BLIND SOURCE EXTRACTION IN GAUSSIAN NOISE

Sergio Cruces†, Andrzej Cichocki§, Luis Castedo‡*

†Signal Processing Group, University of Seville, Spain. E-mail: sergio@viento.us.es
§ Lab. for Open Information Systems, Brain Science Institute, RIKEN, Japan. E-mail: cia@brain.riken.go.jp
‡Electronic and Systems, University of La Coruña, Spain. E-mail: luis@sol.des.fi.udc.es

ABSTRACT

In this paper we address the problem of recovering a subset of sources from a noisy linear mixture. We propose a novel blind source extraction algorithm that is robust with respect to the noise. This robustness is two-fold: on the one hand the algorithm does not lead to biassed estimates and, on the other, it minimizes the amount of signal and noise interference on the estimated sources. In addition, the asymptotic convergence of the algorithm to the extraction solution is demonstrated for almost any source distribution. Finally, the proposed algorithm is shown to be a generalization of the powerful kurtosis FAST-ICA algorithm that enables us to sequentially extract the sources in an arbitrary number of groups.

1. INTRODUCTION

Blind Source Separation (BSS) is the problem of recovering mutually independent unobserved signals (sources) from a linear mixture of them [1, 2, 3]. This problem has recently attracted a lot of interest because of its large number of applications in numerous fields. However, BSS can be very time consuming specially when the number of sources is large (for example, in biomedical signal processing). In these cases it is more practical to recover a subset of the sources only. This is known as Blind Source Extraction (BSE). When combined with a deflation procedure (i.e., the sequential subtraction of the extracted sources from the current observations) BSE algorithms are also able to sequentially extract all the independent sources, thus constituting an attractive alternative to conventional approaches that perform blind separation of sources.

The combined use of BSE and deflation to solve the BSS problem was first proposed in [5] and further explored in [6, 7]. However, in order to avoid replicated

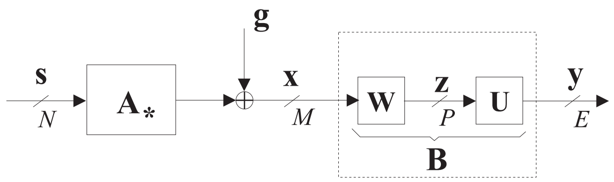

Figure 1: Signal model for blind source extraction.

outputs, the existing algorithms for BSE can only recover the sources sequentially one by one. This is also computationally expensive since it requires a number of deflation steps equal to the number of sources.

In this paper, we overcome the previous limitation and propose a BSE algorithm that, being robust with respect to Gaussian noise, allows us to simultaneously extract an arbitrary number of sources. In combination with deflation the algorithm enables us to sequentially extract the sources in an arbitrary number of groups. In addition, a version of this algorithm is shown to be a generalization of the kurtosis FAST-ICA algorithm [6]. Furthermore, simulations will show that, similarly to FAST-ICA, the proposed BSE algorithm usually extracts at the first stage the *most interesting* sources, i.e., those with largest absolute value of normalized kurtosis.

Let us consider the signal model shown in figure 1. A set of N sources $\mathbf{s} = [s_1[n], \ldots, s_N[n]]^T$ are combined through a linear memoryless system described by a full column rank $M \times N$ mixing matrix $\mathbf{A}_*$ in the presence of a jointly Gaussian and spatially uncorrelated noise process $\mathbf{g} = [g_1[n], \ldots, g_M[n]]^T$ to produce the observations vector

$$\mathbf{x} = \mathbf{A}_*\mathbf{s} + \mathbf{g} \; , \tag{1}$$

where $\mathbf{x} = [x_1[n], \ldots, x_M[n]]^T$ and $M \geq N$. In order to simultaneously extract E non-Gaussian sources (where $E \leq N$) the observations are multiplied by the extracting matrix $\mathbf{B}$ of dimensions $E \times M$ to obtain the

* This work has been partially supported by the Frontier Research Program of Japan. The algorithm MATLAB code is available at http://viento.us.es/˜sergio/BSEGN.html

outputs or estimated sources vector

$$\mathbf{y} = \mathbf{B}\mathbf{x} \tag{2}$$

where $\mathbf{y} = [y_1[n], \ldots, y_E[n]]^T$. The overall transfer matrix is given by $\mathbf{G} = \mathbf{B}\mathbf{A}_*$.

