



On convolutive Blind Source Separation in a noisy context and a total variation regularization

T.Z. Z Boulmezaoud, M. El Rhabi, H. Fenniri, E. Moreau

► To cite this version:

T.Z. Z Boulmezaoud, M. El Rhabi, H. Fenniri, E. Moreau. On convolutive Blind Source Separation in a noisy context and a total variation regularization. 2010 IEEE 11th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2010), Jun 2010, Marrakech, Morocco. <10.1109/SPAWC.2010.5671015>. <hal-01812533>

HAL Id: hal-01812533

<https://enpc.hal.science/hal-01812533v1>

Submitted on 11 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

ON CONVOLUTIVE BLIND SOURCE SEPARATION IN A NOISY CONTEXT AND TOTAL VARIATION REGULARIZATION

T.Z. Boulmezaoud^(a), M. El Rhabi^(b), H. Fenniri^(c) and E. Moreau^(d)

(a) LMV, Université de Versailles Saint-Quentin-en-Yvelines,
boulmezaoud@math.uvsq.fr

(b) Realeyes3D, SA,
melrhabi@realeyes3d.com

(c) CReSTIC, Université de Reims Champagne-Ardenne,
hassan.fenniri@univ-reims.fr

(d) ISITV, Université du Sud-Toulon-Var,
moreau@univ-tln.fr

ABSTRACT

We propose a new strategy for improving classical Blind Source Separation (BSS) methods. This strategy consists in denoising both the observed signal and the estimated source signal, and is based on the minimization of regularized criterion which takes into account the Total Variation of the signal. We prove by the way that the method leads to a projection problem which is solved by means of projected gradient algorithm. The effectiveness and the robustness of the proposed separating process are shown on numerical examples.

1. INTRODUCTION

Blind Source Separation (BSS) is one of the most attractive research topics nowadays in the field of signal processing and its applications, as e.g in wireless communication. The goal of BSS is to recover independent sources given only sensor observations that are unknown linear mixtures of the unobserved independent source signals. The principle of BSS is to transform a multivariate random signal into an ideal signal which has mutual independent components in the statistical sense (see [1]), this transformation is achieved up to a permutation (and/or) a filtering operator.

The case of noisy observations has not received much attention in the BSS literature. Actually, most of existing methods are based on noise-free models and could be highly sensitive to observation noises. Including noise in the model changes seriously the structure of the estimation problem and makes it more difficult to tackle. In this paper, we consider the convolutive BSS separation in a noisy context. Here, the observations vector $\mathbf{x}^0 \in \mathbb{R}^N$, N the number of observations, is

obtained from the sources $\mathbf{s} \in \mathbb{R}^N$ through the mixing system up to an (unknown) additive noise $\mathbf{n} \in \mathbb{R}^N$:

$$\begin{aligned} \mathbf{x}^0(t) &= A \star \mathbf{s}(t) + \mathbf{n}(t) \\ &= \mathbf{x}(t) + \mathbf{n}(t), \end{aligned} \quad (1)$$

where \star denotes the convolutive product, $\mathbf{x} = A \star \mathbf{s}(t)$ is the noise-free mixed vector, A is the mixing operator and $\mathbf{s}(t)$ is the independent component source vector. The BSS consists in finding a demixing system defined by :

$$\mathbf{s}^0(t) = B \star \mathbf{x}^0(t), \quad (2)$$

where the vector $\mathbf{s}^0 \in \mathbb{R}^N$ is the output signal vector (estimated source vector) and B the separating operator. Thus, the estimated source signal obtained by a direct BSS can be written into the form

$$\begin{aligned} \mathbf{s}^0(t) &= B \star \mathbf{x}^0(t) \\ &= B \star (A \star \mathbf{s}(t)) + B \star \mathbf{n}(t) \\ &= \hat{\mathbf{s}}(t) + \mathbf{n}_1(t), \end{aligned} \quad (3)$$

where $\hat{\mathbf{s}}(t) = B \star (A \star \mathbf{s}(t))$ and $\mathbf{n}_1 = B \star \mathbf{n}$. That is, the "noisy" estimated source \mathbf{s}^0 is the sum of $\hat{\mathbf{s}}$, the "ideal" estimated source, and the noise \mathbf{n}_1 . Ideally, we would like to retrieve $\hat{\mathbf{s}}$ by denoising \mathbf{s}^0 , from given observations.

