BLIND SOURCE SEPARATION USING ORDER STATISTICS

Jani Even and Eric Moisan

Laboratoire des Images et des Signaux, groupe Non-Linéaire (UMR 5083)
ENSIEG, Domaine universitaire, BP 46, 38405 Saint Martin d'Hères Cedex , France
jani.even@lis.inpg.fr, eric.moisan@lis.inpg.fr

ABSTRACT

This paper shows the possibility to blindly separate instantaneous mixtures of sources by means of a criteria exploiting order statistics. Properties of higher order statistics and second order methods are first underlined. Then a brief description of the order statistics shows that they gather all these properties and a new criteria is proposed. Next an iterative algorithm able to simultaneously extract all the sources is developed. The last part is comparison of this algorithm with well known methods (JADE and SOBI). The most striking result is the possibility to exploit together independence and correlation owing to order statistics.

1. THE CONTEXT OF BLIND SOURCE SEPARATION

The problem of blind source separation is to estimate a set of signals called sources using observations of some mixtures of these signals [1]. The present work is devoted to the problem of instantaneous mixtures of N sources. Moreover the number of sensors (i.e. of observations) is equal to the number of sources.

Notation conventions:

$$\begin{bmatrix} s_1(t) \\ \vdots \\ s_N(t) \end{bmatrix} \Rightarrow \boxed{\text{A}} \Rightarrow \begin{bmatrix} v_1(t) \\ \vdots \\ v_N(t) \end{bmatrix} \Rightarrow \boxed{\text{M}} \Rightarrow \begin{bmatrix} \hat{s}_1(t) \\ \vdots \\ \hat{s}_N(t) \end{bmatrix}$$

The sources $\{s_i\}_{i\in[1,N]}$ are mixed by the constant matrix A (size $N \times N$). The estimated sources $\{\hat{s}_i\}_{i\in[1,N]}$ are obtained by applying the separation matrix M (size $N \times N$) to the sensor data $\{v_i\}_{i\in[1,N]}$. The goal of blind separation is to adjust the matrix M so that $\{\hat{s}_i\}_{i\in[1,N]}$ correspond to $\{s_i\}_{i\in[1,N]}$ up to a permutation and a scale factor on each source.

In the past few years, blind source separation has received a lot of interest. Two kinds of methods have emerged. On the one hand, higher order statistics can separate sources provided that at most one is Gaussian [2]. On the other hand, second order methods can separate signals as long as they have different autocorrelation functions [3]. Order statistics built on short realizations of a stochastic process accumulate information on both aspects, distribution and correlation. So, a criteria which exploits these statistics must permit to separate processes whose distributions are different and even Gaussian processes, provided that their autocorrelation functions are not the same.

Section 2 is a succinct reminder of usual methods. Section 3 presents briefly order statistics that are used to obtain a separation criteria in section 4. The proposed algorithm is given in section 5. Simulations results and comparison to two usual methods are shown in section 6.

2. EXISTING SEPARATION METHODS

This section reminds principles of usual methods to separate sources. It focuses on necessary assumptions and points out the cases where each method does not work.

2.1. Higher order statistics

A first approach is the use of statistical independence among the sources which can be exploited with higher order statistics [4] [5] [2] [6].

To achieve separation, these methods try to recover the statistical independence of the sources. This independence is measured by a function of higher order statistics. The restriction is that at most one source can have a Gaussian density [2], because if two sources are Gaussian, it is impossible to distinguish them by using their higher order statistics. In that case independence (equivalent to their decorrelation) is a necessary condition but it is not sufficient for separation.

2.2. Second order statistics

Another approach is based only on the second order statistics assuming that the sources have different autocorrelation functions [7] [3] [6].

An algebraic decomposition of covariance matrices of the observations for various delays gives an estimation of the mixing matrix A. The only condition required is that it is possible to choose a set of covariance matrices such that

their decompositions give diagonal matrices that are different. These covariance matrices for various delays are different if and only if the sources have different autocorrelation functions.