Denoting the autocorrelation matrix of the observations as $\mathbf{R}_{xx} = E[\mathbf{x}\mathbf{x}^H]$, we will consider the general case where the rank of $\mathbf{R}_{xx}$ is P with $N \leq P \leq M$. Let $\mathbf{W}$ be a matrix of dimension $P \times M$ that whitens the observations. Defining $\mathbf{z} = \mathbf{W}\mathbf{x}$ as the spatially decorrelated observations its autocorrelation matrix is $\mathbf{R}_{zz} = \mathbf{I}_P$, i.e., the identity matrix of rank P. Let us also consider that the extracting system $\mathbf{B}$ factors as $\mathbf{B} = \mathbf{U}\mathbf{W}$ where $\mathbf{U}$ is a $P \times E$ matrix with orthogonal columns.

Along the paper we will use the following notation. The sources vector splits in two disjoint parts $\mathbf{s} = [\mathbf{s}_L^T | \mathbf{s}_R^T]^T$ where $\mathbf{s}_L$ contains the E sources to be extracted and $\mathbf{s}_R$ represents the remaining $N-E$. Correspondingly, matrix $\mathbf{A}_*$ will be decomposed as $\mathbf{A}_* = [\mathbf{A}_{*L} | \mathbf{A}_{*R}]$ where $\mathbf{A}_{*L}$ and $\mathbf{A}_{*R}$ contain the columns associated to $\mathbf{s}_L$ and $\mathbf{s}_R$, respectively. $(\cdot)^+$ will denote the Moore-Penrose pseudo-inverse operator, $\mathbf{C}_{a,b}^{1,3}$ will denote the fourth order cross-cumulant matrix of elements $[\mathbf{C}_{a,b}^{1,3}]_{ij} = Cum(a_i[n], b_j^*[n], b_j[n], b_j^*[n])$ and $\mathbf{S}_a$ will be a diagonal matrix of cumulants signs $[\mathbf{S}_a]_{ii} = sign([\mathbf{C}_{a,a}^{1,3}]_{ii})$. We will also assume, without loss of generality, that the sources are properly scaled so that $\mathbf{C}_{s,s}^{1,3} = \mathbf{S}_s$.

The structure of the paper is as follows. Section 2 summarizes some previous results of a cumulant based algorithm for blind source separation, whereas Section 3 is devoted to the development of the proposed blind source extraction technique. Subsections 3.1 and 3.2 present two particular implementations of the blind extraction algorithm and their relations. In Subsection 3.3 we present the deflation step and the stability of the algorithm is analyzed in Subsection 3.4. Simulations of its performance are provided in Section 4. Finally, Section 5 presents the conclusions.

2. BLIND SOURCE SEPARATION IN GAUSSIAN NOISE

Let us start considering the conventional BSS problem, i.e., the simultaneous extraction of all the sources ($E = N$). Let us denote our current estimate of the mixing system as $\mathbf{A}^{(n)}$ and introduce the concept of *local robust estimate* to designate a matrix $\hat{\mathbf{A}}^{(n)}$ that provides an adequate direction for updating $\mathbf{A}^{(n)}$ towards $\mathbf{A}_*$. It was shown in [8, 9] that $\hat{\mathbf{A}}^{(n)} = \mathbf{C}_{x,y}^{1,3}\mathbf{S}_y$ is a *local robust estimate* of the mixing system. Since the iterative correction of the current estimate in the direction of the robust estimate leads to an estimation improvement, we can use the following algorithm [8] to estimate the mixing system

$$\mathbf{A}^{(n+1)} = \mathbf{A}^{(n)} + \mu^{(n)}(\mathbf{C}_{x,y}^{1,3}\mathbf{S}_y - \mathbf{A}^{(n)}) \tag{3}$$

Since the mixing matrix $\mathbf{A}$ is not square in general, we need only to estimate any full column-rank projection $\mathbf{P}_A\mathbf{A}_*$ of the mixing system to recover the sources. $\mathbf{P}_A = \mathbf{A}^{(n)}(\mathbf{A}^{(n)})^+$ represents the projection matrix onto the subspace spanned by the columns of $\mathbf{A}^{(n)}$. As a consequence, the BSS problem can be interpreted as the estimation of the pseudo-inverse of the mixing system. Rewriting recursion (3) in terms of $\mathbf{B}^{(n)} = (\mathbf{A}^{(n)})^+$ we arrive at

$$\mathbf{B}^{(n+1)} = (\mathbf{I}_N + \mu^{(n)}(\mathbf{C}_{y,y}^{1,3}\mathbf{S}_y - \mathbf{I}_N))^{-1}\mathbf{B}^{(n)} \tag{4}$$

where $\mathbf{I}_N$ is the $N \times N$ identity matrix. We will refer to this recursion as the Cumulant-based Iterative Inversion (CII) algorithm [8]. Note that any $\mathbf{B} = (\mathbf{P}_A\mathbf{A}_*)^+$ leads to the separation (i.e., $\mathbf{B}\mathbf{A}_* = \mathbf{I}_N$).