Several authors treat the BSS in the noisy case. In [2], they proposed a two-step approach by combining the fractional lower order statistic for the mixing estimation and minimum entropy criterion for noise-free source components estimation, the performance of their method depends on a non linear function properly chosen with relation to the source distribution and the characteristic of the noise. In [3], they proposed a whitening procedure to reduce the noise effect. Both methods are proposed for the linear instantaneous case. In [4], in the linear convolutive case, the authors proposed an algorithm

(a) 45, Avenue des Etats-Unis 78035, Versailles, Cedex. France,

(b) 217 Bureaux de la Colline F-92213 - Saint-Cloud Cedex France

(c) Moulin de la Housse, B.P. 1039, 51687, Reims cedex 2

(d) LSEET UMR 6017, La Valette du Var, France

robust against noise. They assume that noise could be decomposed in coherent and incoherent contributions. In their method to increase robustness of their BSS algorithms against uncorrelated noise, bias removal techniques must be considered.

In general, the BSS framework for convolutive mixture was presented for the noiseless case. The robustness of these methods against the noise is often checked afterward. Our objective here is to propose a new strategy which takes into account the presence of the noise in order to remove it. Our method is mainly based on a double action; (a) a denoising of the observed signal \mathbf{x} before demixing, and a simultaneous BSS-denoising procedure the aim of which is to get a noiseless estimation of source. Both of the actions lead to a regularized optimization problem. Notice that Total variations methods are also used in image processing [5, 6, 7] with some relative success.

The paper is organized as follows. Section 2 motivates our choice and define the denoising and the regularized separating criterion. Section 3 proposes a discretization of the problem and a pre-processing process, shows how to estimate the noise-free observation and presents the regularized separating algorithm. Finally, a discrete form of the criterion, then a stochastic form are given with some numerical results illustrating this work in Section 4.

2. THE METHOD

Our aim here is to get a good approximation of $\hat{\mathbf{s}}$ from (3). The method we propose is decomposed into two steps: (a) step 1: denoising the observed signal (b) step 2: a blind source separation combined with a denoising of the estimated source. Let us sketch the main features of each step.

2.1. Denoising the observed signal

Let x_i^0 , $1 \leq i \leq N$, be the noisy observed stochastic signals which write into the form $\mathbf{x}^0 = \mathbf{x} + \mathbf{n}$. We would reconstruct the *ideal* observed signal \mathbf{x} from \mathbf{x}^0 by means of the following variational problem

$$x_i = \arg \min_{w_i \in X} \frac{1}{2} \mathbf{E} (|w_i - x_i^0|^2) + \lambda \mathbf{E} (\theta (|w_i'|)). \quad (4)$$

where $\lambda > 0$ is a penalization parameter, θ is a well chosen function, $w'(t)$ the first derivative of w and X is an appropriate space. The Euler-Lagrange equation corresponding to this optimization problem writes

$$w_i - \lambda \left(\frac{\theta'(|w_i'|)}{|w_i'|} w_i' \right)' = x_i^0. \quad (5)$$

Thus, in practice, the function θ is chosen to encourage smoothing in regions where the variation of the signal are weak, that is $|w'| \approx 0$, and to preserve discontinuities where $|w'|$

is strong. The Total variation case corresponds to the choice $\theta(t) = t$ which will be adopted here. One can make other choices like $\theta(t) = \sqrt{1+t^2}$.

The natural space for treating the continuous variational problem (4) when $\theta(t) = t$ is $BV([0, T])^N$, the space of bounded variation functions (see [8], [9]). The treatment of the continuous problem is mathematical question which is beyond the scope of this paper. Here, we consider only the corresponding discrete problem which will be detailed and solved in section 3.1 hereafter.