Therefore such methods are unable to separate processes whose correlations are the same, even when their distributions are different.

Tacking into account that order statistics built on a short observation set give information on both statistical distribution and correlation of the sources [8], a criteria based on order statistics must be able to perform the separation of mixtures, provided that the sources distributions are not Gaussian or that their autocorrelations are different.

3. ORDER STATISTICS

The order statistics of a discrete signal x are extracted from a sliding window of size T ($T \ll$ number of points of the signal). At time t, the T last points of the signal are available: $[x_t, x_{t-1}, \ldots, x_{t-T+1}]$.

After sorting these values we have a realization of the T order statistics:

$$x^{(t)}_{(1:T)} \leq x^{(t)}_{(2:T)} \leq \cdots \leq x^{(t)}_{(T:T)}$$

Let us denote $X_{(r:T)}$ the random variables:

$$X_{(1:T)} \leq X_{(2:T)} \leq \ldots \leq X_{(T:T)}$$

Let $P_{\underline{X}}(\underline{x})$ be the joint probability density function of the random variables $X_1, \ldots, X_T$ whose realizations are the $x_1, \ldots, x_T$ on a window of size T.

Then the probability density function of the r^{th} order statistic derives from its cumulative probability function:

$$\begin{aligned} F_{X_{(r:T)}}(u) &= proba\{X_{(r:T)} \leq u\} \\ &= \int \cdots \int_{\Delta_r} P_{\underline{X}}(\underline{x}) dx_1 \ldots dx_T \end{aligned} \quad (1)$$

where $\Delta_r = \{(x_1, \ldots, x_T)$ such that at least r variables among the x_i are in $]-\infty, u]\}$.

Then the probability density function is given by the derivative:

$$P_{X_{(r:T)}}(u) = \frac{\partial}{\partial u} F_{X_{(r:T)}}(u)$$

When the signal x is white, the $X_1, \ldots, X_T$ are independent identically distributed random variables and their probability density function can be factorised. Then the probability density function of the r^{th} order statistic (1) can be written (see [8] for detail):

$$P_{X_{(r:T)}}(u) = C \times F_X^{r-1}(u)[1 - F_X(u)]^{T-r} P_X(u) \quad (2)$$

with $F_X(u)$ and $P_X(u)$ the cumulative density function and probability density function of the stochastic process X, and $C = \frac{1}{B(r, T-r+1)}$

where $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ is the beta function.

For the white signal, equation (2) shows that the information on the law is present in the r^{th} order statistics. But it is obvious with equation (1) that both the correlation and the law of the X_i are present in the r^{th} order statistic because the joint probability density function of the X_i appears.

The problem is to exploit this information to separate independent sources.

4. THE CRITERIA

The main idea of this section was already introduced by Donoho for the deconvolution in [9].

Due to the central limit theorem, when sources are independent stationary processes, the densities of linear mixtures are closer to Gaussian than those of the sources.

Here our goal is to find estimated sources whose distributions are as far as possible of the Gaussian one. So we need a tool to measure a 'kind of distance' between the estimated sources distributions and the normal distribution.

We have to note that the blind separation of linear mixtures has two indeterminacies: A permutation of the sources cannot be detected and the power of the sources cannot be estimated. To cope with the power indeterminacy, the estimated signals are forced to have a unit power. So the density used as a reference is the normal density with zero mean and unit variance ($\mathcal{N}(0,1)$). Our major purpose is the use of the 'information' that is available in the order statistics which are non parametric estimators of the quantiles ($\mathcal{E}\{X_{(r:T)}\} = Q_X(\frac{r}{T+1})$). So the 'distance' will use the quantile function of the estimated sources and compare it to the quantile function of the normal distribution.

Previous works have already used the order statistics in this field. In [10], Pham used order statistics to estimate a discretized form of mutual information in order to perform the separation. In [11], a novel method to perform ICA trough order statistics is presented. A Gaussianity measure based on the cumulative density function is also introduced. The authors choose one specific order statistic that is discriminating for the extraction of one source and a deflation approach is used to achieve separation. We have also introduced a criteria that uses only extreme statistics (maximum and minimum) in [12] [13].

