MULTICHANNEL SIGNAL PROCESSING USING SPATIAL RANK COVARIANCE MATRICES

S. Visuri[1] *H. Oja*[2] *V. Koivunen*[1]

[1]Signal Processing Lab.
Tampere Univ. of Technology
P.O. Box 553, FIN-33101 Tampere
Finland

[2]Dept. of Statistics
University of Jyväskylä
P.O. Box 35, FIN-40351 Jyväskylä
Finland

ABSTRACT

This paper addresses the problem of estimating the covariance matrix reliably when the assumptions, such as Gaussianity, on the probabilistic nature of multichannel data do not necessarily hold. Multivariate spatial sign and rank functions, which are generalizations of univariate sign and centered rank, are introduced. Furthermore, spatial rank covariance matrix and spatial Kendall's tau covariance matrix based new robust covariance matrix estimators are proposed. Efficiency of the estimators is discussed and their qualitative robustness is demonstrated using a empirical influence function concept. The use and reliable performance of the proposed methods is demonstrated in color image filtering, image analysis, principal component analysis and blind source separation tasks.

1. INTRODUCTION

A growing number of signal processing applications require processing of multichannel data. The application domains include biomedical signal processing such a EEG, color image processing and array processing. The components in multichannel signal may have different scales and they may be correlated. Hence, it is desirable to develop techniques that take into account the correlation structure and scale. Such techniques typically require the estimation of a covariance matrix. The efficiency of the sample covariance matrix deteriorates drastically if the signal is contaminated by non-Gaussian noise. In particular, if the data are contaminated by outliers, poor estimates are obtained. Therefore, there is a need for techniques that produce nearly optimal results at nominal conditions and reliable estimates if the assumptions on the underlying signal and noise model are not completely valid.

In this paper, we investigate the problem of robust covariance estimation. Techniques based on the concepts of multivariate sign function and rank function are introduced for estimating the covariance matrix for the multichannel signal. Let us express a $p \times p$ covariance matrix Σ in terms of its eigenvalue decomposition [2] $\Sigma = \lambda U C U^T$ where the scalar λ^p is the generalized variance, the matrix U contains the eigenvectors and the diagonal matrix C the standardized eigenvalues such that $det(C) = 1$. The proposed techniques estimate the covariance matrix in pieces, i.e., U, C and λ separately and then construct the complete covariance matrix. The obtained covariance matrix, its eigenvectors U, standardized eigenvalues C or eigenvalues $\Lambda = \lambda C$ may then be used in various signal processing applications. To illustrate the practical use of the proposed estimators, we introduce a new algorithm for color image filtering.

This paper is organized as follows. In section 2, the spatial rank covariance matrix (RCM) and the spatial Kendall's tau covariance matrix (TCM) are introduced. Methods for constructing robust covariance matrix estimates using the RCM and TCM are given. The efficiency properties of the proposed estimates are discussed and the qualitative robustness of the TCM is demonstrated using the finite sample influence function concept. Section 3 gives several signal processing examples where the proposed estimators are employed. The new algorithm for color image filtering is developed. The robustness of the covariance estimators is illustrated by a principal component analysis example and by an image analysis example where the position and orientation of an object in an image is determined. In addition, an estimation method based on the RCM is applied to a whitening transform in a blind source separation task. Finally, section 4 concludes the paper.

2. SPATIAL RANK COVARIANCE MATRICES

We begin by giving definitions for multivariate spatial sign and rank concepts used in this article. For a p-variate data set $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, the *spatial rank function* is [9]

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

where $\mathbf{S}$ is the *spatial sign function*

$$\mathbf{S}(\mathbf{x}) = \begin{cases} \frac{\mathbf{x}}{||\mathbf{x}||}, & \mathbf{x} \neq \mathbf{0} \\ \mathbf{0} & \mathbf{x} = \mathbf{0} \end{cases}.$$

A spatial median $\mathbf{M}(X)$ solves $\mathbf{R}(\mathbf{x}) = \mathbf{0}$ (for the spatial median see for example [11]). The spatial sign and rank functions are multivariate generalizations of the univariate sign and centered rank functions.

