CONVOLUTIVE BSS USING A PENALIZED MUTUAL INFORMATION CRITERION

Mohammed El Rhabi, Guillaume Gelle, Hassan Fenniri, and Georges Delaunay,

DéCom-LAM Laboratory, University of Reims Champagne-Ardenne Moulin de la Housse B.P 1039 Reims cedex, France (Europe)
phone: +33 3 26918600, fax: +33 3 26913106, email: surname.name@univ-reims.fr

ABSTRACT

The Blind Separation problem of linear time dependent mixtures is addressed in this paper. We have developed a novel algorithm based on the minimization of the mutual information plus a penalized term which ensures an *a priori* normalization of the estimated sources (outputs). The criterion minimization is done using a well known gradient approach. Finally, some numerical results are presented to illustrate the performance of the penalized algorithm comparing to the Babaie-Zadeh approach presented in [4].

1. INTRODUCTION

In the last years, blind source separation (BSS) became a classical problem in signal processing due to the wide range of engineering applications that could benefit from such techniques. A general class of blind signal separation problem is the linear blind source separation where the mixing system is a linear time dependent (or not) function. Such a model is named convolutive mixture and the separation in regards with, convolutive BSS. The principle of BSS is to transform a multivariate random signal into an ideal signal which have mutual independent components in the statistical sense (see [1, 2]). So, BSS is achieved by maximizing the distance between the pdf of the ideal signal and the pdf of the multivariate observed signal. Note that in practice, the pdfs are unknown and must be estimated. It has been shown in [3] that this distance can be easily related to the maximization of a contrast function like the mutual information between the ideal signal and the observations.

This above mentioned approach (mutual information) is used by Babaie-Zadeh *et al* in [4]. The authors have proposed a new method to separate convolutive mixture based on the minimization of a delayed output mutual information where each mutual information term is minimized using the Marginal and the Joint score function. We propose here an extension of this method using a penalized mutual information criterion which carries out blind source separation regardless of the permutation problem and the scale indeterminacy. Moreover, this criterion allows us to use a direct gradient method without any constraint on the displacements and so an efficient optimization.

The paper is organized as followed. **Section 2** recalls the principle of BSS, and presents the model. **Section 3** introduces the mutual information and the penalized separation criterion. The algorithm is presented in **Section 4**. Finally, a discrete form of the criterion, then a stochastic form are presented with some numerical results illustrating this work in **section 5**.

This work was supported by The Champagne-Ardenne region.

2. PRINCIPLE OF CONVOLUTIVE BSS

The mixing model can be introduced as follows (in the noise free case):

$$x(t) = \mathscr{A} * s(t), \tag{1}$$

where $*$ denotes the convolutive product, $\mathscr{A}$ is the mixing operator, $x(t)$ the observation vector, and $s(t)$ the independent component source vector.

Then, the separating system is defined by :

$$y(t) = \mathscr{B} * x(t), \tag{2}$$

where the vector $y(t)$ is the output signal vector (estimated source vector) and $\mathscr{B}$ the separating operator and can be implemented as in figure 1

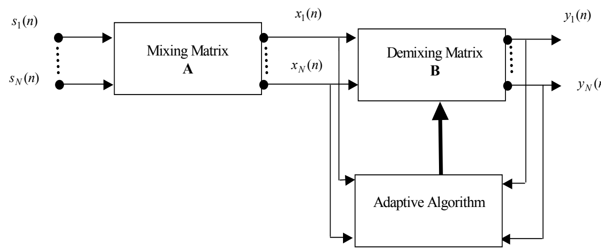

Figure 1: Mixing and separating systems.

In the discrete form, (3) and (4) become:

$$x(n) = [\mathscr{A}(z)]s(n) = \sum_k \mathscr{A}_k s(n-k), \tag{3}$$

and

$$y(n) = [\mathscr{B}(z)]x(n) = \sum_k \mathscr{B}_k x(n-k), \tag{4}$$

where $\mathscr{A}_k$ and $\mathscr{B}_k$ are respectively the corresponding $\mathscr{A}$ and $\mathscr{B}$ z-transform matrix.
If we assume $\mathscr{A}$ is left-invertible and statistically independent sources, then the problem consists in finding $\mathscr{B}$ and y for a given x such that:

$$y(n) = [\mathscr{B}(z)] * x(n) = [\mathscr{B}(z)][\mathscr{A}(z)]s(n), \tag{5}$$

where $\mathscr{B}$ satisfy $[\mathscr{B}(z)][\mathscr{A}(z)] = [\mathscr{P}\mathscr{H}(z)]$, and $\mathscr{P}$ is a permutation operator (and/or) $\mathscr{H}$ a *filtering* operator.