Taking into account that high-order cumulants of joint Gaussian processes are all equal to zero, we can state the following theorem

Theorem 1 *Even in presence of additive Gaussian noise in the mixture, the Cumulant-based Iterative Inversion (CII) algorithms presented in equations (3) and (4) asymptotically lead to unbiased estimates of the mixing and separating systems when a sufficient long data record is available to estimate the involved cumulants.*

3. BLIND SOURCE EXTRACTION: EXTRACTING SOURCES IN GROUPS

The key property that allows the extraction of part of the sources is that the iteration (3) can be decoupled independently for the left and right columns of matrix $\mathbf{A} = [\mathbf{A}_L | \mathbf{A}_R]$. This makes possible the estimation of the left part $\mathbf{A}_L$ without any knowledge of the right part $\mathbf{A}_R$. Indeed, when $E \neq N$ the iteration for $\mathbf{A}_L$ is given by

$$\mathbf{A}_L^{(n+1)} = \mathbf{A}_L^{(n)} + \mu^{(n)}(\mathbf{C}_{x,y}^{1,3}\mathbf{S}_y^3 - \mathbf{A}_L^{(n)}) \tag{5}$$

the outputs of the separation system

$$\mathbf{y} = \mathbf{B}^{(n)}\mathbf{A}_{*L}\mathbf{s}_L + \mathbf{B}^{(n)}\mathbf{A}_{*R}\mathbf{s}_R + \mathbf{B}^{(n)}\mathbf{g} \tag{6}$$

have three different components. The first component $\mathbf{B}^{(n)}\mathbf{A}_{*L}\mathbf{s}_L$ comes from the desired sources. The second term $\mathbf{B}^{(n)}\mathbf{A}_{*R}\mathbf{s}_R$ can be regarded as interference coming from the remaining sources. The third term $\mathbf{B}^{(n)}\mathbf{g}$ is the noise component at the outputs. When designing the separating system $\mathbf{B}^{(n)}$ our aim is:

1. Preserve the signals of interest at the outputs, i.e.,

$$\mathbf{B}^{(n)}\mathbf{A}_L^{(n)} = \mathbf{I}_E \quad (7)$$

2. Use the remaining degrees of freedom in $\mathbf{B}^{(n)}$ to minimize components at the outputs coming from the noise and the rest of sources.

There are two different ways to perform this process according to the existence or not of a prewhitening step.

3.1. Extraction without Prewhitening

Since the channel noise and perhaps part of the sources are Gaussian, an adequate solution to the previous problem is to minimize the Mean Square Error (MSE) between the estimated sources $\mathbf{A}_L^{+(n)}\mathbf{A}_{*L}\mathbf{s}_L$ and the outputs $\mathbf{y}$. But, under the minimum MSE criteria, this is tantamount to minimize the output power subject to some signal preserving constrains, i.e.,

$$\min_{\mathbf{B}} \; tr\{\mathbf{R}_{yy}\} \;\text{ subject to }\; \mathbf{B}^{(n)}\mathbf{A}_L^{(n)} = \mathbf{I}_E \quad (8)$$

We can solve this constrained minimization problem by means of the complex Lagrange multipliers and obtain

$$\mathbf{B}^{(n)} = (\mathbf{A}_L^{H(n)}\mathbf{R}_{xx}^{+}\mathbf{A}_L^{(n)})^{-1}\mathbf{A}_L^{H(n)}\mathbf{R}_{xx}^{+} \quad (9)$$

Then, in order to extract the sources, we only have to aternatively iterate equations (5) and (9), that constitute the kernel of the first implementation of the BSE algorithm. The algorithm recovers an arbitrary number $E < N$ of sources without any prewhitening step. This is done, however, at the extra cost of having to initially compute the pseudo-inverse of the correlation matrix of the observations. The problem of avoiding replicated sources at the outputs is considered in the following lemma.