2.2. The simultaneous BSS-denoising procedure

The purpose of this second step is to reconstruct an estimated source signal \mathbf{s} from the partially denoised observed signal \mathbf{x} . Following the arguments of the last section, our model consists in minimizing the criterion

$$J(\hat{\mathbf{s}}) = J_{sep}(\hat{\mathbf{s}}(B; \mathbf{x})) + J_{reg}(\hat{\mathbf{s}}(B; \mathbf{x})), \quad (6)$$

with respect to B . Here $\hat{\mathbf{s}}(B; \mathbf{x}) = B \star \mathbf{x}$, J_{sep} is the separating criterion and J_{reg} is a regularization term of the form

$$J_{reg}(\hat{\mathbf{s}}) = \sum_{i=1}^N \left(\frac{\gamma}{2} \mathbf{E} (|\hat{s}_i - s_i^0|^2) + \mu \mathbf{E} (\theta (|\hat{s}_i'|)) \right) \quad (7)$$

where \mathbf{E} is the mathematical expectation, θ is defined as in section 2.1, \hat{s}_i and s_i^0 are respectively the i^{th} component of $\hat{\mathbf{s}}$, \mathbf{s}^0 for $i \in \{1, \dots, N\}$. \hat{s}' is the first derivative of $\hat{\mathbf{s}}$. Notice that the first term controls the variation of v while the second is the fidelity term. The reals $\mu > 0$ and $\gamma > 0$ are small regularization parameters.

Here, because we deal 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 system. That is why additional constraints must be stated to ensure the mutual independence of the output signal components $y_i(n), i \in \{1, \dots, N\}$. To make it easier to understand, let us consider now a bidimensional random vector $\mathbf{y} = (y_1(n), y_2(n))^T$. 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 . As in [10, 11], we define the separating criterion J_{sep} by:

$$J_{sep}(\hat{\mathbf{s}}) = \sum_{\mathbf{q}} I(\hat{\mathbf{s}}^{\mathbf{q}}), \quad (8)$$

where $\mathbf{q} = (q_1 = 0, q_2, \dots, q_N)$ is an integer vector and $\hat{\mathbf{s}}^{\mathbf{q}}(n) = (\hat{s}_1(n - q_1), \dots, \hat{s}_N(n - q_N))$.

I is the mutual information which could be written as follows:

$$I(\hat{\mathbf{s}}) = \int_{\mathbb{R}^N} \mathbf{p}_{\hat{\mathbf{s}}}(t) \ln \left(\frac{\mathbf{p}_{\hat{\mathbf{s}}}(t)}{\prod_{i=1}^N p_{\hat{s}_i}(t)} \right) dt, \quad (9)$$

where $p_{\hat{s}}$ define the joint probability density function (joint pdf) and $p_{\hat{s}_i}$, $i \in \{1, \dots, N\}$, the marginal probability density function of the i^{th} component of \hat{s} (marginal pdf). Actually, the separation is obtained when the components of \hat{s}^q become independent.

3. DISCRETIZATION

In the sequel, continuous signals are sampled at a period T_e ; to each continuous signal u , we associate a vector $(u^0, \dots, u^m) \in X = \mathbb{R}^{m+1}$ defined by $u^k = u(kT_e)$, $0 \leq k \leq m$. This vector is still denoted by u . The space $X = \mathbb{R}^{m+1}$ is equipped with the euclidian inner product $\langle u, v \rangle = \sum_{k=0}^m u^k v^k$, for all $x = (u^0, \dots, u^m) \in X$ and $\hat{s}^k = (v^0, \dots, v^m) \in X$. The first derivative of $u \in X$, written u' , belongs to X and is defined by

$$(u')^k = \frac{u^{k+1} - u^k}{T_e} \text{ if } k < m, (u')^m = 0. \quad (10)$$

We define also the backward derivative of $u \in X$, written u^* , by :

$$(u^*)^0 = \frac{u_i^k}{T_e}, (u^*)^m = -\frac{u_i^{k-1}}{T_e}$$

and

$$(u^*)^k = \frac{u^k - u^{k-1}}{T_e} \text{ if } 0 < k < m.$$

The backward derivatives $(\cdot)^*$ is the adjoint operator of $(\cdot)'$. That is, for all $u \in X$, we have

$$\forall v \in X, \langle u^*, v \rangle = -\langle u, v' \rangle.$$

All these definitions are extended in a natural way to elements of X^N .