3. INDEPENDENCE CRITERION

Let $y = (y_1, ..., y_N)^T$ a random vector and consider p_y, the joint probability density function (joint pdf) and $p_{y_i}, i \in \{1, ..., N\}$, the marginal probability density function of the i^{th} component of y (marginal pdf).

In the BSS context, the mutual information can be written as follow:

$$I(y) = \int_{\mathbb{R}^N} p_y(t) \ln \left(\frac{p_y(t)}{\Pi_{i=1}^N p_{y_i}(t)} \right) \mathrm{d}t, \tag{6}$$

It is well known that (6) is nonnegative and equal to zero if and only if the components are statistically independent.

With convolutive mixtures, it is easy to show that the independence between two scalar sources $y_1(n)$ and $y_2(n)$ (for all n) are not sufficient to separate the system. That's why additional constraints must be stated to ensure the mutual independence of the output signal components $y_i(n)$, $i \in \{1,...,N\}$. To make it easier to understand, let us consider now a bidimensional random vector $y(n) = (y_1(n), y_2(n))^T$. The independence of the components $y_1(n)$ and $y_2(n')$ is needed for all n and n' to ensure the separation, in a different way the independence of $y_1(n)$ and $y_2(n-m)$, for all n and at all lags m.

Babaie-Zadeh *et al* take this last remark into consideration and propose in [4] the minimization of the following separation criterion:

$$J = \sum_q I(y^q(n)) \tag{7}$$

with $y^q(n) = (y_1(n-q_1), y_2(n-q_2), ..., y_N(n-q_N))^T$ when $y(n)= (y_1(n), y_2(n), ..., y_N(n))^T$ is the random output vector, and $q = (q_1, q_2, ..., q_N)^T$ an integer vector (with $q_1 = 0$). Actually, the separation is obtained when the components of y^q become independent. In order to prevent the convergence to the trivial solution and to overcome the scale indeterminacy, for each $k \leq N$, the k^{th} output component is subject to a normalization constraint. From a mathematical viewpoint, and in terms of separating matrix $\mathscr{B}$ can be written :

$$\mathscr{B} = (\mathscr{B}_0, ..., \mathscr{B}_p) \in \mathscr{M}_{N,(p+1)N} \simeq \mathbb{R}^{(p+1)N^2} \tag{8}$$

which means that each row of $\mathscr{B}$ belongs to the variety $V = \{L \in \mathbb{R}^{(p+1)N}$; energy of $L = 1\}$ of the euclidian space $\mathbb{R}^{(p+1)N}$. So, the Babaie-Zadeh *et al* method is based on a *tangential* gradient *i.e.* at each iteration, the matrix is scaled, and the displacement is done in the opposite direction of tangential gradient such that the rows of $\mathscr{B}$ remain on the variety V. The main drawback of this method is that the convergence of the algorithm could be affected by the normalization process.

In this paper we propose another approach which consists in overcoming the normalization constraint by adding a penalization term to the criterion. This allows us to use a direct gradient method without any constraint on the displacements toward the optimum and so a more efficient optimization. The numerical results presented at the end of this paper confirms this fact.

Moreover, one of the difficulties in the criterion (7) minimization is that the mutual information depends on the densities of the random variables, which must be estimated from the data. To minimize the criterion (7), several methods are proposed to estimate the pdf of the random variable. The approach in [5] consists in writing an expansion like Edgeworth or Gram-Charlier series of the pdf. Another approach is to calculate the (stochastic) gradient with respect to the separating matrix (see [6] and recently [4]). This approach points out the relevance of the score function (the log-derivative of the density defined below) of a random variable.

