# BLIND SEPARATION FROM $\varepsilon$-CONTAMINATED MIXTURES

*V. Koivunen*[1] *P. Pajunen* *J. Karhunen* *E. Oja*

Laboratory of Computer and Information Science
Helsinki University of Technology
P.O. Box 2200, FIN-02150 Espoo,
Finland

**Abstract**
*This paper deals with the problem of Blind Source Separation (BSS). BSS algorithms typically require that observed data are prewhitened. The data are here assumed to be contaminated by highly deviating samples. Hence, covariance matrix used for whitening and determining the number of signals is estimated unreliably. We propose a method where data are first whitened in a robust manner. Sources are then separated using an iterative least squares algorithm. The proposed method is compared to a method based on sample estimates and the influence of outliers is analysed.*

## 1. INTRODUCTION

Blind source separation has important applications, e.g., in speech and array signal processing. In BSS, a collection of observed linear combinations (mixtures) of source signals are processed in order to find the unobservable sources. Most separation techniques make strict assumptions on the number of sources and mixtures as well as statistics of the signal. Highly deviating observations, i.e. outliers, may make these assumptions inaccurate: they may make the observed density appear asymmetric about the mean, inflate the variances, change the correlation structure as well as add new insignificant signal components to the data.

Commonly BSS methods require prewhitening that de-correlates and normalizes the observed data. Whitening is typically performed based on eigenanalysis of sample covariance matrix. Outliers cause error by attracting the mean towards them and perturb the eigenvalues and eigenvectors significantly. This also has implications for information theoretical criteria such as MDL (see [7]) that are used in estimating the number of signals.

This paper is organized as follows. The BSS problem is defined and a method for source separation is presented in section 2. The problems of prewhitening in the face of noise is addressed as well. In section 3, examples and quantitative results on whitening, estimating the number of signals and separating sources in $\varepsilon$-contaminated additive Gaussian noise are given.

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

## 2. BLIND SOURCE SEPARATION

BSS has been under busy investigation both in signal processing and in neural network communities (see e.g., [2, 5, 3, 1]). The unobservable sources and the observed mixtures are related by

$$\mathbf{x}_k = A\mathbf{s}_k + \mathbf{v}_k \tag{1}$$

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, $\mathbf{x}$ is a column vector of $n$ mixtures, $\mathbf{v}$ is additive noise vector and $k$ is time index. The matrix $A$ is assumed to be of full rank and sources are typically assumed to be zero mean and stationary. Blindness refers to the fact that no prior information about the mixing coefficients is available. All the sources are commonly assumed to have either positive or negative kurtosis. One of the sources may be Gaussian (zero kurtosis) if the noise are non-Gaussian and if the noise are Gaussian none of the sources may be Gaussian.

The separation task at hand is to estimate a separating matrix $H$ so that the original sources are recovered. Observed $\mathbf{x}$ are typically whitened prior to separation. Whitening allows for solving the separation problem easier 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 $\mathbf{y}$ of unknown sources $\mathbf{s}$ is given by

$$\hat{\mathbf{s}} = \mathbf{y} = \hat{H}^T\mathbf{x}. \tag{2}$$

The estimate can be obtained only up to a permutation of $\mathbf{s}$, i.e., the order of the sources may change. If $n > m$, the number of source signals may be estimated using criteria such as the MDL (see [7]). The data are first whitened and the number of signals is estimated based on robust covariance estimate. In this paper, the separation is done by employing a least squares algorithm.

### 2.1. Covariance estimation and whitening

Sample covariance matrix does not perform whitening reliably in the face of outliers because unreliable estimates of both the mean and covariance matrix are obtained. In this section, the influence of highly deviating observations on covariance matrix estimate and computed eigenvalues and eigenvectors is studied. Then a method for estimating the covariances reliably in the presence of outliers is presented. This robust approach stems from generalized maximum likelihood principle [4].

The noise $\mathbf{v}$ in (1) are assumed to be additive $\varepsilon$-contaminated distributed as

$$F = (1-\varepsilon)\, F_0 + \varepsilon\, G, \qquad (3)$$

where the actual distribution $F$ is a mixture of outliers with unknown distribution $G$ and the nominal noise distribution $F_0$. The fraction of outliers is $\varepsilon$ ($< 0.5$). It is important to note that additive noise is assumed to act on the observed *mixtures* instead of the sources.