4.1. Extraction of one source

To estimate only one source $\hat{s}$ from the N mixtures, it is necessary to identify N coefficients m_i such that:
$\hat{s}(t) = \sum_{i=1}^{N} m_i v_i(t)$

Equation (3) gives the criteria to be minimized for the estimation of only one signal $\hat{s}$ from the N mixtures. The integral is approximated with order statistics.

$$\begin{aligned}\mathcal{J}_1(\hat{s}) &= -\int_0^1 [Q_{\hat{S}}(u) - Q_{\mathcal{N}}(u)]^2 du \\ &\approx -\frac{1}{T+1}\sum_{r=1}^{T}[Q_{\hat{S}}(\frac{r}{T+1}) - Q_{\mathcal{N}}(\frac{r}{T+1})]^2 \\ &\approx -\frac{1}{T+1}\sum_{r=1}^{T}[\widehat{\mathcal{E}\{\hat{S}_{(r:T)}\}} - Q_{\mathcal{N}}(\frac{r}{T+1})]^2 \quad (3)\end{aligned}$$

where $\widehat{\mathcal{E}\{\hat{S}_{(r:T)}\}}$ is the estimation of the expectation of $\hat{S}_{(r:T)}$ by empirical mean.

We can see that if $\hat{s}$ is Gaussian $\mathcal{J}_1$ asymptotically vanishes.

Experiments have shown that under the unit power constraint on $\hat{s}$ this quantity presents minima when it points to a source.

Particularly, there is a global minimum for one of the N sources if they have different distributions. It is necessary to add a constraint to this simple criteria $\mathcal{J}_1$ to point out simultaneously all N sources.

4.2. Extraction of all sources

Actually the algorithm used to recover all the sources is performing a spatial whitening along with the separation. It estimates as many sources as observed mixtures:
$\hat{s_i}(t) = \sum_{j=1}^{N} m_{ij} v_j(t)$ with $i \in [1, N]$.

It is natural to introduce the sum of the criteria $\mathcal{J}_1$ for all the estimated sources

$$\mathcal{J}_N = \sum_{i=1}^{N} \mathcal{J}_1(\hat{s}_i)$$

The addition of a new term $\mathcal{J}_\perp$, see equation (4), to the criteria $\mathcal{J}_N$ becomes necessary in order to perform the whitening procedure along with the separation.

$$\mathcal{J}_\perp = \frac{1}{2}[trace(R_{\hat{S}}) - log(det(R_{\hat{S}})) - N] \quad (4)$$

where $R_{\hat{S}}$ denotes the covariance matrix of the outputs.

The resulting criteria is $\mathcal{J}$

$$\mathcal{J} = \sum_{i=1}^{N} \mathcal{J}_1(\hat{s}_i) + \frac{1}{2}[trace(R_{\hat{S}}) - log(det(R_{\hat{S}})) - N] \quad (5)$$

The question is now to find the N^2 coefficients m_{ij} which minimize this criteria. This is achieved by an iterative algorithm.

5. RELATIVE GRADIENT OF THE CRITERIA AND ITERATIVE ALGORITHM

Given a first matrix M_0, a relative gradient descent is used to converge to the minimum by adapting the separation matrix M,

$$M_{k+1} = (I_N - \mu \, \nabla\mathcal{J}(M_k)) \times M_k$$

where the subscript k denotes the iteration's number, I_N is the identity matrix of size $N \times N$, μ is the adaptation step and $\nabla\mathcal{J}(M_k)$ is the relative gradient of the criteria with respect to M.