In the following we describe how the spatial rank covariance matrix $RCM = n^{-1} \sum_i \mathbf{R}(\mathbf{x}_i)\mathbf{R}(\mathbf{x}_i)^T$ and the spatial Kendall's tau covariance matrix $TCM = n^{-2} \sum_{i,j} \mathbf{S}(\mathbf{x}_i - \mathbf{x}_j)\mathbf{S}(\mathbf{x}_i - \mathbf{x}_j)^T$ can be used in robust covariance matrix estimation. A covariance matrix Σ can be expressed in terms of its eigenvalue decomposition as follows

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

where the scalar λ^p is the generalized variance, the orthogonal matrix U contains the eigenvectors and C is the matrix of the standardized eigenvalues ($det(C) = 1$). It can be shown [7], that for so called elliptically symmetric distributions [5] the eigenvectors of the theoretical RCM and TCM are the same as the eigenvectors of the ordinary covariance matrix. There is also a one-to-one correspondence between the standardized eigenvalues of the theoretical TCM and the eigenvalues of the usual covariance matrix.

The estimation strategy using the RCM or TCM may now be given as follows:

1. Calculate the RCM or TCM of the data. Find the corresponding eigenvector estimates, that is, matrix $\hat{U}$.
2. Estimate the marginal values (eigenvalues, principal values) of
$$\hat{U}^T\mathbf{x}_1, \ldots, \hat{U}^T\mathbf{x}_n$$
using any robust univariate scale estimate. Write $\hat{\Lambda} = diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ for the estimates.
3. The covariance matrix estimate is
$$\hat{\Sigma} = \hat{U}\hat{\Lambda}\hat{U}^T.$$

As explained in this section, it is also possible to construct both eigenvector and standardized eigenvalue estimates, say $\hat{U}$ and $\hat{C}$, from the spatial TCM. To estimate the scale, use first the matrix $\hat{U}\hat{C}\hat{U}^T$ to construct robust (squared) Mahalanobis type of distances from the spatial median $\mathbf{M}(X)$,

$$d_i = (\mathbf{x}_i - \mathbf{M}(X))^T\hat{U}\hat{C}^{-1}\hat{U}^T(\mathbf{x}_i - \mathbf{M}(X)),\ i = 1, \ldots, n.$$

The generalized variance (λ^p) may then be estimated robustly by $h_p \times Med(D)^p$ where the correction factor h_p can be selected to guarantee the convergence for the underlying distribution.

Our simulation results [12, 13] show that, in the bivariate normal case, the efficiencies of the estimates based on the two spatial rank covariance matrices are quite high (about 90 percent for sample size 50) as compared to those based on the usual covariance matrix. For heavy tailed distributions the spatial methods are much more efficient.

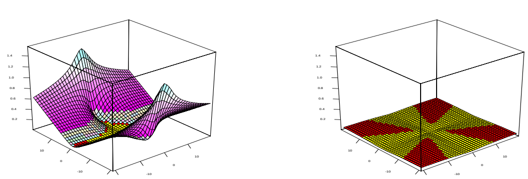

Figure 1: The empirical influence function plots for the first eigenvector of the sample covariance matrix (left) and for the first eigenvector of the TCM (right). The influence of an outlier to the TCM is bounded.

In addition to being highly efficient, the robustness should be considered. An empirical influence function is a standard tool for describing the qualitative robustness of an estimator. Roughly speaking it measures the influence of one additive observation to the estimator. For a robust estimator the empirical influence function should naturally be bounded. Figure 1 shows the empirical influence function plots for the sample covariance matrix and for the TCM. The effect of one added observation to the direction of the eigenvector corresponding to the largest eigenvalue of the estimated covariance matrix was measured. For the sample covariance matrix, the influence of one additional observation is unbounded whereas the empirical influence function of the TCM is bounded.

Table 1: Average component mean square errors for the filtering results shown in Figure 2.