In this paper, we propose the following penalized criterion:

$$\begin{aligned}\tilde{J} = \textstyle\sum_q I(y^q(n)) + \\ \lambda \textstyle\sum_{i=1}^N \sum_q (E[(y_i^q(n) - E[y_i^q(n)])^2] - 1)^2,\end{aligned} \tag{9}$$

where $E[.]$ is the mathematical expectation and λ a positive parameter (the penalization parameter).

It is trivial to show that the criterion (9) reaches its minimum with normalized independent component outputs, since we choose $\lambda > 0$. The first term in the criterion is to minimize the criterion (7) of the estimated sources, while the second term demands that the output signal in each source have unit energy on average. In a sense, this constraint adds a signal normalization feature to the algorithm. In others words, this criterion overcomes the scale indeterminacy and prevents the algorithm from converging to the trivial solution ($y = \boldsymbol{0}$).

4. A PENALIZED ALGORITHM

In this section, we apply the gradient approach to separate convolutive mixtures based on the minimization of the criterion (9). To separate the sources by means of FIR filters with maximum degree p, the de-mixing system will be:

$$y(n) = \sum_{k=0}^{p} \mathscr{B}_k x(n-k), \tag{10}$$

where the infinite summation in (4) is replaced by a finite one.

To estimate the matrices $\mathscr{B}_k$ leading to estimate sources outputs, we calculate the gradients of $\tilde{J}$ with respect to each $\mathscr{B}_k$. So, we define the Joint Score Function (JSF), the Marginal Score Function (MSF) and the Score Function Difference (SFD) respectively by:

$$\begin{aligned}\varphi_y(y) &= \left(-\frac{\frac{\partial p_y(y)}{\partial y_1}}{p_y(y)}, ..., -\frac{\frac{\partial p_y(y)}{\partial y_N}}{p_y(y)}\right), \\ \psi_y(y) &= \left(-\frac{p'_{y_1}(y_1)}{p_{y_1}(y_1)}, ..., -\frac{p'_{y_N}(y_N)}{p_{y_N}(y_N)}\right), \text{ and} \\ \beta_y(y) &= \psi_y(y) - \varphi_y(y).\end{aligned}$$

4.1 The gradient

Let $\mathscr{B}_k$ a matrix, $\mathscr{E}$ a "small" matrix, to calculate the gradient with respect to $\mathscr{B}_k$ of $\tilde{J}$. We set $\hat{B}_k = \mathscr{B}_k + \mathscr{E}$ a matrix in a neighborhood of $\mathscr{B}_k$.

From (10), we have by definition:

$$\hat{y}(n) = [\hat{B}(z)]x(n) = y(n) + \mathscr{E}x(n-k),$$

Setting $h(n) = \mathscr{E}x(n-k)$, we have:

$$\hat{y}^q(n) = y^q(n) + h^q(n).$$

Then we can state the following proposition :

Proposition 1 *Let us consider $\tilde{J}$ defined by (9), then:*

$$\begin{aligned}\tilde{J}(\hat{y}^q(n)) - \tilde{J}(y^q(n)) = \langle \mathscr{E}, E\left\{\beta_{y^q}^{-q}(n) x(n-k)^T\right\}\rangle \\ \lambda \langle \mathscr{E}, E\left\{w(n)x(n-k)^T\right\}\rangle + o(\mathscr{E}),\end{aligned}$$

where $w = (w_1, ..., w_N)$ with $w_i = 4(E[y^{q2}_i] - 1)y_i$, $o(\mathscr{E})$ denotes higher order terms in $\mathscr{E}$ and $\langle C, D\rangle = trace(CD^T)$ is the matrix inner product.

Proof —See the appendix. $\diamond$

Formally, we get

$$\frac{\partial \tilde{J}(y^q(n))}{\partial \hat{B}_k} = E\left\{\beta_{y^q}^{-q}(n)x(n-k)^T + \lambda w(n)x(n-k)^T\right\}.$$

4.2 Algorithm

¿From the proposition (1), we derive the following algorithm:

- Step 1 : for $k = 0, ..., p$ and given $\hat{B}_k^0$

$$\hat{B}_k^n = \hat{B}_k^{n-1} - \mu \frac{\partial \tilde{J}}{\partial \hat{B}_k^{n-1}},$$

- Step 2: update y^n such that:

$$y^n = [\hat{B}^n(z)]x(n) = \sum_{k=0}^{p} \hat{B}_k^n x(n-k),$$

- Step 3: repeat until convergence.