The relative gradient can be split in two parts:
$\nabla\mathcal{J}(M_k) = \nabla\mathcal{J}_\perp(M) + \nabla\mathcal{J}_N(M)_{antisym}$

A skew-symmetric projection of the relative gradient of $\mathcal{J}_N$, in equation (6), ensures that the process of minimizing $\mathcal{J}_N$ will not interfere with the orthogonalization (this is done in [4] [14]).

$$\nabla\mathcal{J}_N(M)_{antisym} = \frac{\partial\mathcal{J}_N}{\partial M}M^T - M\frac{\partial\mathcal{J}_N}{\partial M}^T \quad (6)$$

The gradient of $\mathcal{J}_N$ is:

$$\frac{\partial\mathcal{J}_N}{\partial M} = \left[2\sum_{r=1}^{T} D \times [\mathcal{E}\{\hat{S}_{i(r:T)}\} - Q_{\mathcal{N}}(\frac{r}{T+1})]\right] \quad (7)$$

where

$$D = \mathcal{E}\{V_{ij(r:T)}\}$$

In equation (7), $\mathcal{E}\{V_{ij(r:T)}\}$ is the mean of the values of the observation j corresponding with the r^{th} order statistic of the estimated source i.

$$\hat{S}_{i(r:T)} = \sum_{j=1}^{N} m_{ij} V_{ij(r:T)} \quad (8)$$

It is important to see that $V_{ij(r:T)}$ is not the order statistic of rank r of the observation j but the contribution of this observation to the r^{th} order statistic of the estimated source $\hat{S}_i$. The relation between $\hat{S}_{i(r:T)}$ and $V_{ij(r:T)}$ is given by equation (8).

The relative gradient of $\mathcal{J}_\perp$ (detailed calculus in [4]) is symmetric, as it can be seen in equation (9).

$$\nabla\mathcal{J}_\perp(M) = R_{\hat{S}} - I_N \quad (9)$$

So the relative gradient of the criteria $\mathcal{J}$ (see equation 5) is given by

$$\nabla\mathcal{J}(M) = R_{\hat{S}} - I_N + \frac{\partial\mathcal{J}_N}{\partial M}M^T - M\frac{\partial\mathcal{J}_N}{\partial M}^T \quad (10)$$

The main points of the iterative algorithm used to minimize $\mathcal{J}$ are:

First, the separation matrix is initialized at $M_0 = \Lambda \times I_N$ where Λ is a diagonal matrix introduced for normalization of the outputs ($\lambda_{ii} = \frac{1}{\sqrt{\mathcal{E}\{v_i^2\}}}$).
A step $\mu > 0$ is chosen.
Then the matrix M is updated until convergence by the following actions:
1.Estimation of all sources at step k by:
$\hat{S} = M_k \times V$
2.Calculus of the gradient $\nabla\mathcal{J}(M_k)$ for estimated sources $\hat{S}$
3.Adaptation of M:
4. $M_{k+1} = (I_N - \mu\,\nabla\mathcal{J}(M_k)) \times M_k$
5. Test of convergence:
If $\|M_{k+1} - M_k\| > \epsilon$, return to 1

This iterative algorithm minimizes $\mathcal{J}$, and at convergence, the matrix M_{final} performs the separation. Assuming that unknown sources s_i have unit power, the product $M_{final} \times A$ is expected to be a permutation matrix.

6. SIMULATIONS

6.1. Simulations parameters

Extensive numerical simulations have been done in order to study the algorithm's behavior. The synthetic sources are noises with various distributions and correlations.

Here is an example where the proposed algorithm is compared to well known methods: SOBI that performs separation using only second order statistics [3] and JADE that uses higher order statistics [5].

6.1.1. Performance measures

In the present case since sources and mixing matrix A are simulated, we can use this knowledge to evaluate the performances of the separation. In blind context, the covariance matrix of estimated sources should be used.

A measure of the effectiveness of the separation is provided by the ratio of residual power to the power of the estimated source for each output $\hat{s}_i$. The product $M_{final} \times A$ gives a matrix $B = \{b_{ij}\}$. Taking into account the unit power constraint of the output $\hat{S}_i$, $\sum_{j=1}^{N} b_{ij}^2 = 1$. The source S_l which has the largest power contribution to the output $\hat{S}_i$ is such that: $max_j\{b_{ij}\}^2 = b_{il}^2$. Then the power contribution of the other sources to the same output $\hat{S}_i$ is: $(\sum_{j=1}^{N} b_{ij}^2) - b_{il}^2 = 1 - max_j\{b_{ij}\}^2$.

