Blind Source Separation using Penalized Mutual Information criterion and Minimal Distortion Principle

Mohammed El Rhabi, Hassan Fenniri and Georges Delaunay

Abstract—One of the most important problems in Blind Source Separation of convolutive mixtures is the filtering ambiguity. One way to address this is to modify the separation algorithm to enforce some constraint. In this paper, according to our previous studies recently introduced in [4], we propose to combine the Minimum Distortion Principle and the Mutual Information criterion in order to reduce the ambiguity filtering. It consists to choose an optimal separator among an infinite number of valid separators that can extract the source signals in a certain sense according to the Minimal Distortion Principle. Our simulations and experimental results confirm that these issues have substantial impact on the performance of the proposed BSS algorithm.

I. INTRODUCTION

Independent component analysis (ICA), introduced in [1], [2], and blind source separation (BSS, [3]) are related data analysis problems that have received considerable attention in the signal processing community during the last few years. Source separation consists in recovering a set of unobservable signals (sources) from a set of observed mixtures. In the convolutive blind source separation, each source can be recovered up a filtering, in addition with the well known permutation ambiguity. This indeterminacy may be unacceptable, since it can strongly distort the sources. So, a post-processing could be necessary to estimate the effect of each source on each sensor. In a precedent paper [4], we proposed a new algorithm based on a criterion where each mutual information term is minimized using the Marginal and the Joint score function regardless to the scale indeterminacy. In this contribution, we propose to use this algorithm taking account of the Minimal Distortion Principle introduced in [5] and which allows us to choose an optimal separator (in a sense define below) among an infinite number of valid separators that can extract the source signals. It is important to note that this optimal separator is uniquely determined and does not depend on the sources properties, it depends only on the mixing operator. Moreover, the proposed criterion allows us to use a direct gradient method without any constraint on the displacements and so an efficient optimization. Thus, our approach allows to restore directly the contribution of the sources to the sensor signals without post-processing. In addition, this proposed algorithm was applied to separate a mixture of i.i.d. sequences and harmonic sequences like the signals of vibrations resulting from rotating machines (in harmony with [6]).

The paper is organized as followed. Section 2 states the BSS problem, and presents the model. Section 3 introduces the mutual information and the proposed 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.

II. PROBLEM FORMULATION

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

\[ x(t) = A * s(t), \]

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

Then, the separating system is defined by :

\[ y(t) = B * x(t), \]

where the vector \( y(t) \) is the output signal vector (estimated source vector) and \( B \) the separating operator. The system can be implemented as in figure 1.

In the discrete form, (1) and (2) become:

\[ x(n) = [A(z)]s(n) = \sum_{k} A_k s(n - k), \]

\[ y(n) = [B(z)]x(n) = \sum_{k} B_k x(n - k), \]

where \( A_k \) and \( B_k \) are respectively the corresponding \( A \) and \( B \) z-transform matrices.

![Fig. 1. Block diagram of a separating system.](image-url)
If we assume $\mathcal{A}$ is left-invertible and statistically independent sources, then the problem consists in finding $\mathcal{B}$ and $\mathbf{y}$ for a given $\mathbf{x}$ such that:

$$y(n) = [\mathcal{B}(z)] \mathbf{x}(n) = [\mathcal{B}(z)] [\mathcal{A}(z)] \mathbf{s}(n),$$

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

III. SEPARATION CRITERION AND MINIMAL DISTORTION PRINCIPLE

Let $\mathbf{y} = (y_1, \ldots, y_N)^T$ a random vector and consider $p_{\mathbf{y}}$, the joint probability density function (joint pdf) and $p_{y_i}, i \in \{1, \ldots, N\}$, the marginal probability density function of the $i^{th}$ component of $\mathbf{y}$ (marginal pdf). In the BSS context, the mutual information can be written as follow:

$$I(\mathbf{y}) = \int_{\mathbb{R}^N} p_{\mathbf{y}}(\mathbf{t}) \ln \left( \frac{p_{\mathbf{y}}(\mathbf{t})}{\prod_{i=1}^N p_{y_i}(t_i)} \right) \, d\mathbf{t},$$

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 $n$) is not sufficient to separate the sources because we are dealing with random process and not random variable. That is why additional constraints must be stated to ensure the mutual independence of the output signal components $y_1(n), i \in \{1, \ldots, N\}$. To make it easier to understand, let us consider now a bidimensional random vector $\mathbf{y}(n) = (y_1(n), y_2(n))$. 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 $m$.

However, it was emphasized previously that it is possible to restore the source original up to a linear filtering. It is possible, however in the case of instantaneous mixtures to reduce the shape indeterminacy in the model (5) by setting a simple constraint on $\mathbf{x}$ where its components are generally assumed to have unit variance, naturally checked since the penalized term adds a signal normalization feature to the algorithm.