Method	MSE
VM	99.5
RCM	88.2

3. APPLICATION EXAMPLES

As an example of the use of the spatial rank covariance matrices in multichannel filtering, the RCM is applied to attenuating noise in RGB color images. The algorithm employing the RCM proceeds as follows:

1. In a processing window of n vectors $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, estimate the covariance matrix robustly using estimation strategy described in section 2 (use Median Absolute Deviation (MAD) to estimate the marginal variances). Write $\hat{\Sigma}$ for the estimated covariance matrix. Find also the spatial median $\mathbf{M}(X)$ of the data by solving the equation $\mathbf{R}(\mathbf{x}) = \mathbf{0}$.
2. Compute the (squared) robust Mahalanobis type of distances
$$d_i = (\mathbf{x}_i - \mathbf{M}(X))^T\hat{\Sigma}^{-1}(\mathbf{x}_i - \mathbf{M}(X))\ i = 1, \ldots, n$$
and find the median $M(D)$ of these distances.
3. The output is the half sample mean of the data
$$\mathbf{T}(X) = \frac{\sum_{i=1}^n w_i\mathbf{x}_i}{\sum_{i=1}^n w_i}$$
where
$$w_i = \begin{cases} 1 & \text{if} \quad d_i \leq M(D) \\ 0 & \text{if} \quad d_i > M(D) \end{cases}.$$

The filtering algorithm was applied to a 256×256 RGB color image using a 3×3 processing window. A test image was created by adding multivariate Gaussian noise ($\sigma_1^2 = 400, \sigma_2^2 = 225, \sigma_3^2 = 100$ and $\rho = 0.7$) to the original image. Furthermore, 10% of the samples were randomly replaced by outliers, having the minimum or maximum value. The original image, the noisy image and the filtered image are shown in Figure 2. In quantitative comparison in terms of mean square error (MSE) criterion, the proposed filtering method outperforms the Vector Median (VM) [1], see Table 1. It also preserves the edges and details significantly better than VM as can be seen from Figure 2.

In our second example we consider Discrete Karhunen-Loève Transform (DKLT) and Principal Component Analysis (PCA) which are widely used in various signal processing applications.

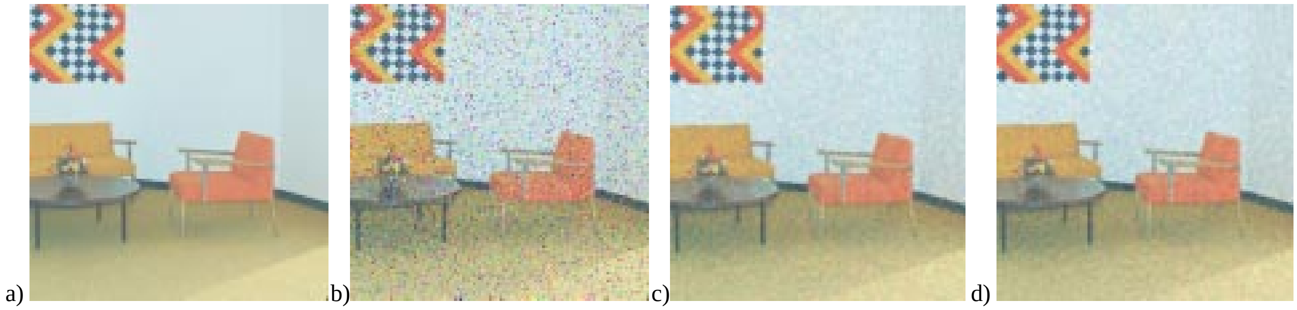

Figure 2: a) The original Room 256×256 RGB color image and b) the noisy image where multivariate Gaussian noise with unequal component variances ($\sigma_1^2 = 400,\ \sigma_2^2 = 225,\ \sigma_3^2 = 100, \rho = 0.7$) is added. Moreover, 10% of the samples are replaced by outliers which have the minimum or maximum signal value with equal probability. c) The filter output using the vector median algorithm. d) The filter output using the algorithm employing the RCM.

These techniques employ eigenanalysis of the covariance or autocorrelation matrix. Their applications areas include: data compression, image analysis, pattern recognition, detection, spectrum estimation and array processing. Due to the large number of applications, it is of interest to see how the performance of PCA or DKLT deteriorates in the face of outliers.