5. NUMERICAL RESULTS

In this section, we deal with two observations obtained by a convolutive mixture of two sources. We give a comparison between the criterion (7) (as it is presented in [4]) and our criterion (9). We will use the separation criterion (9) in its discrete form, i.e the finite summation over $q_i \in \{-M, ..., M\}$ takes the place of the infinite one over $q_i \in \mathbb{Z}$, where $M = 2p$ (p is the maximum degree of the separating filters). Since this criterion is computationally expensive, we use its stochastic version. In other words, at each iteration, m is randomly chosen from the set $\{-M, ..., M\}$.

As performance criterion, we have used the output Signal to Noise Ratio (SNR) defined by:

$$SNR_i = 10\log_{10}\left(\frac{E[y_i^2]}{E[(y_i|_{\{s_i=0\}})^2]]}\right) \qquad (11)$$

where $y_i|_{\{s_i=0\}} = \{([\mathscr{B}(z)][\mathscr{A}(z)]s(n))_i\}_{s_i=0}$.

5.1 Example 1.

The mixing system is chosen as follows:

$$\mathscr{A}[z] = \begin{bmatrix} 1+0.2z^{-1}+0.1z^{-2} & 0.5+0.3z^{-1}+0.1z^{-2} \\ 0.5+0.3z^{-1}+0.1z^{-2} & 1+0.2z^{-1}+0.1z^{-2} \end{bmatrix}$$

The maximum degree of the FIR is equal to 2 ($p = 2, M = 4$). The number of observations is taken equal to 500, and the SFD are estimated using the Pham's method described in [7]. The experiment is repeated 50 times with different realizations of the random sources, the average standard deviation is ensured by the penalized algorithm (1.0009 for the first estimated source and 1.0005 for the second). The figure 2 shows the averaged SNRs versus iterations for the two algorithms, the adapting step-size is equal to $\mu = 0.3$ for the Babaie-Zadeh algorithm and to $\mu = 0.08$ for our algorithm (with a penalization parameter taken to $\lambda = 1$).

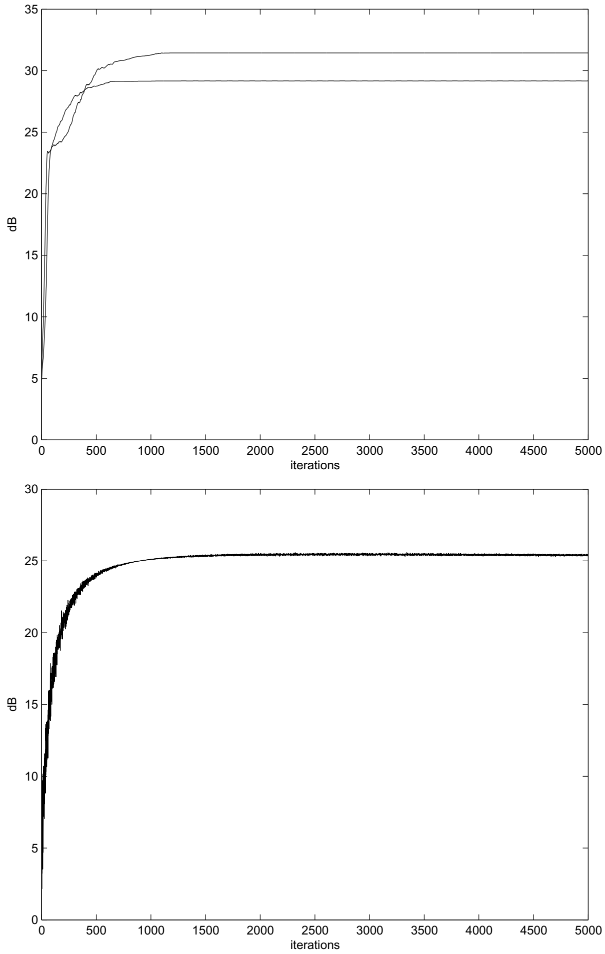

Figure 2: Averaged output SNRs versus iterations with Pham's estimation of the SFD : our penalized algorithm (at the top), Baibaie-Zadeh algorithm (at the bottom) for the mixing of two random signals (without permutations between the components).