Some other simple constraints could be assumed, for example on the diagonal terms of the mixing matrix $[\mathcal{A}(z)]$ which are usually supposed to be equal to unity. This assumption can be easily related to the fact that the sensors are as close to the sources as possible [7]

Nevertheless, in the general case of more than two sources, this is not a sufficient condition and it is necessary to add other constraints to estimate the filtering indeterminacy effect of each source on each sensor. To circumvent this problem, a very attractive approach is proposed in [5] by K. Matsuoka et al. They introduced a principle, called the Minimal Distortion Principle which allows to choose an optimal separator among an infinite number of valid separators that can extract the source signals in a certain sense. The optimal choice is made such that the observed signals are the least subjected to distortion by the separator. Namely, a valid separator $[\mathcal{B}(z)]$ is defined as follow:

$$[\mathcal{A}(z)][\mathcal{B}(z)] = [\mathcal{P} \mathcal{D}(z)]$$

where $\mathcal{P}$ is a permutation matrix, $[\mathcal{D}(z)]$ is an arbitrary nonsingular diagonal matrix.

So, we can state the Minimal Distortion Principle as follows: the separator must be chosen such as its outputs become as close to the sensors outputs as possible.

In other words, the optimal separator $[\mathcal{B}(z)^{\text{opt}}]$ is the valid separator which minimizes:

$$E[\|\mathbf{y}(n) - \mathbf{x}(n)\|^2] = E[\|\mathcal{B}(z)[\mathbf{x}(n) - \mathbf{x}(n)]\|^2]$$

where $\| . \|$ represents the Euclidean norm.

So, taking this last principle into consideration, the criterion proposed in [4] leads to:

$$\mathcal{J}_{MD}(n) = \sum_q I(\mathbf{y}^q(n)) + \lambda \sum_{i=1}^N \sum_q (E[(y_i^q(n) - E[y_i^q(n)])^2] - 1)^2 + \gamma \sum_q E[\|\mathbf{y}^q(n) - \mathbf{x}^q(n)\|^2],$$

where $\mathbf{x}^q(n) = (x_1(n-q_1), x_2(n-q_2), \ldots, x_N(n-q_N))^T$ when the vector $\mathbf{x}(n)$ is the random observed vector and $\mathbf{q}$ is a vector of integers. It is trivial to show that the criterion (8) reaches its minimum with normalized independent component outputs which are close to the observation outputs, since we choose $\lambda > 0$ and $\gamma > 0$.

Indeed, the Minimal Distortion Principle forces the estimated sources to be as close as possible to its contribution on sensors, but the normalization process is not necessary ensured; nevertheless, this normalization is very important to avoid the algorithm from exploding, that is why, we retain the normalization term in the criterion (8).

IV. PROPOSED ALGORITHM

In this section, we apply the gradient approach to separate convolutive mixtures based on the minimization of the criterion (8). To separate the sources by means of FIR filters with maximum degree $p$, the de-mixing system will be:

$$\mathbf{y}(n) = \sum_{k=0}^p \mathcal{B}_k \mathbf{x}(n-k),$$

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

To approximate the matrices $\mathcal{B}_k$ leading to estimated source outputs, we calculate the gradients of $\mathcal{J}_{MD}$ with respect to each $\mathcal{B}_k$.

So, the derivation leads to multivariate score functions, namely the Joint Score Function (JSF), the Marginal Score
Function (MSF) and the Score Function Difference (SFD) defined respectively by:

\[
\varphi_y(y) = (-\frac{\partial p_y(y)}{p_y(y)}, \ldots, -\frac{\partial p_y(y)}{p_y(y)}),
\]

\[
\psi_y(y) = (-\frac{p_y(y)}{p_y(y)}, \ldots, -\frac{p_y(y)}{p_y(y)}),
\]

and \(\beta_y(y) = \psi_y(y) - \varphi_y(y)\) .

A. The gradient

Let \(B_k\) a matrix, \(E\) a “small” matrix, to calculate the gradient with respect to \(B_k\) of \(\tilde{J}_{\text{MD}}\). We set \(B_k = B_k + E\) a matrix in a neighborhood of \(B_k\).

