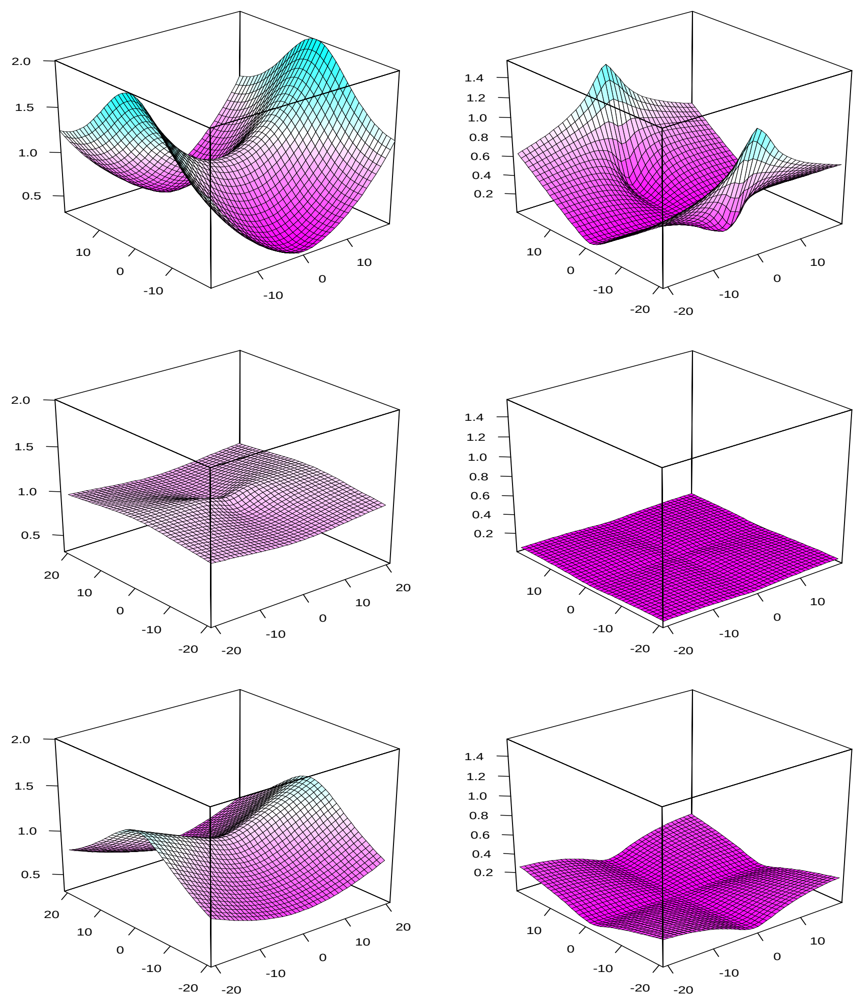

Figure 2: Sensitivity plots for the sample covariance matrix (top row), the spatial Kendall's tau matrix (middle row) and the Oja RCM (bottom row). Used criteria are the ratio of the condition numbers (left column) and the direction of the first eigenvector (right column)

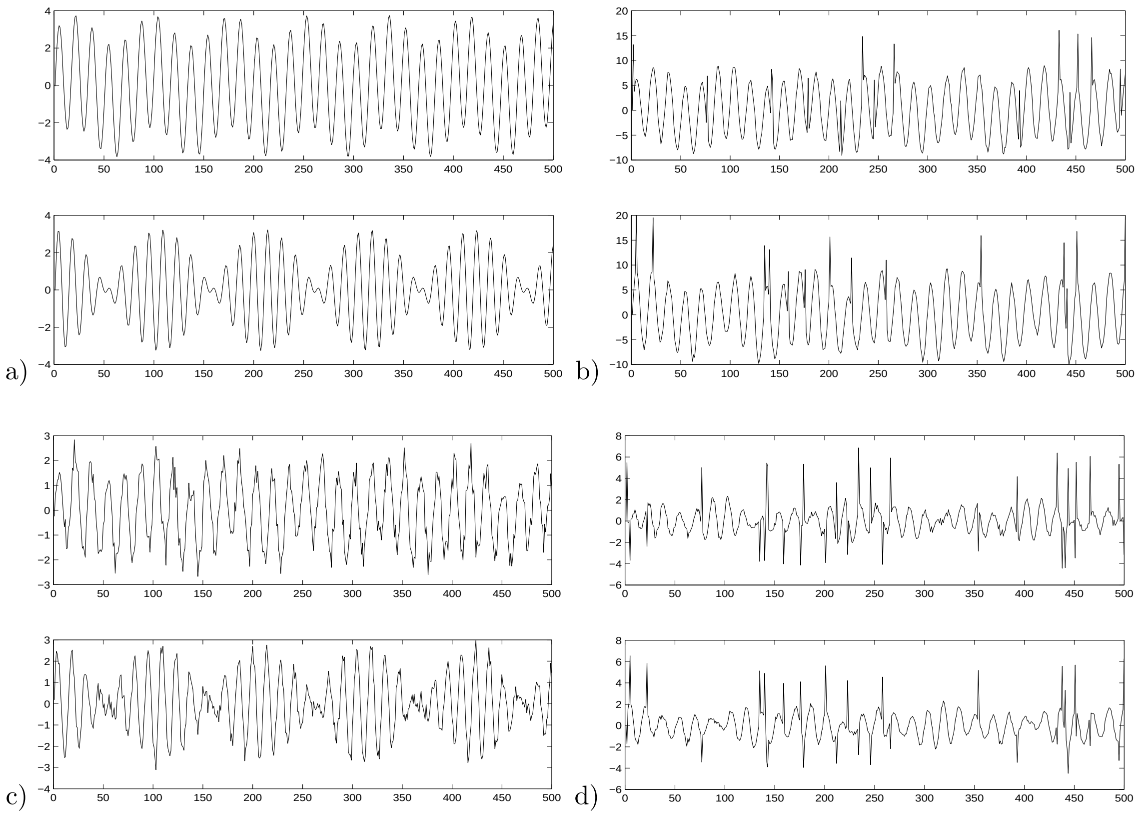

Figure 3: An example of BSS from noisy sequences: a) Noise free source signals, b) ϵ-contaminated mixtures ($\epsilon = 0.05$), c) separation result using robust whitening, d) separation result using whitening based on sample covariance matrix.