5.2 Example 2.

Here, the mixing system is:

$$\mathscr{A}[z] = \begin{bmatrix} 1 & \mathscr{A}_{12}(z) \\ \mathscr{A}_{21}(z) & 1 \end{bmatrix},$$

where

$$\begin{aligned} \mathscr{A}_{12}(z) &= 0.01z^{-1} - 0.09z^{-2} + 0.05z^{-3} \\ &\quad -0.18z^{-4} + 0.59z^{-5} + 0.46z^{-6} \\ \mathscr{A}_{21}(z) &= 0.1z^{-1} - 0.15z^{-2} + 0.55z^{-3} \\ &\quad +0.42z^{-4} - 0.12z^{-5} + 0.04z^{-6} \end{aligned}$$

The score functions are estimated using the Pham's method, the maximum degree of the FIR is taken equal to 6 ($p = 6$, $M = 12$). The adapting step-sizes and the penalized parameter are taken as in the previous test. The average standard deviation is well kept (1.0000 for the first estimated source and 1.0002 for the second).

Figure 3 shows the averaged SNRs (taken also over 50 repetitions of the experiment) versus iterations for both algorithms.

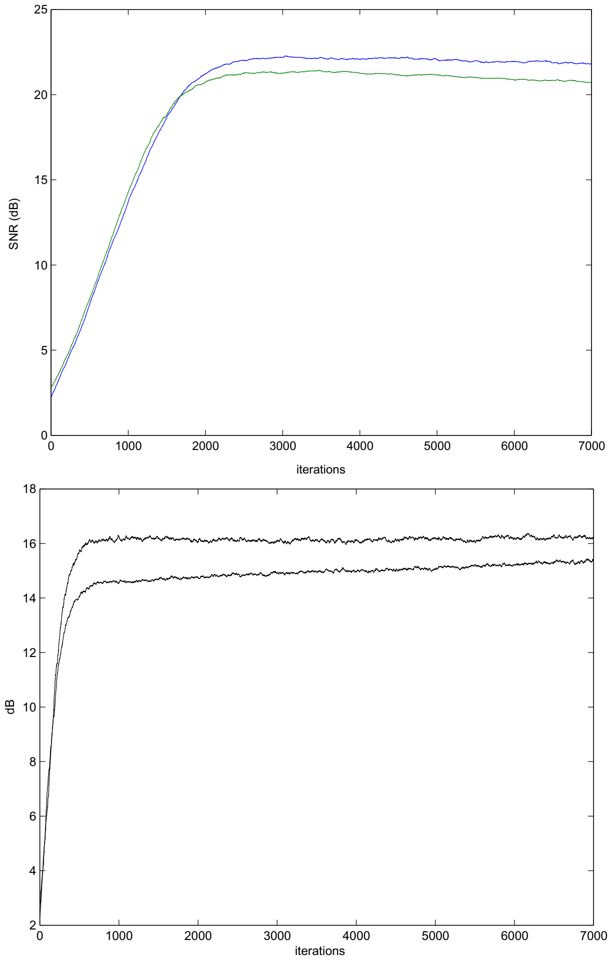

Figure 3: Averaged output SNRs versus iterations with Pham's estimation of the SFD : our penalized algorithm (at the top), Baibaie-Zadeh algorithm (at the bottom) for the mixing of two random signals.

6. CONCLUSION

A new convolutive BSS algorithm based on the minimization of a penalized mutual information criterion is presented. Unlike Babaie-Zadeh *et al* approach which uses implicitly a tangential gradient method, our algorithm is based on a global one due to the addition of a penalization term. So, the simulations results have shown that this implementation is proving to be much efficient in terms of stability and SNR's performance.