Principal component analysis looks for a few linear combinations which can be used to summarize the data, losing in the process as little information as possible, see [8]. First task in PCA is to transform the observed data so that the components are uncorrelated and the first principal component explains the largest possible amount of the information in the data, the second component explains the second largest amount of variation of the data and so on.

Let $\mathbf{x}$ be a p-variate random variable with a covariance matrix $\Sigma = U^T \Lambda U$ where the ordered eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$ of the matrix Σ are on the diagonal of the matrix Λ and the matrix U is the corresponding eigenvector matrix having the eigenvectors on the columns. Now the covariance matrix of the transformed random variable

$$\mathbf{y} = U^T \mathbf{x}$$

is Λ so in the theory we know the transform leading to the observations with properties described above. The amount of the variation explained by the first $k \leq p$ principal components (first k elements of the r.v. $\mathbf{y}$) is $(\lambda_1 + \cdots + \lambda_k)/(\lambda_1 + \cdots + \lambda_p)$.

In practice, the matrix U above has to be estimated from the observations. Usual estimator is the eigenvector matrix of the sample covariance matrix. This estimators is optimal in the case of Gaussian distribution but it is well known that it can give very misleading results if the underlying distribution has heavy tails or the data are contaminated by outliers. In such cases, robust estimators such as the method proposed in this paper are worth considering.

To demonstrate the difference in the behavior of the conventional sample covariance matrix and robust estimators we generated a random sample of 100 observations from the 5-variate normal distribution with the covariance matrix $\Sigma_1 = diag(100, 50, 25, 1, 0.99)$ and the symmetry center $\boldsymbol{\mu} = \mathbf{0}$. We contaminated the data by adding 5 observations from the normal distribution with the covariance matrix $\Sigma_2 = diag(100, 50, 25, 1000, 1000)$ and the symmetry center $\boldsymbol{\mu} = \mathbf{0}$.

The results of the PCA using the conventional sample estimate and the robust estimate for covariance matrix are reported in Table 2 for the original and contaminated data. Robust estimate for the covariance matrix was produced using the RCM and the technique described in the section 2. For simplicity we have only considered in our results the standard deviation of the principal components and the amount of the total variation explained by the $k \leq p$ first principal components. The results show that only 5 outlying observations add an irrelevant dimension to the data when we use the estimate based on the normal distribution theory. On the other hand, outlying observations do not have any significant influence to the PCA based on the robust estimator. From this simple example it is easy to see that if the assumption of the Gaussianity is not necessarily true, the normal theory based PCA and the signal processing methods based on it may give strongly misleading results. Therefore, robust covariance matrix estimators should be considered.

The deteriorating influence of outliers to the sample estimates may be qualitatively demonstrated by using a simple image analysis example. The task at hand is to determine the position and orientation of an aircraft in an image. The aircraft is segmented from the background using simple thresholding method. As a result, there remains a few outlying points that are not from the surface of the airplane. After segmenting the airplane from the background we have just a two dimensional data set of pixel coordinates. Therefore the center and the orientation of the airplane can be determined using any estimates for the two dimensional location and scatter (orientation is obtained from the ordered eigenvectors of an estimated covariance matrix).

In Figure 3 we have illustrated the results using the sample estimates and the robust estimates by drawing the tolerance ellipses obtained from the sample covariance matrix and from the TCM. Ellipses are located at the estimates of the center: the sample mean vector and the spatial median. The direction of the first main axis of the tolerance ellipse gives the estimated main direction of the object. The outlying points influence the sample estimates severely so that the position and the orientation are obtained incorrectly whereas the robust estimates give us much more reliable results.