3.1. The denoising of the discrete observed signal

In this section, we show how to estimate in practice x from the observation x^0 . Recall that this estimation is obtained by solving the optimization problem (4). We start with the following proposition which gives a simple characterization of the solution

Proposition 1 *The problem (4) has a unique solution given by*

$$x = x^0 - \Pi_{\lambda G} x^0, \quad (11)$$

where $\Pi_{\lambda G}$ is the orthogonal projection operator on the convex set λG with $G = \{v^* | v \in X^N, |v_i^k| \leq 1, \forall (i, k)\}$.

Thus, we are lead to compute the projection operator $\Pi_{\lambda G}$ on the convex λG . In other words, we solve the following problem:

$$x = \arg \min_{v \in D^N} \|\lambda v^* - x^0\|^2 \quad (12)$$

where D is the convex given by $D = \{v \in X |, \forall k \leq m, |v^k| \leq 1\}$. In order to solve the problem (12), one can use a projected gradient method, that is

Data: x^0 the observation vector
Result: x the denoised observation vector
initialization : Given $\varepsilon > 0$, $x_0 = x^0$ and $\rho > 0$
do
 • update x :
 for $i=1, \dots, N$
 $x_i^{p+1} = \Pi_D (x_i^p - 2\lambda\rho(\lambda(x_i^p)^* - x_i^0)')$
 until $\|x^{p+1} - x^p\| < \varepsilon$
 $x = x^{p+1}$.

where Π_D is the orthogonal projection on D , given by:

$$\forall v \in X, (\Pi_D v)^k = \begin{cases} \frac{v^k}{|v^k|}, & \text{if } |v^k| \geq 1, \quad k = 0, \dots, m. \\ v^k, & \text{else} \end{cases} \quad (13)$$

Afterwards, the filtered observation vector x will be considered as the denoised version of the observation vector x^0 . Now, we investigate the BSS step.

3.2. An algorithm for the simultaneous BSS-denoising algorithm

In this section, we apply the gradient approach to separate convolutive mixtures based on the minimization of the criterion (6). Let us assume that the discrete separating system form of (2) is defined by:

$$\hat{s}(n) = \sum_{k=0}^L B_k x(n-k), \quad (14)$$

where B_k are FIR filters with maximum degree L . To estimate the matrices B_k leading to estimate sources outputs, we calculate the gradients of (6) according to each B_k . So, the derivation of the separating term of (6) 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:

$$\begin{aligned} \varphi_{\hat{s}}(\hat{s}) &= \left(-\frac{\partial p_{\hat{s}}}{\partial \hat{s}_1}, \dots, -\frac{\partial p_{\hat{s}}}{\partial \hat{s}_N} \right), \\ \psi_{\hat{s}}(\hat{s}) &= \left(-\frac{p_{\hat{s}_1}(\hat{s}_1)}{p_{\hat{s}_1}(\hat{s}_1)}, \dots, -\frac{p_{\hat{s}_N}(\hat{s}_N)}{p_{\hat{s}_N}(\hat{s}_N)} \right), \\ \beta_{\hat{s}}(\hat{s}) &= \psi_{\hat{s}}(\hat{s}) - \varphi_{\hat{s}}(\hat{s}). \end{aligned}$$

We state the following proposition

Proposition 2 Let us consider J defined by (6). Then,

$$\frac{\partial J(\hat{\mathbf{s}}(n))}{\partial B_k} = \mathbf{E} \left(\left(\beta_{\hat{\mathbf{s}}^q}^{-q}(n) - \mu \left(\frac{\theta'(\hat{\mathbf{s}}'(n)) \hat{\mathbf{s}}'(n)^*}{|\hat{\mathbf{s}}'(n)|} \right)^* \right) \mathbf{x}(n-k)^T + \gamma \sum_{l=1}^L B_l (\mathbf{x} - \mathbf{x}^0)(n-l)(\mathbf{x} - \mathbf{x}^0)(n-k)^T \right).$$

Then, we can derive the following algorithm:

Data: \mathbf{x}^0 the observation vector
Result: $\hat{\mathbf{s}}$ the estimated source vector
initialization : Compute $\mathbf{x} = \mathbf{x}^0 - \Pi_D \mathbf{x}^0$. Given $\varepsilon > 0$, B_k^0 , $k \in \{1, \dots, L\}$, $\hat{\mathbf{s}}^0 = \mathbf{x}$ and $\tau > 0$
do
 • update B_k
 for $k=1, \dots, L$,
 $B_k^{p+1} = B_k^p - \tau \frac{\partial J(\hat{\mathbf{s}}^p(n))}{\partial B_k^p}$;
 • update $\hat{\mathbf{s}}$
 $\hat{\mathbf{s}}^{p+1}(n) = \sum_{k=0}^L B_k^{p+1} \mathbf{x}(n-k)$,
until $\|B_k^{p+1} - B_k^p\| < \varepsilon$

4. NUMERICAL RESULTS

In this section, we dealt with two sorts of sample, namely observations obtained by a convolutive mixture of two real, non-Gaussian and independent sources with zero means and BPSK signals which are used, for example, in trellis coding for wireless communications.

The performance comparison is given using the output signal-to-noise ratio (SNR) defined by:

$$SNR = 10 \log_{10} \left(\frac{y_i^2}{r_i^2} \right),$$

$$\text{where } r_i = \left(\sum_{k=1}^L B_k \left(\sum_{l=1}^{n-k} A_l s(n-k-l) \right) \right) \Big|_{s_i=0}.$$

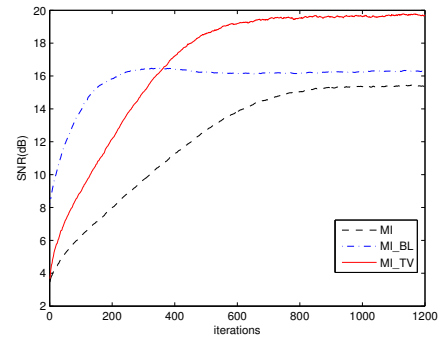
The mixtures were separated using mutual information (MI) method shown in ([10, 11]), MI method where the data are pre-whitening (MI_BL) and our proposed algorithm (MI_TV).

We use the separation criterion (6) in its discrete form, i.e. the finite summation over $q_i \in \{-M, \dots, M\}$ takes the place of the infinite one over $q_i \in \mathbb{Z}$, where $M = 2L$. Since the three criteria are computationally expensive, we implemented here their stochastic version i.e. at each iteration, m was randomly chosen from the set $\{-M, \dots, M\}$. The SFD are estimated using the Pham's method described in [12].

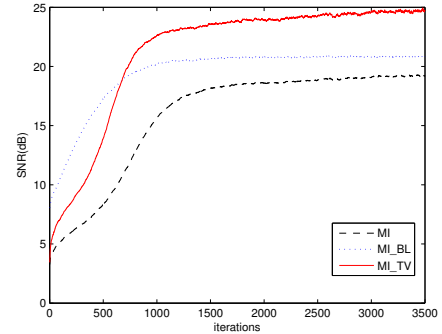
4.1. Example 1

Here, the source signals were linearly mixed through a randomly generated RIF filters of length 6. The number of observations taken equals 1000. The experiment is repeated 100 times with different realizations of the random sources. For each experiment, we add a gaussian noise to the mixing system ($RSB = -30$). Figure 1 shows the averaged SNRs versus iterations for the three algorithms, the adapting step-size is equal to $\tau = 0.2$ for all algorithm and concerning our algorithm parameters: $\mu = 0.001$ and $\gamma = 0.00001$

We notice that the convergence of the algorithms is quite comparable. But in terms of SNRs performances, our algorithm gives better separation.



(a) Results for two random signals



(b) Results for two BPSK signals

Fig. 1. Average output SNRs versus iterations.