Appendix

Proof of the Proposition 1. Namely the following result

$$\begin{aligned}\tilde{J}(\hat{y}^q(n)) - \tilde{J}(y^q(n)) &= E\left\{\beta_{y^q}(n)^T h^q(n)\right\} + \\ &\quad 4\lambda E\left\{y^q(n)^T h^q(n)\right\} + \mathrm{o}(h^q), \\ &= E\left\{\beta_{y^q}^{-q}(n)^T \mathscr{E}x(n-k)\right\} + \\ &\quad \lambda E\left\{w(n)^T \mathscr{E}x(n-k)\right\} + \mathrm{o}(\mathscr{E}),\end{aligned}$$

We can find a proof of the first part of the right hand side in [4]. We propose to show the second part, we set:

$$\tilde{J}_2(y^q) = \sum_{i=1}^{N} \sum_{q_i \in \mathbb{Z}} (E[(y_i^q - E[y_i^q])^2] - 1)^2,$$

Without loss of generality, we assume : $E[y_i^q] = 0$, $\forall i \in \{1,\ldots,N\}$. First, we need the following lemma:

Lemma 1 *If y denotes a random process and h a "small" random process then*

$$(E[(y+h)^2]-1)^2 - (E[y^2]-1)^2 = 4\left(E[y^2]-1\right)E[yh] + o(h). \tag{12}$$

Proof—

$$\begin{aligned}&\forall i \in \{1,\ldots,N\}, \\ &(E[(y+h)^2]-1)^2 - (E[y^2]-1)^2 = \\ &\left(E[(y+h)^2] - 1 + E[y^2] - 1\right)\left(E[(y+h)^2] - E[y^2]\right) \\ &= \left(E[2y^2 + 2hy + h^2] - 2\right)\left(E[(y+h)^2 - y^2]\right) \\ &= \left(E[2y^2 + 2hy + h^2] - 2\right)\left(E[h(y+h)]\right) \\ &= \left(2E[y^2] + 2E[yh] - 2 + E[h^2]\right)\left(E[hy] + E[h^2]\right) \\ &= 4E[y^2]E[hy] - 4E[hy] + \mathrm{o}(h) \\ &= 4\left(E[y^2]-1\right)E[yh] + \mathrm{o}(h).\end{aligned} \tag{13}$$

◇

Applying the lemma to each component of y^q and h^q and by summation over i $\{1,\ldots,N\}$, we obtain :

$$\begin{aligned}\tilde{J}_2(\hat{y}^q) - \tilde{J}_2(y^q) &= \textstyle\sum_{i=1}^N 4\left(E[y_i^{q2}] - 1\right)E[y_i^q h_i^q] + \mathrm{o}(h_i^q) \\ &= \textstyle\sum_{i=1}^N 4\left(E[y_i^{q2}] - 1\right)E[y_i h_i] + \mathrm{o}(h_i), \\ &= E[w^T \mathbf{h}] + o(\mathbf{h}).\end{aligned} \tag{14}$$

which ends the proof. ◇

REFERENCES

[1] P. Comon, *Independent Component Analysis, a new concept ?*, Signal Processing, vol. 36, no. 3, pp. 287 – 314, 1994.

[2] A. Hyvarinen, J. Karhunen, E. Oja, *Independent Component Analysis*, Wiley, New York, 2001.

[3] Jean-François Cardoso, *Blind signal separation: statistical principles*, Proceedings of the IEEE, vol. 90, n. 8, pp. 2009-2026, Oct. 98, special issue on blind identification and estimation, R.-W. Liu and L. Tong editors.

[4] M. Babaie-Zadeh, C. Jutten and K. Nayebi, *Separating convolutive mixtures by mutual information minimization*, in Proceedings of IWANN2001, Granada, Spain, Juin 2001, pp. 834 – 842.

[5] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Distribution Theory, vol. 1, Griffin, 1977.

[6] A. Taleb and C. Jutten, *Entropy optimization, application to blind source separation*, in ICANN, Lausanne, Switzeland, October 1997, pp. 529 – 534.

[7] D.T Pham *Fast Algorithm for Estimating Mutual Information, Entropies and Score Functions*, Proceeding of ICA 2003 Conference , Nara, Japan April 2003.