So we choose to measure the performance by the following ratio computed for each estimated source

$$P(\hat{s}_i) \quad = \quad \frac{1 - max_j\{b_{ij}\}^2}{max_j\{b_{ij}\}^2}$$

The N residual powers only give information on the fact that an estimated source is close or not to a source. Moreover, in order to recover all N sources, the product $M_{final} \times A$ must look like a permutation matrix. This last point can be examined in the simulations because real mixing matrix A is known.

6.1.2. Data

Four zero mean unit variance signals are used as sources in the following example. The first three sources are white: s_1 has a Gaussian distribution, s_2 a Uniform distribution and s_3 a Laplacian distribution. Fourth signal s_4 is a Gaussian filtered with an $AR1$ low-pass filter ($s_4(t) = 0.7 \times s_4(t-1) + \epsilon(t)$ where $\epsilon(t) \rightsquigarrow \mathcal{N}(0,1)$). All the signals are 10000 points long.

Mixing matrix is:

$$A = \begin{bmatrix} 0.57 & 0.94 & 0.5 & 0.13 \\ -0.37 & 0.75 & 0.31 & -0.79 \\ 0.57 & 0.25 & -0.83 & -0.42 \\ -0.92 & 0.53 & -0.72 & 0.89 \end{bmatrix}$$

After convergence we can compare the products $M_{final} \times A$ for each method.

For the analyses of performance, 100 runs were made. The residual power are the mean of these 100 experiments. Whereas the product $M_{final} \times A$ is the one obtained by the last simulation, because the mean of the matrix $M_{final} \times A$ is not representative of the separation.

6.2. Results

6.2.1. SOBI

As expected, second order methods can only perform separation if the sources have different autocorrelation functions. These methods cannot separate s_1, s_2 and s_3. The only source that can be recovered is s_4.

The mean residual powers are (s_1 to s_4):
-9.9 dB, -8.84 dB, -9.32 dB and -77.6 dB

$$M_{final} \times A = \begin{bmatrix} 0.27 & 0.95 & 0.13 & 0.01 \\ -0.01 & 0.01 & 0 & -1 \\ 0.96 & -0.26 & -0.15 & -0.03 \\ -0.1 & 0.18 & -0.98 & 0.02 \end{bmatrix}$$

The 2^{nd} line of the matrix $M_{final} \times A$ shows that the 2^{nd} estimated signal has converged to the 4^{th} source. The three other outputs are still mixtures of the three other sources. They have been spatially whitened but not separated, since the resulting mixing matrix for these three signals is an orthonormal matrix.

6.2.2. JADE

Higher order methods are able to separate a mixture if at most one source is Gaussian. The two Gaussian distributed sources s_1 and s_4 cannot be separated with these methods.

The mean residual powers are:
-21.1 dB, -69.7 dB, -64.5 dB and -21.3 dB

$$M_{final} \times A = \begin{bmatrix} -0.63 & 0.02 & 0 & 0.79 \\ 0 & 1 & 0 & -0.01 \\ 0.01 & -0.01 & 1 & 0 \\ 0.77 & 0 & -0.02 & 0.62 \end{bmatrix}$$

The 2^{nd} and 3^{rd} lines of the matrix $M_{final} \times A$ show that the 2^{nd} and 3^{rd} estimated signals have converged to the 2^{nd} and 3^{rd} sources (the ones that have non Gaussian distribution). The two other output signals are still mixtures of the two Gaussian sources. They have been spatially whitened so that they are independent but they differ from the sources, since the resulting mixing matrix for these two signals is an orthonormal matrix.

6.2.3. Order Statistics

As expected, the use of order statistics, that give information on both correlation and distribution of the sources, enable the separation of the four sources.