4.2. Example 2

Now, the mixing A is a FIR with maximum degree equal to 2, randomly chosen. The number of observations taken equals 2000. The experiment is repeated 25 times with different realizations of the random sources. For each experiment, we

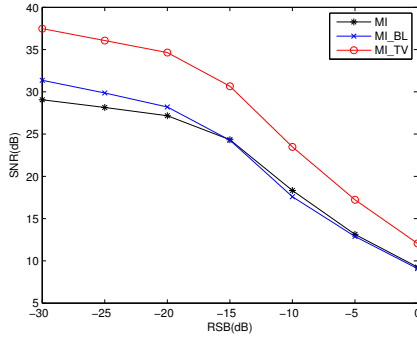
add a gaussian noise at different RSB level, where the RSB is defined such that:

$$RSB = 10 \log_{10} \left(\frac{P_n}{P_s} \right),$$

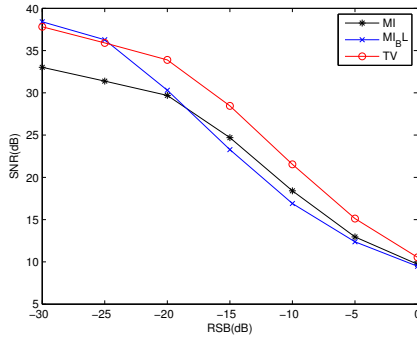
where P_n and P_s are respectively the noise and signal power.

Figure 2 shows the averaged SNRs versus the RSB for the three algorithms, the adapting step-size is equal to $\tau = 0.1$ for all algorithm and concerning our algorithm parameters: $\mu = 0.001$ and $\gamma = 0.00001$.

In this example, we can show that our algorithm outperformed MI and MI.BL algorithms in terms of SNRs.



(a) Results for two random signals



(b) Results for two BPSK signals

Fig. 2. Average output SNRs versus RSB.

5. CONCLUSION

A robust algorithm for convolutive BSS in presence of noise is presented. This algorithm uses variational frameworks to control the noise. It is based on two steps: a pre-processing, reducing the noise on the data and the minimization of a regularized mutual information criterion. This regularization prevents from the error model and allows to denoise the estimate

output. So, the simulation results have shown that this implementation is proving to be very efficient in terms of stability, robustness against noise and SNR's performance.

6. REFERENCES

- [1] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.
- [2] M. Sahmoudi, H. Snoussi, and M. G. Amin, "Robust approach for blind source separation in non-gaussian noise environments," in *Proceedings of ISCCSP, Marrakesh, Morocco*. IEEE/EURASIP, March 2006.
- [3] A. Belouchrani and A. Cichocki, "Robust whitening procedure in blind source separation context," *Electronics Letters*, vol. 36, N24, pp. 2050–2051, Novembre 2000.
- [4] R. Aichner, H. Buchner, and W. Kellermann, "Convolutional blind source separation for noisy mixtures," in *E. Hansler and G. Schmidt (eds.), Topics in Speech and Audio Processing in Adverse Environments*, Springer-Verlag, Berlin/Heidelberg, pp. 469–524, August 2008.
- [5] L.I. Rudin, S. Oshers, and E. Fatemi, "Non linear total variation based noise removal algorithm," in *Annual international conference N11, Los Alamos NM, Etats-Unis N 1-4*, 1992, vol. 60, pp. 259–268.
- [6] A. Chambolle, "An algorithm for total variation minimization and application," *Journal of Mathematical Imaging and vision*, vol. 20 issue 1-2, pp. 89–97, 2004.
- [7] A. Haddad and Y. Meyer, "An improvement of rudin-osher-fatemi model," *Appl. Comput. Harmon. Anal.*, vol. 22(3), pp. 319–334, 2007.
- [8] L.C. Evans and R.F. Gariepy, "Measure theory and fine properties of function," *Studies in Advanced Mathematics*, CRC Press, Boca Raton, FL, 1992.
- [9] A. Cohen, D. Wolfgang, I. Daubechies, and R. DeVore, "Harmonic analysis of the space bv," *Rev. Mat. Iberoamericana*, vol. 19(1), pp. 235–263, 2003.
- [10] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Convolutional mixtures by mutual information minimization," in *Proceedings of IWANN, Granada, Spain*, 2001, pp. 834–842.
- [11] M. El Rhabi, G. Gelle, H. Fenniri, and G. Delaunay, "A penalized mutual information criterion for blind separation of convolutive mixtures," *Signal Processing*, vol. 84, pp. 1979–1984, 2004.
- [12] D.T. Pham, "Fast algorithm for estimating mutual information, entropies and score functions," in *Proceedings of ICA 2003 Conference, Nara, Japan*, 2003.