Lema 1 *The blind extraction algorithm formed by equations (5) and (9) is unstable for that extraction solutions where the sources appear replicated at the outputs.*

Proof: This is a direct consequence of the fact that the signal preserving constraint $\mathbf{B}^{(n)}\mathbf{A}_L^{(n)} = \mathbf{I}_E$ is impossible to fulfil when $\mathbf{A}_L^{(n)}$ is not full column rank. We can observe that due to the matrix inversion present in equation (9), as long as some columns of $\mathbf{A}_L$ become dependent (which is a necessary condition for having replicated outputs), the inverse term $(\mathbf{A}_L^{H(n)}\mathbf{R}_{xx}^{+}\mathbf{A}_L^{(n)})^{-1}$ sharply diverges and, therefore, the proposed BSE iteration cannot converge to such deceptive solutions.□

3.2. Extraction with Prewhitening

Let us consider now the case of including an initial prewhitening stage in the algorithm. Defining $\mathbf{V}_L = \mathbf{W}\mathbf{A}_L$ as the left part of the overall estimated mixing after prewhitening, we can multiply equation (5) by $\mathbf{W}$ to obtain

$$\mathbf{V}_L^{(n+1)} = \mathbf{V}_L^{(n)} + \mu^{(n)}(\mathbf{C}_{z,y}^{1,3}\mathbf{S}_y - \mathbf{V}_L^{(n)}) \quad (10)$$

Similarly as before, the desired extraction matrix is also a solution of the following constrained minimization problem

$$\min_{\mathbf{U}} \; tr\{\mathbf{R}_{yy}\} \;\text{ subject to }\; \mathbf{U}^{(n)}\mathbf{V}_L^{(n)} = \mathbf{I}_E \quad (11)$$

Since now $\mathbf{R}_{zz} = \mathbf{I}$, the minimum MSE solution is simply given by

$$\mathbf{U}^{(n)} = (\mathbf{V}_L^{(n)})^{+} \quad (12)$$

where $\mathbf{V}_L^{+} = (\mathbf{V}_L^{H}\mathbf{V}_L)^{-1}\mathbf{V}_L^{H}$. Equations (10) and (12) are the kernel of a second implementation (presented in Table 1) of the BSE algorithm.

The following lemma illustrates the relation between the two previous approaches to BSE:

Lema 2 *The BSE algorithm with prewhitening given by equations (10) and (12) is a particular case of the BSE algorithm without prewhitening given by (5) and (9). Moreover, both are equivalent when the initialization for the estimated mixing system is given by* $\mathbf{A}_L^{(0)} = \mathbf{W}^{+}\mathbf{V}_L^{(0)}$.

Proof: The demonstration of this lemma comes from the fact that the inverse of the correlation matrix $\mathbf{R}_{xx}^{+}$ can be factorized as $\mathbf{R}_{xx}^{+} = \mathbf{W}^{H}\mathbf{W}$. Substituting this result in equation (9) and taking into account that $\mathbf{V}_L^{(n)} = \mathbf{W}\mathbf{A}_L^{(n)}$ we obtain

$$\begin{aligned}\mathbf{B}^{(n)} &= (\mathbf{W}\mathbf{A}_L^{(n+1)})^{+}\mathbf{W} \\ &= \mathbf{U}_L^{(n)}\mathbf{W} \quad (13)\end{aligned}$$

which proves the lemma since $\mathbf{G}_L^{(n)} = \mathbf{B}^{(n)}\mathbf{A}_{*L} = \mathbf{U}_L^{(n)}\mathbf{V}_{*L}$.□

When considering the extraction of a single source we obtain the following interesting lemma.