In our third example we consider the whitening transform in the blind source separation problem. 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. Such operation is called

Table 2: Results from the PCA using the conventional and robust covariance matrix estimator. The results for the corrupted data are printed in **bold.**

Criterion	Method	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5
Standard deviation	Conventional	9.48	6.69	4.50	1.04	0.93
		9.77	**7.92**	**6.73**	**5.65**	**4.47**
	Robust	10.85	6.62	4.91	1.02	0.81
		10.93	**6.37**	**5.30**	**1.21**	**0.79**
Cumulative Proportion	Conventional	0.57	0.86	0.99	0.99	1.00
		0.37	**0.62**	**0.80**	**0.92**	**1.00**
	Robust	0.63	0.86	0.99	1.00	1.00
		0.63	**0.84**	**0.99**	**1.00**	**1.00**

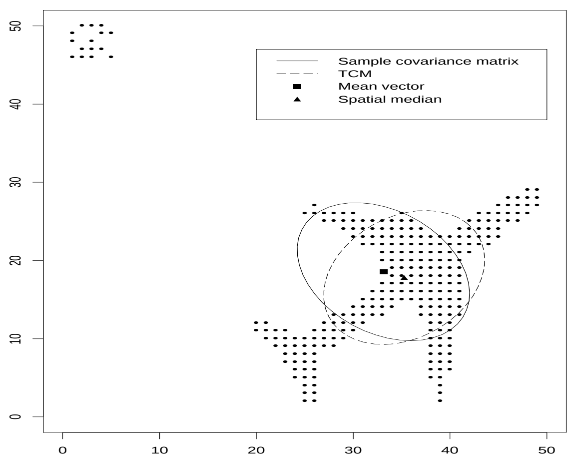

Figure 3: The pixel coordinates, the estimates for the center and the 50% tolerance ellipses obtained from the sample covariance matrix and the TCM.

the whitening transform. Whitening transformation is often assumed to be performed before applying any estimator to multichannel signals.

The theoretical basis for the whitening transform is very similar to the theory reviewed in the principal component analysis example. Let again $\mathbf{x}$ be a p-variate random variable with covariance matrix $\Sigma = U^T \Lambda U$ where matrices U and Λ are as before. Let the transform matrix V be given in terms of eigenvalues and eigenvectors of the covariance matrix as follows

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

The covariance matrix of the transformed random variable

$$\mathbf{y} = V\mathbf{x}$$

is an identity matrix I_p so, again, in theory we know the transformation leading to the desired properties of the marginal variables. If the sample covariance matrix is used for whitening, very unreliable results can be obtained in the face of outliers. Therefore, robust covariance estimators should be considered.

In Blind Source Separation (BSS) problem, 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 [4, 10]. 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.

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 an additive noise vector and k is a 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 centered at the origin and decorrelated by the whitening transform. Whitening allows for solving the separation problem more easily. Uncorrelated components with variance $\sigma^2 = 1$ are used as an input to the actual separation. By projecting the observed data into subspace spanned by eigenvectors corresponding to m largest eigenvalues, we will have $n = m$ and the separating matrix will be orthogonal ($H^{-1} = H^T$). The projection into signal subspace will attenuate some noise as well. 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 [4].

An example of the separation is depicted in Figure 4. In our simulation, 5 source signals and 7 mixtures with randomly generated coefficient matrix A were used. The 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 actual separation is performed using

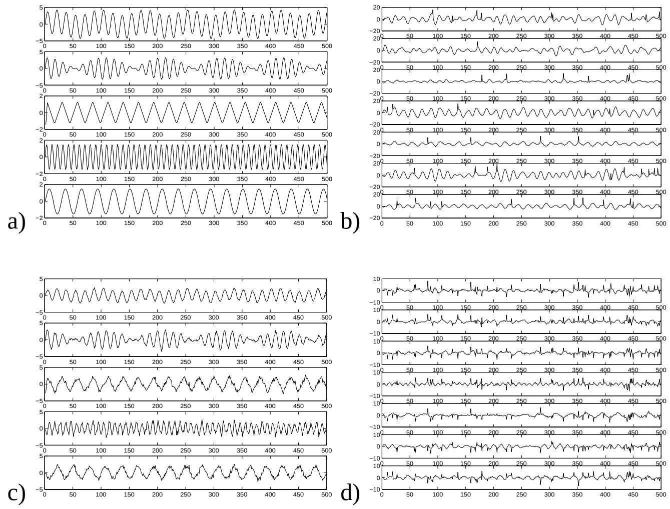

Figure 4: An example of the BSS from noisy sequences: a) Noise free source signals, b) the noisy mixtures, c) separation result using robust centering and whitening, d) separation result using centering and whitening based on the sample mean and the sample covariance matrix.

a least squares algorithm, see [6] for details. Separation is preceded by robust centering and whitening which allow for performing the separation reliably whereas centering and whitening using the sample mean and the sample covariance matrix followed by the same separation method produce incomprehensible results.