From (9), we have by definition:

\[
y(n) = [B(z)x(n)]e(y(n)) + E(x(n) - k),
\]

Setting \(h(n) = E(x(n) - k)\), we have:

\[
y'(n) = y'(n) + h'(n)
\]

Then if we consider \(\tilde{J}_{\text{MD}}\) defined by (8), we have shown in [6] the following result:

\[
\tilde{J}_{\text{MD}}(y'(n)) - \tilde{J}_{\text{MD}}(y(n)) = \langle E, E \{\beta_{y'}(y) x(n-k)^T\} \rangle
\]

\[
+ \lambda E \{w(n)x(n-k)^T\} + 2\gamma E \{(y(n) - x(n))x(n-k)^T\} + o(E),
\]

where \(w = \{w_1, \ldots, w_N\}\) with \(w_i = 4\{E[y_i^2] - 1\}y_i\), \(o(E)\) denotes higher order terms in \(E\) and \((C, D) = \text{trace} CD^T\) is the matrix inner product (associated with the Schur norm).

It follows from the previous result:

\[
\frac{\partial J_{\text{MD}}(y'(n))}{\partial B_k} = E \{\beta_{y'}(y) x(n-k)^T\} + \lambda E \{w(n)x(n-k)^T\} + 2\gamma E \{(y(n) - x(n))x(n-k)^T\} .
\]

B. The algorithm

From (11), we derived the following algorithm:

for \(k = 0, \ldots, p\) and given \(\hat{B}_k\)

\[B_k = \hat{B}_k - \mu \frac{\partial J_{\text{MD}}}{\partial B_k},\]

update \(y^n\) such as:

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

return \(y^n\)

There are some practical problems associated with the implementation of this algorithm. It requires the estimation of the score functions (10) which are easily approximated by a polynomial or Pham’s approach (see [8]). We can notice that the cost calculus of this algorithm depends strongly on this approximation.

From a theoretical viewpoint, it is possible to handle convolutive mixtures with long impulse responses but the algorithm becomes very expensive and time-consuming.

V. Numerical results

Two examples are presented to illustrate the performances of our algorithm by comparison with the penalized algorithm (see [4]). The separation criterion (8) is used in its discrete form, i.e., the finite summation over \(q_i \in \{-M, \ldots, M\}\) 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 \(\{-M, \ldots, M\}\).

In order to illustrate the performance of the criterion (8), in the first example, we dealt with 3 observations obtained by a convolutive mixture of three random idd sources, non-gaussian and independent sources with zero means. The separation performance is given using the output Signal to Noise Ratio (SNR) defined by:

\[
\text{SNR}_i = 10 \log_{10} \left( \frac{E[y_i^2]}{E[(y_i|s_i=0)]^2} \right)
\]

where \(y_i|s_i=0 = \{(B(z))A(z)s(n))\}_{s_i=0}\)

The mixing operator \([A(z)]\) (minimal phase FIR filters) is randomly chosen. The length of the observation signals is equal to 3000 samples, and the SFD are estimated using the polynomial estimator. The experiment is repeated 20 times with different random sources realizations. We set unmixing filter length \(p = 12\) taps.

In the simulation, the adapting step is equal to \(\mu = 0.01\), the penalization parameters are equal to \(\lambda = 3\) and \(\gamma = 0.01\).

The figure 2 shows the averaged SNA’s versus iterations for the penalized algorithm coupled with the Minimal Distortion Principle. We clearly notice that the convergence of this algorithm is quite stable, and that it provides a good quality of separation (up to 25 dB). Note also that this algorithm is able to separate mixtures independently of the sources number.

In the same context, we try to separate only with the penalized algorithm (see [4] and figure 3); here also, we can see a quite good quality of separation but the SNR’s are lesser than the previous test (less than 22 dB). In addition, when we would like to separate mixtures more than 3 sources “we can observe this phenomenon more and more “. We can remark that in the 2D context, we have a similar separation.

In the second example, as we do not have a real experiment with more than two sources, we simulate 3 observations signals by an artificial convolutive mixtures of 3 real rotating machine vibrations signals, (see figure 4 and [6]). The mixing system A is randomly chosen. Each mixing filter is FIR of length 12. The PSD of the mixtures can be shown on the figure 5. The results shown in figure 6 indicate that this approach gives satisfactory results for the different frequencies plus harmonics. Moreover, we can point out that no frequency channel is permuted. We can also notice that the output SNRs for this example is equal to
in the sense that the estimated sources are the least subjected to distortion among the set of all the valid separators. We show that the implementation of such approach improves appreciably the quality of the sources separation. The simulation results presented in section V have shown that this approach is proving to be efficient in terms of stability and SNR’s performances. Future work will be oriented towards a real system signal processing with many sources.

VI. CONCLUSIONS

A new convolutive BSS algorithm based on the minimization of a mutual information criterion penalized by an additional Minimal Distortion Principle term is presented. This algorithm is implemented by a global gradient method (without any constraint on its displacement) which ensures an efficient optimisation.

The introduction of the Minimal Distortion Principle allows the mitigation of the filtering indeterminacy inherent to the BSS problem, the optimal separator is only “optimal”

REFERENCES