The mean residual powers are:
-63.4 dB, -75.8 dB, -71.9 dB and -66.6 dB

$$M_{final} \times A = \begin{bmatrix} 0 & 1 & 0.01 & 0.01 \\ -0.01 & 0 & 0 & -1 \\ -0.03 & 0.02 & -1 & 0.01 \\ -1 & 0.01 & 0.04 & 0.03 \end{bmatrix}$$

Generally after 600 iterations, the matrix M_{final} has converged and the product $M_{final} \times A$ is close to a permutation matrix, so that each line points out one source.

To obtain such a result, it is necessary to adjust two parameters. As usual, the step μ must be positive and not too big to avoid divergence. The compromise is the same as in every adaptive algorithm: Speed of convergence versus variance after convergence. The significant parameter is the size T of the sliding window on which data are sorted.

6.3. The size of the sliding window

The first thing that appears is the possibility to separate sources that have different autocorrelation functions and those which have non Gaussian distributions present in a same mixture.

The new criteria was built to measure a "kind of distance to the Gaussianity", and thus focuses on the distributions of the signals. Moreover the use of order statistics as estimation of the quantile function makes the hypothesis that signals samples are independent (white processes). But for colored processes, the use of order statistics on short window introduces an information on the correlation. The criteria is able to separate two Gaussian sources with different autocorrelation by using this information.

Extensive numerical simulations have underlined the importance of the size T of the sliding window. With a large window, the approximation of the integral made in equation (3) is good. Two signals with different distributions are easily separated by the criteria. On the contrary, a small window size provides a poorer approximation of the integral but gives better results for Gaussian distributed sources that have different autocorrelations.

For mixtures where all sorts of sources appear a compromise must be done.

The three figures 1, 2 and 3 represent the evolution of the coefficients of the matrices $M_k \times A$ versus k. The rows of the matrices are individually plotted. These three simulations are done with the same data that the 100^{th} run of the example, but the size T of the window is different for each figure.

The first graph is for the uniform signal, the second the filtered Gaussian, the third the laplacian signal and the last is the white Gaussian.

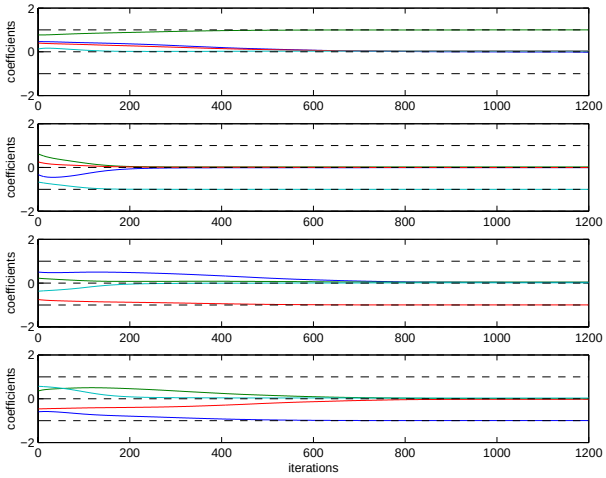

Fig. 1. Convergence of the coefficients of $M_k \times A$ with a sliding window of size $T = 7$

Figure 1 shows that the convergence is quicker for the colored source in the second graph, when a short sliding window is used.

Figure 2 shows that the large sliding window tends to favor sources with non Gaussian distribution (first and third graphs).

Compromise must be done by choosing an intermediate sliding window size, in figure 3, where T as been chosen equal to 17.

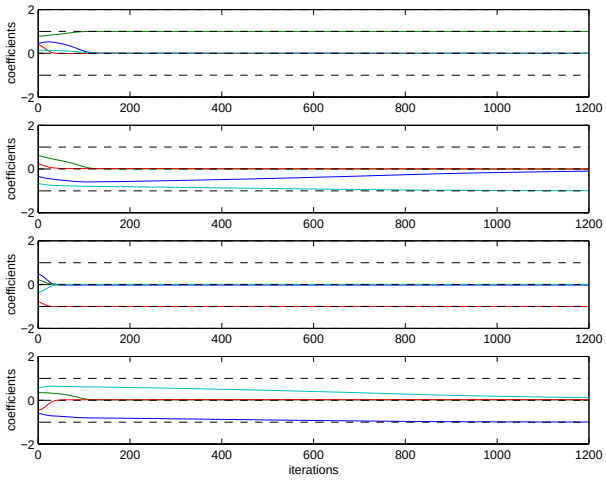

Fig. 2. Convergence of the coefficients of $M_k \times A$ with a sliding window of size $T = 35$