Lema 3 *The blind source extraction algorithm of equations (10),(12) is a generalization of the kurtosis FAST-ICA algorithm developed by Hyvarinen and Oja in [6] for the simultaneous extraction of E sources.*

Proof: Since we have previously demonstrated that equations (10),(12) enable us to simultaneously extract several sources, it is readily seen that the proposed algorithm reduces to the kurtosis FAST-ICA algorithm for a single source extraction. Indeed, let us observe that when $E = 1$, the BSE algorithm of equations (10),(12) simplify to

$$\begin{aligned} \mathbf{V}_L^{(n+1)} &= \mathbf{V}_L^{(n)} + \mu^{(n)}(\mathbf{C}_{z,y}^{1,3}\mathbf{S}_y - \mathbf{V}_L^{(n)}) \quad (14) \\ \mathbf{U}^{(n+1)} &= \mathbf{V}_L^{T(n+1)}/\|\mathbf{V}_L^{(n+1)}\|_2^2 \quad (15) \end{aligned}$$

and, for $\mu^{(n)} = 1$, the proposed algorithm matches the well-known kurtosis FAST-ICA algorithm [6]. There are some minor and not important differences: the normalization term in the FAST-ICA algorithm $\|\mathbf{V}_L^{(n+1)}\|_2$ appears squared in (15), and the scalar cumulant sign $\mathbf{S}_y$ in equation (14) does not exist in the FAST-ICA algorithm.□

3.3. Deflation

Once the first set of sources has been estimated, we can try to perform a new group extraction by previously removing the recovered sources from the mixture. This deflation process is obtained from the orthogonal projection operator onto the complementary space to the columns of $\mathbf{V}_L = \mathbf{W}\mathbf{A}_L$. The orthogonal projector given by the matrix $\mathbf{P}_\perp = \mathbf{I}_P - \mathbf{P}_V = \mathbf{I}_P - \mathbf{V}_L^{(n)}\mathbf{U}_L^{(n)}$ can be applied to the whitened observations. Thus,

$$\begin{aligned} \mathbf{z_d} &= \mathbf{P}_\perp \mathbf{z} \\ &= \mathbf{z} - \mathbf{V}_L^{(n+1)}\mathbf{y} \quad (16) \end{aligned}$$

performs the deflation. Analogously, the deflation takes the form

$$\mathbf{x_d} = (\mathbf{I}_M - \mathbf{A}_L^{(n)}(\mathbf{A}_L^{(n)})^+)\mathbf{x} \quad (17)$$

for the algorithm without prewhitening.

3.4. Asymptotic Convergence

The following theorem summarizes the results of the convergence analysis for the BSE algorithm with prewhitening.

Theorem 2 *As long as the fourth order cumulants of the sources are non zero, the only necessary and sufficient asymptotic stability condition for the following BSE algorithm*

$$\begin{aligned} \mathbf{V}_L^{(n+1)} &= \mathbf{V}_L^{(n)} + \mu(\mathbf{C}_{z,y}^{1,3}\mathbf{S}_y - \mathbf{V}_L^{(n)}) \quad (18) \\ \mathbf{U}^{(n+1)} &= (\mathbf{V}_L^{(n+1)})^+ \quad (19) \end{aligned}$$

to converge to an extraction solution is given by

$$\mu < \frac{1}{2} \quad (20)$$

Proof: In order to analyze the algorithm convergence and its asymptotic behaviour at the extraction solution we define the following deviation

$$\epsilon^{(n)} = \mathbf{V}_L^{H(n)}(\mathbf{V}_{*L}^H)^+\mathbf{D}^{(n)} - \mathbf{I}_E \quad (21)$$

where $\mathbf{V}_* = \mathbf{W}\mathbf{A}_*$ and $\mathbf{D}^{(n)}$ is a diagonal matrix with elements $D_{ii}^{(n)} = e^{-j\angle[\mathbf{V}_L^{H(n)}(\mathbf{V}_{*L}^H)^+]_{ii}}$. It is clearly seen that at the extraction solution ($\mathbf{V}_L = \mathbf{V}_{*L}\mathbf{D}$) the deviation term ϵ will converge to zero. The purpose of the matrix $\mathbf{D}^{(n)}$ is to remove the phase indeterminacy of the direct terms at convergence. Therefore, $\epsilon^{(n)}$ will always have real and positive diagonal terms.

Defining $\bar{\epsilon}^{(n+1)} = \mathbf{V}_L^{H(n+1)}(\mathbf{V}_{*L}^H)^+\mathbf{D}^{(n)} - \mathbf{I}_E$ (whose diagonal terms are no longer real), and taking into account iteration (10) we obtain

$$\bar{\epsilon}^{(n+1)} = (1-\mu)\epsilon^{(n)} + \mu\left(\mathbf{S}_y\mathbf{C}_{y,z}^{3,1}(\mathbf{V}_{*L}^H)^+ - \mathbf{I}_E\right)\mathbf{D}^{(n)} \quad (22)$$