Figure 1 illustrates some of the problems encountered in whitening in the presence of highly deviating observations. Points labeled with $a, b, c, d, e$ may be

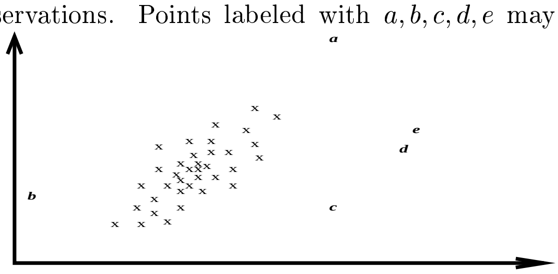


Figure 1: Highly deviating observations $a$,$b$,$c$,$d$,$e$ influence the estimated variances and correlations and consequently the eigenvalues and eigenvectors significantly.

considered outlying observations and they significantly influence the estimated covariance matrix. Point $a$ inflates the variances but has little effect on correlation. Point $b$ reduces the correlation and inflates the variance along the horizontal axis. Point $c$ has little effect on variance but reduces correlation. Points $b$ and $c$ add an insignificant dimension to data. If one draws an ellipse of equidistant points using Mahalanobis distance from the center of the data set, points $d$ and $e$ blow up this ellipse significantly. In addition, these observations attract the mean towards them and make the density appear asymmetric. Consequently, the eigenvalues and eigenvectors of the covariance matrix $C$ are influenced. The eigenvalue spread

$$\lambda_{max}(\widehat{C})/\lambda_{min}(\widehat{C}).$$

and the whole eigenvalue spectra of the estimated covariance matrix as well as the directions of the eigenvectors may drastically differ from true ones.

Whitening transform $W$ can be defined by terms of eigenvalues and eigenvectors of the covariance matrix $C$ as follows

$$W = \Lambda^{-1/2} U^T, \qquad (4)$$

where $\Lambda$ is diagonal matrix of the eigenvalues and $U$ is a matrix of eigenvectors. In case the noises can be assumed to be zero mean and Gaussian, techniques such as MDL estimate the number of signal components using the ratio of geometric to arithmetic mean of the eigenvalues [7]. Unreliable estimate results in both cases if the eigenvalues and eigenvectors are significantly perturbed.

Matrices $U$ and $\Lambda$ are obtained here by estimating covariance matrix iteratively so that highly deviating observations are downweighted. Estimates of the mean $\mu$ and covariance $C$ are computed as follows:

$$\widehat{\mu} = \frac{\sum_{i=1}^{n} w_i \mathbf{x}_i}{\sum_{i=1}^{n} w_i} \qquad (5)$$

$$\widehat{C} = \frac{\sum_{i=1}^{n} w_i^2 (\mathbf{x}_i - \widehat{\mu})(\mathbf{x}_i - \widehat{\mu})^T}{\sum_{i=1}^{n} w_i^2 \ - 1}. \qquad (6)$$

The weights $w_i$ are recomputed at each iteration using distances $d_i$

$$d_i^2 = (\mathbf{x}_i - \widehat{\mu})^T \widehat{C}^{-1} (\mathbf{x}_i - \widehat{\mu})$$

and the computation is iterated so that distances are computed using the current estimates of $\mu$ and $C$. The weights are computed by:

$$w_i = \begin{cases} 1 & \text{if } d_i = 0 \\ w(d_i) = \frac{sin(d_i/b)}{d_i/b} & \text{if } d_i \leq d_{thr} \\ 0 & \text{if } d_i > d_{thr}, \end{cases} \qquad (7)$$

where constant $b = d_0/\pi$ and $d_0 \geq d_{thr}$ are tuning parameters controlling the shape of the weighting function and the distance where the influence of an observation goes to zero. The weighting function is depicted in Figure 2. If $d_{thr} = d_0$ weights go smoothly to zero, otherwise they drop abruptly to zero at $d_{thr}$. If time ordering of data is of importance, outlying sample have to be substituted by a reasonable value so that no new signal components are introduced. Here they are replaced by a weighted sum of a few of its neighboring values. The weights $w(d_i)$ above are employed in the summation.

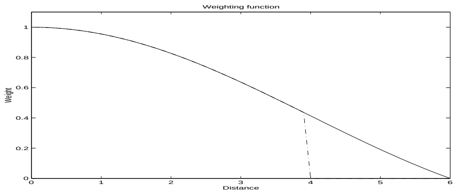


Figure 2: Two different weighting functions. The dotted line depicts a weighting function where $d_0 > d_{thr}$

### 2.2. Separating matrix estimation

An iterative least squares batch algorithm for source separation is employed. The relationship the between matrix $H$ and input $X$ may be written

$$X = HG$$

where $G = G(Y) = [g(y(1)) \ \cdots \ g(y(n))]$ and $H$ appears within $G$ as well. The error $\mathbf{e}$ is defined by

$$\mathbf{e} = \mathbf{x} - Hg(H^T\mathbf{x}) \tag{8}$$

where $g(\bullet)$ is an appropriate nonlinear function, for example, $tanh()$ for negatively kurtotic source signals. The least squares estimate of $H$ is obtained as follows [6]

$$\widehat{H} = XG^+,$$

where $G^+$ is the Moore-Penrose pseudoinverse of $G$. The cost function used above is a Bussgang equalization cost. In each iteration, values of $\mathbf{y} = \hat{H}^T\mathbf{x}$ needed for forming matrix $G$ are computed using the current estimate of $H$.