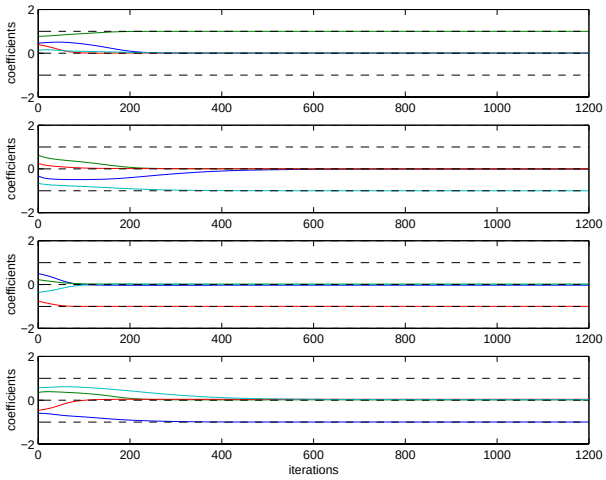

Fig. 3. Convergence of the coefficients of $M_k \times A$ with a sliding window of size $T = 17$

Conclusion

In this paper a new criteria based on order statistics is introduced. An iterative algorithm able to separate instantaneous mixtures is developed. The simulations and comparisons with usual methods show that order statistics allow to blindly separate instantaneous mixtures by exploiting both the distribution and the correlation of the sources. The justification of the role played by correlation is still a burning issue. The next point is to extend this work to the convolutive mixtures.

7. REFERENCES

[1] C. Jutten, *Calcul neuromimétique et traitement du signal, Analyse en Composantes Principales*, Ph.D. thesis, INPG, 1987.

[2] P. Comon, "Independent component analysis, a new concept ?," *Signal Processing*, vol. 36, 1994.

[3] A. Belouchrani, K. Abed-Meraim, and J.F. Cardoso, "A blind source separation technique using second order statistics," *IEEE Trans. on SP*, vol. 45, pp. 434–444, Feb. 1997.

[4] J.F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. on Sig. Proc.*, vol. 44, no. 12, pp. 3017–3030, 1996.

[5] J.F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[6] J.-L. Lacoume, P.O. Amblard, and P. Comon, *Statistiques d'ordre supérieur pour le traitement du signal*, Masson, 1997.

[7] A. Belouchrani, *Séparation autodidacte de sources: algorithmes, performances et applications à des signaux expérimentaux*, Ph.D. thesis, ENST, 1995.

[8] H.-A. David, *Order Statistics, second edition*, Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., 1981.

[9] D. Donoho, "On minimum entropy deconvolution," *Proceedings of the second Applied Time Series Symposium, 1980*, pp. 565–608, 1981.

[10] D.-T. Pham, "Blind separation of instantaneous mixture of sources based on order statistics," *IEEE Trans. on Sig. Proc.*, vol. 2, no. 48, pp. 363–375, 2000.

[11] Y. Blanco, S. Zazo, and J.C. Principe, "Alternative statistical gaussianity measure using the cumulative density function," *ICA 2000*, pp. 537–542, 2000.

[12] J. Even, "Séparation adaptative de sources à l'aide de statistiques d'ordre," Rapport de DEA, INPG, 2000.

[13] J. Even and E. Moisan, "Critère de séparation de sources basé sur les statistiques d'ordre," *Gretsi 2000*, 2000.

[14] J.F. Cardoso, "Blind signal separation: statistical principles," *Proceeding of the IEEE*, vol. 10, no. 9, pp. 2009–2025, 1998.