The second term in the right side can be further simplified using the properties of the cumulants, i.e.,

$$\mathbf{C}_{y,z}^{3,1} = [{}^3\mathbf{G}_L^{(n)}|{}^3\mathbf{G}_R^{(n)}]\mathbf{C}_{s,s}^{3,1}\mathbf{V}_*^H \quad (23)$$

where $\mathbf{C}_{s,s}^{3,1} = \mathbf{S}_s$ by hypothesis and $({}^3\mathbf{G}_L^{(n)})$ denotes the Hadamard product $\mathbf{G}_L^{(n)} \times \mathbf{G}_L^{*(n)} \times \mathbf{G}_L^{(n)}$. Due to the orthogonality between the columns of $\mathbf{V}_* = [\mathbf{V}_{*L}|\mathbf{V}_{*R}]$ we observe that $\mathbf{V}_*^H(\mathbf{V}_{*L}^H)^+$ simplifies to $[\mathbf{I}_E|\mathbf{0}]$. Then, substituting (23) in (22) yields

$$\bar{\epsilon}^{(n+1)} = (1-\mu)\epsilon^{(n)} + \mu\left(\mathbf{S}_y({}^3\mathbf{G}_L^{(n)})\mathbf{S}_{s_L} - \mathbf{I}_E\right)\mathbf{D}^{(n)} \quad (24)$$

Now, let us analyze the asymptotic behaviour around the extraction solution. Let σ be a small perturbation matrix such that $\mathbf{V}_L^{(n)} = \mathbf{V}_{*L}\mathbf{D}^{(n)} + \sigma^{(n)}$. Then, from the definition of $\epsilon^{(n)}$

$$\epsilon^{(n)} = \sigma^{H(n)}(\mathbf{V}_{*L}\mathbf{D}^{(n)})^{H+} \quad (25)$$

we proceed to calculate the expansion of $\mathbf{G}_L^{(n)} = \mathbf{V}_L^{+(n)}\mathbf{V}_{*L}$ in terms of ϵ. With the help of equation (25) and after some straightforward calculus we obtain

$$\mathbf{G}_L^{(n)} = (\mathbf{I}_E - \epsilon^{H(n)})\mathbf{D}^{H(n)} + o(\epsilon) \quad (26)$$

In the neighborhood of the solution we have $\mathbf{S}_y = \mathbf{S}_{s_T}$. Also, since $diag(\epsilon^{(n)})$ is real, ${}^3(\mathbf{G}_L^{(n)})$ expands as

$${}^3(\mathbf{G}_L^{(n)}) = \left(\mathbf{I}_E - 3\mathrm{diag}(\epsilon^{(n)})\right)\mathbf{D}^{H(n)} + o(\epsilon) \quad (27)$$

Substituting (27) in (24) we obtain

$$\bar{\epsilon}^{(n+1)} = (1-\mu(1+3\mathrm{diag}(\epsilon^{(n)}))\epsilon^{(n)} + o(\epsilon) \quad (28)$$

Then, keeping the linear part of the expansion (which determines the asymptotic behaviour) and taking modulus yields

$$|\bar{\epsilon}_{ij}^{(n+1)}| \quad = \quad |1-\mu(1+3\delta_{ij})| \cdot |\epsilon_{ij}^{(n)}| \qquad (29)$$

for all $i,j=1,\ldots,E$ where δ_{ij} is the Kronecker delta. Since $\forall i,j$ the elements $[\mathbf{V}_L^{H(n+1)}(\mathbf{V}_{*L}^H)^{+}\mathbf{D}^{(n+1)}]_{ij}$ and $[\mathbf{V}_L^{H(n+1)}(\mathbf{V}_{*L}^H)^{+}\mathbf{D}^{(n)}]_{ij}$ share the same modulo and $\epsilon_{ii}^{(n+1)}$ is always real and positive, it is straightforward to show that $|\epsilon_{ij}^{(n+1)}| \leq |\bar{\epsilon}_{ij}^{(n+1)}|$. Then, we obtain

$$|\epsilon_{ij}^{(n+1)}| \quad \leq \quad |1-\mu(1+3\delta_{ij})| \cdot |\epsilon_{ij}^{(n)}| \qquad (30)$$

for all $i,j=1,\ldots,E$. From this result it is immediately seen that the necessary and sufficient asymptotic stability condition for convergence of the diagonal and off-diagonal terms is given by

$$\mu < \frac{1}{2} \qquad (31)$$

□

The step-size $\mu^{(n)}$ in equation (32) of Table 1 has been selected to fulfil this stability condition (see [10] for more details).