## 3. EXAMPLES

In this section, performance of the proposed BSS method is investigated in simulation. Example separation results are shown and analyzed quantitatively using MSE measure. Moreover, the robust whitening is analyzed quantitatively by comparing the eigenvalue spectra and directions of eigenvectors of estimated covariance matrix to theoretical values and values computed from sample covariance matrix. The number of signals is estimated using the MDL criterion. An example of separation from $\varepsilon$-contaminated mixtures is given in Figure 3. In our simulation, $m = 5$ source signals and $n = 7$ mixtures with randomly generated mixing coefficients in matrix $A$ are used. Each observed mixture of 500 samples is contaminated with zero mean additive Gaussian noise with variance $\sigma^2 = 0.4$. Moreover, 10% of the multivariate samples are randomly replaced by outliers with a large amplitude. The noisy data are first whitened which is followed by the separation. The number of signals was estimated using the MDL criterion. The robust method estimated consistently over different realizations that there are 5 signals components whereas the estimate based on sample covariance matrix was 4 in the example above and it varied between 4 and 6 depending on realization. Figure 4 shows the eigenvalue spectra obtained for the data shown in Figure 3.

Computation of the MSE between the recovered and original sources over large number of realizations is difficult because the order or sign of the sources may change in the separation. Consequently, one needs to find matching pairs of estimated and original sources and compute the MSE between pairs yielding the minimum error. The SNR between the true and estimated source signals $\mathbf{s}_i$ and $\mathbf{y}_i$ is given in terms of $MSE(\mathbf{s}_i, \mathbf{y}_i)$ as follows

$$SNR(\mathbf{s}_i, \mathbf{y}_i) = -10 \ log_{10} MSE(\mathbf{s}_i, \mathbf{y}_i).$$

SNR's are averaged over all components and over 10 realizations of 500 samples. The obtained SNR's are given in Table 1.

Table 1: Obtained mean and minimum SNR's using robust and conventional whitening. The mean SNR are computed over 10 realizations and over all sources.

| | Robust $\hat{C}$ | Sample $\hat{C}$ |
|---|---|---|
| mean SNR | 35.66 $dB$ | 29.48 $dB$ |
| min SNR | 33.55 $dB$ | 27.05 $dB$ |

Outlying observations tend to offset the mean and inflate the eigenvalues as well as rotate the eigenvectors of the covariance matrix. As a result, there is error in the whitening transform which applies scaling and rotation transformations to the observed data. The covariance matrices estimated from noisy data are compared to covariance matrix of noise-free mixtures based on eigenvalues and eigenvectors. Eigenvalues are compared using the ratio between the sum of eigenvalues and eigenvectors by determining the angles between corresponding eigenvectors. The angle between eigenvectors $\mathbf{u}$ and $\mathbf{v}$ is determined by $\angle(\mathbf{u}, \mathbf{v}) = cos^{-1}|\mathbf{u}^T\mathbf{v}|$. The eigenvalues are ordered such that $\lambda_1 > \lambda_2 > \cdots > \lambda_m \geq \cdots \geq \lambda_n$ and the related eigenvectors respectively. One needs to be careful in comparing the directions because the order of the eigenvectors may change due to outliers. Only the $m = 5$ eigenvectors associated with the signal subspace are used in the comparison. The results on eigendecomposition based on 500 samples and 10 realizations are given in Table 2. The performance of sample covariance estimator relative to robust estimator matrix is at worst for small contaminated samples. Both methods yield reasonably good estimates of the largest eigenvalue and the related

Table 2: The influence of outlying samples on eigenvalues and eigenvectors. The difference in eigenvalues is expressed using the ratio of the sum of eigenvalues of estimated and true covariance matrices and the difference in eigenvectors by the angle between the estimated and true eigenvectors.

| | Robust $\hat{C}$ | Sample $\hat{C}$ |
|---|---|---|
| mean $\sum \hat{\lambda}_i / \sum \lambda_i$ | 1.01 | 1.17 |
| max $\sum \hat{\lambda}_i / \sum \lambda_i$ | 1.04 | 1.31 |
| mean $cos^{-1}\lvert u_i^T \hat{u}_i \rvert$ | 3.0 degr. | 9.7 degr. |
| max $cos^{-1}\lvert u_i^T \hat{u}_i \rvert$ | 9.2 degr. | 40.1 degr. |

eigenvector. Sample $\hat{C}$ starts to produce poor results for subsequent smaller eigenvalues and related eigenvectors. In our experience, the proposed method has a performance close to optimal also in nominal conditions, i.e., if there are no outliers present in the data set.