The robust estimates were obtained using the spatial median and the robust covariance matrix estimation based on the RCM. Marginal variances were estimated using the MAD. Because the marginal distributions in this example are strongly non-Gaussian (we do not know the right consistency correction factor for the MAD), we obtained our final estimates in a following way:

1. We computed the weights w_i based on robust mahalanobis distances

$$d_i = (\mathbf{x}_i - \mathbf{M}(X))^T \hat{\Sigma}_0^{-1} (\mathbf{x}_i - \mathbf{M}(X))$$

   where M(X) is the spatial median of the data $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and $\hat{\Sigma}_0$ is the covariance matrix estimate obtained using the RCM and the MAD:s for the marginal variances. We used a redescending weight function giving a zero weight for the observations with sufficiently large d_i (see [3]).

2. The final estimates were the weighted mean and covariance matrix

$$\mathbf{T}(X) = \frac{\sum_{i=1}^{n} w_i \mathbf{x}_i}{\sum_{i=1}^{n} w_i}$$

   and

$$\hat{\Sigma} = \frac{\sum_{i=1}^{n} w_i (\mathbf{x}_i - \mathbf{T}(X))(\mathbf{x}_i - \mathbf{T}(X))^T}{\sum_{i=1}^{n} w_i - 1}.$$

The final step given above is called one step reweighting and it is a standard procedure in robust covariance matrix estimation.

4. CONCLUSION

The main purpose of this paper was to introduce the concepts of spatial sign and rank and demonstrate their use in different multichannel signal processing tasks. Robust covariance matrix estimates obtained from the spatial rank covariance matrix and the spatial Kendall's tau covariance matrix were used in RGB color image filtering, principal component analysis, discrete Karhunen Loève Transform and Blind source separation problem. In addition, we showed how the methods based on the sample covariance matrix give strongly misleading results in the face of outliers.

At the end it seems worth mentioning that the robust covariance matrix estimation methods introduced in this paper are easy to implement and have relatively low computational complexity. Future work extends the use of spatial sign and rank concepts to the autocorrelation/autocovariance matrix estimation.

5. REFERENCES

[1] J. Astola, P. Haavisto, and Y. Neuvo. Vector median filters. *Proceedings or the IEEE*, 78(4):678–689, 1990.

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

[3] N. A. Campbell. Robust procedures in multivariate analysis I: Robust covariance estimation. *Applied Statistics*, 29(3):231–237, 1980.

[4] J-F. Cardoso. Source separation using higher order moments. In *Proc. of Int'l Conf. on Acoust. Speech and Signal Proc.*, pages 2109–2112, 1989.

[5] K.-T. Fang, S. Kotz, and K. W. Ng. *Symmetric multivariate and related distributions*. Chapman and Hall, London, 1990.

[6] V. Koivunen, P. Pajunen, J. Karhunen, and E. Oja. Blind separation from ϵ-contaminated mixtures. In *EUSIPCO'98*, pages 1817–1820, 1998.

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

[8] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, London, 1979.

[9] J. Möttönen and H. Oja. Multivariate spatial sign and rank methods. *Nonparametric Statistics*, 5:201–213, 1995.

[10] E. Oja. The nonlinear pca learning rule in independent component analysis. *Neurocomputing*, 17:25–45, 1997.

[11] C. Small. A survey of multidimensional medians. *International Statistical Review*, 58(3):263–277, 1990.

[12] S. Visuri. Robust covariance matrix estimation and its applications to signal processing. Master's thesis, University of Oulu, 1998.

[13] S. Visuri, V. Koivunen, and H. Oja. Sign and rank covariance matrices. Conditionally accepted to the *Journal of Statistical Planning and Inference*, 1999.