It is interesting to note that, similarly to CII algorithms [8] and in contrast with other BSE algorithms, the local stability condition derived from the theorem does not depend on the distribution of the sources as long as their fourth-order cumulants are non zero.

4. SIMULATIONS

In this section we present the results of a computer simulation to illustrate the performance of our proposed approach. We considered a total of 50 sources of two different classes. There are five interesting random binary sources ($\mathbf{s}=\pm 1$ with equal probability) located at positions $\{5,15,25,35,45\}$. The remaining 45 sources are all random uniform processes in the interval $[-1,1]$. The normalized kurtosis of the interesting sources is -2 whereas the normalized kurtosis of the remaining sources is $-6/5$. We mixed these sources by a 200×50 matrix of random uniform coefficients. Then, we added to the mixture a zero mean Gaussian noise with unity power obtaining a set of 200 observations. Note that the input signal to noise ratio is very low (only 0dB). The sample histograms of the different sources and channel noise can be observed in figure 2 whereas the sample histograms of the observations are shown in figure 3.

We applied the blind source extraction algorithm of Table 1 with parameters $\gamma = 5 \times E^2 \times 10^{-6}$, $\eta = 0.9$ and $p=1$, to estimate the five binary sources ($E=5$). After only 12 iterations the algorithm extracted the whole set of desired sources. The extraction can be checked in different ways: looking at the histograms of outputs which are shown in figure 4 or looking at the coefficients of the overall transfer function from the sources to each output which are presented in figure 5. From this latter figure it is readily seen that the transfer functions for each row are almost zero except at one of the places of the interesting source locations, i.e., at $\{5,15,25,35,45\}$.

This and other simulations show that the extracted sources order usually corresponds to a balance between the sources with largest absolute value of normalized kurtosis and the sources with an important initial component in the selected outputs. This fact is also in accordance with the convergence behaviour of FAST-ICA algorithms.

5. CONCLUSIONS

In this paper we have proposed a novel blind source extraction algorithm which is able to recover simultaneously an arbitrary number ($E < N$) of sources from a linear mixture. In the presence of noise in the observations, the algorithm minimizes the noise and signal interference on the recovered sources. Additionally, the algorithm is robust in the sense that for a Gaussian noise the obtained estimates are asymptotically unbiased. The algorithm is shown to be a generalization of the kurtosis FAST-ICA algorithm for blind source extraction. Finally, we have proved that the algorithm asymptotic convergence does not depend on the source distributions as long as the fourth order cumulants of the sources are not zero.

6. REFERENCES

[1] Cardoso, J.-F. (1998) Blind Signal Separation: Statistical Principles. *Proceedings of the IEEE* **86**(10):2009–2025.

[2] Amari, S.I. & Cardoso, J.F. (1997) Blind source separation – semiparametric statistical approach, *IEEE Transactions on Signal Processing* **45**(11):2692–2697.

[3] Jutten, C. & Herault, J. (1991) Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture, *Signal Processing*, **24**:1–10.

[4] Bell, A.J. & Sejnowski, T.J. (1996) An information maximization approach to blind separation and blind deconvolution, *Neural Computation* **7**:1129–1159.

Table 1: BSE algorithm with prewhitening.

1. Prewhitening: $\mathbf{z} = \mathbf{W}\mathbf{x}$

2. Initialization: $E \leq rank(\mathbf{R}_{zz})$, $\eta < 1$, γ, p, $\mathbf{V}_L^{(0)}$, $\mathbf{U}^{(0)} = \mathbf{V}_L^{+(0)}$ and $\mathbf{y} = \mathbf{U}^{(0)}\mathbf{z}$.

3. Step-size:

$$\mu^{(n)} = min\left\{\frac{2\eta}{1+3\eta}, \frac{\eta}{1+\eta\|\mathbf{C}_{y,y}^{1,3}\|_p}\right\} \qquad (32)$$

4. Estimation & extraction:

$$\begin{aligned} \mathbf{V}_L^{(n+1)} &= \mathbf{V}_L^{(n)} + \mu^{(n)}(\mathbf{C}_{z,y}^{1,3}\mathbf{S}_y - \mathbf{V}_L^{(n)}) \\ \mathbf{U}^{(n+1)} &= (\mathbf{V}_L^{(n+1)})^{+} \\ \mathbf{y} &= \mathbf{U}^{(n+1)}\mathbf{z} \end{aligned}$$

5. $n = n + 1$,
   UNTIL ($\|\mathbf{C}_{y,y}^{1,3}\mathbf{S}_y - \mathbf{I}_E\|_p < \gamma$) RETURN TO 3.