## 4. CONCLUSION

The problem of BSS in the face of outliers was investigated. The observed data were whitened in a robust manner and sources were separated using an iterative least squares algorithms. The influence of outliers was analysed in detail. In separation, a significant improvement in SNR over a method based on sample estimates was achieved.

## 5. REFERENCES


[1] Bell, A.J., Sejnowski, T.J., "An information-maximisation approach to blind separation and blind deconvolution", Neural Computation, Vol. 7, No. 6., 1995, pp. 1129-1159.

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

[3] Comon, P., "Independent Component Analysis – A New Concept?", Signal Processing, Vol. 36, No. 3, 1994, pp. 287-314.

[4] Huber, P., "Robust Statistics", J. Wiley & Sons, 1981.

[5] Jutten, C., Herault, J., "Blind Separation of Sources, Part I: an Adaptive Algorithm Based on Neuromimetic Architecture", Signal Processing, Vol. 24, No. 1, 1991, pp. 1-10.

[6] Karhunen, J., Pajunen, P., "Blind Source Separation Using Least-Squares Type Adaptive Algorithms", Proc. of the Intl. Conf. on Acoustics, Speech, and Signal Processing (ICASSP-97), Vol. IV, pp. 3361-3364.

[7] Wax, M., Kailath, T., "Detection of Signals by Information Theoretic Criteria", IEEE T. Acoustics, Speech, and Signal Processing (T-ASSP), Vol. 33, No. 2, 1985, pp. 387-392.


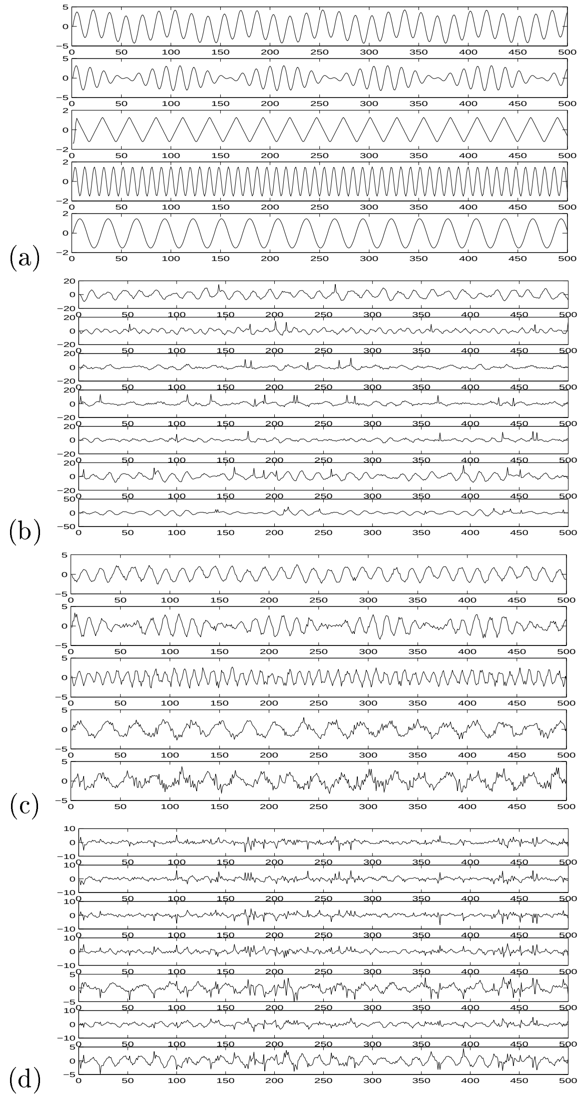


Figure 3: An example of BSS from noisy sequences: (a) Noise free source signals, (b) $\varepsilon$-contaminated mixtures, (c) separation result using robust whitening and (d) separation result using whitening based on sample covariance matrix.

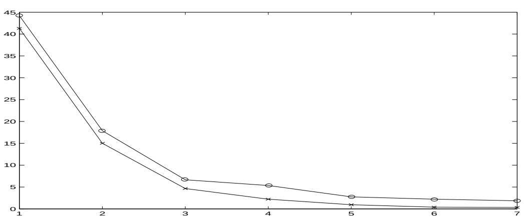

Figure 4: The changes in eigenvalue spectrum for the example in Fig. 3: $x$'s are the spectrum obtained using robust estimate and $o$'s using sample estimate of covariance.