6. IF deflation
   STORE $\mathbf{y}$,
   $\mathbf{z} = \mathbf{z} - \mathbf{V}_L^{(n)}\mathbf{y}$,
   RETURN TO 2.
   ELSE END.

[5] Delfosse, N. & Loubaton, P. (1995) Adaptive blind separation of independent sources: A deflation approach, *Signal Processing*, **45**:59–83.

[6] Hyvarinen, A. & Oja, E. (1997) A fast fixed-point algorithm for independent component analysis, *Neural Computation*, **9**:1483–1492.

[7] Cichocki, A. & Thawonmas, R. & Amari, S. (1997) Sequential blind signal extraction in order specified by stochastic properties, *Electronics Letters*, **33**(1):64–65.

[8] Cruces, S. & Castedo, L. & Cichocki, A. (1999) Equivariant blind separation using cumulants, *Submitted to IEEE Transactions on Signal Processing.*

[9] Cruces, S. & Castedo, L. & Cichocki, A. (2000) Novel Blind Source Separation Algorithms Using Cumulants, *Proceedings of the ICASSP*, Istambul, Turkey.

[10] Cruces, S. (1999) *An unified view of blind source separation algorithms*, Ph.D. thesis, University of Vigo. Signal Processing Dept., Spain.

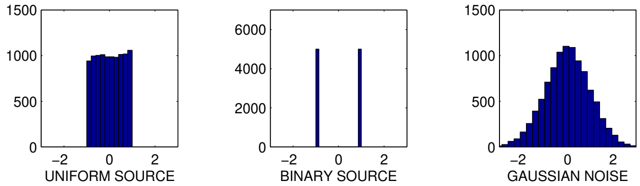

Figure 2: Histogram of the source signals and noise.

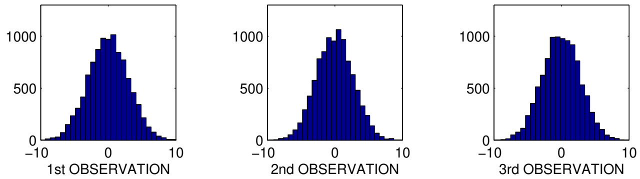

Figure 3: Histograms of some of the 200 observations.

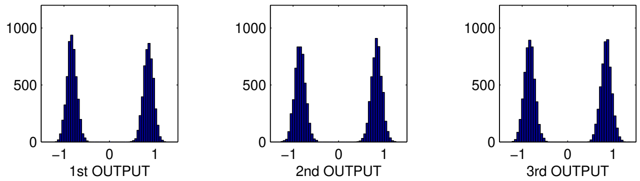

Figure 4: Histograms of the first recovered sources.

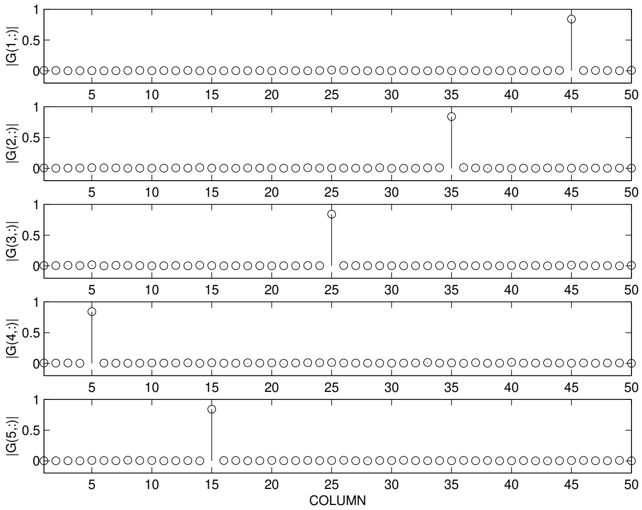

Figure 5: Overall transfer function from the sources to the outputs. The transfer function for each row is basically a delta function whose position determines the recovered source. Note from the delta positions $\{45, 35, 25, 5, 15\}$ that only the interesting sources has been recovered.