A MINIMIZATION-PROJECTION (MP) APPROACH FOR BLIND SEPARATING CONVOLUTIVE MIXTURES

Massoud BABAIE-ZADEH *[2,3], *Christian JUTTEN*[1] *and Kambiz NAYEBI*[3]

[1]Laboratoire des images et des signaux (LIS),
Institut National Polytechnique de Grenoble (INPG), Grenoble, France

[2]Multimedia Laboratory, Iran Telecom Research Center (ITRC), Tehran, Iran
[3]Electrical engineering department, Sharif University of Technology, Tehran, Iran
mbzadeh@yahoo.com, Christian.Jutten@inpg.fr, knayebi@sina.sharif.edu

ABSTRACT

In this paper, a new algorithm for blind source separation in convolutive mixtures, based on minimizing the mutual information of the outputs, is proposed. This minimization is done using a recently proposed Minimization-Projection (MP) approach for minimizing mutual information in a parametric model. Since the minimization step of the MP approach is proved to have no local minimum, it is expected that this new algorithm has good convergence behaviours.

1. INTRODUCTION

Blind Source Separation (BSS) or Independent Component Analysis (ICA) is a relatively new subject in signal processing which has been started in the mid 80's (see [1] and the references in it). The problem consists in retrieving unobserved independent signals from mixtures of them, assuming there is neither information about the original source signals, nor about the mixing system (hence the term *Blind*).

Suppose that N observed signals $x_1(n), \dots, x_N(n)$ are given, which are assumed to be mixtures of N independent source signals $s_1(n), \dots, s_N(n)$ (here, the number of sources is assumed to be equal to the number of observations). In the most simple case, the observed signals are assumed to be a linear instantaneous mixture of the sources (Fig. 1), that is, $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{x}(n) \triangleq (x_1(n), \dots, x_N(n))^T$, $\mathbf{s}(n) \triangleq (s_1(n), \dots, s_N(n))^T$, and $\mathbf{A}$ is the mixing matrix (assumed to be invertible). Using the sole information about the signals, *i.e.* their statistical independence, the separating matrix $\mathbf{B}$ is estimated by maximizing the independence of the estimated sources $\mathbf{y} = \mathbf{B}\mathbf{x}$. It is well-known [2] that

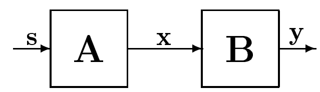

Fig. 1. Mixing and separating systems.

if there is at most one Gaussian source, then the statistical independence of the outputs is sufficient for separating the sources, up to a scale and a permutation indeterminacy.

The instantaneous mixing model may be not accurate enough in many applications. For example, in speech signals, the effects of each speech on two different microphones do not differ only by a scale factor, there is at least some delay. Consequently, the convolutive model arises: the components of the mixing and separating matrices are modeled by linear time invariant (LTI) filters, not scalars. In this model, the mixing system is shown by $\mathbf{x}(n) = [\mathbf{A}(z)]\,\mathbf{s}(n)$ and the separating system by $\mathbf{y}(n) = [\mathbf{B}(z)]\,\mathbf{x}(n)$.

For convolutive mixtures too, it has been shown [3] that the output independence is sufficient for signal separation (up to a filtering and a permutation indeterminacy). However, in convolutive mixtures, the independence of two random processes y_1 and y_2, cannot be reduced to the instantaneous independence of $y_1(n)$ and $y_2(n)$, it requires the independence of $y_1(n)$ and $y_2(n-m)$, for all m [4].

Several methods have been proposed for separating convolutive mixtures. Most of them are based on higher (than 2) order statistics: cancellation of cross-spectra [3], cancellation of second order [5] or higher order cross-moments [6], of higher order cross-cumulants [6, 7, 8], of mutual information of outputs [4], or more generally on a contrast function [9].

Recently, a new approach for minimizing the mutual information in a parametric model has been proposed [10], and applied to BSS in instantaneous (linear and non linear) mixtures. This approach is based on a non-parametric gradient of mutual information [11]. In this paper, using mutual information as an independence criterion, we extend the method and design an algorithm for BSS in convolutive mixtures.

The first author has done main parts of this work when he was with LIS, INPG (Grenoble, France).

This work has been partially funded by the European project Blind Source Separation and applications (BLISS, IST 1999-13077), by the French project Statistiques Avancées et Signal (SASI), and by Iran Telecom Research Center (ITRC).

2. PRELIMINARY ISSUES

2.1. Mutual information

For having a criterion for measuring the independence of some random variables, we first recall that random variables $y_1, \ldots, y_N$ are independent, if and only if $p_{\mathbf{y}}(\mathbf{y}) = \prod_{i=1}^{N} p_{y_i}(y_i)$. Hence, the Kullback-Leibler divergence between $p_{\mathbf{y}}(\mathbf{y})$ and $\prod_{i=1}^{N} p_{y_i}(y_i)$, which is called mutual information of y_i's, can be used as an independence measure:

$$I(\mathbf{y}) = \int_{\mathbf{y}} p_{\mathbf{y}}(\mathbf{y}) \ln \frac{p_{\mathbf{y}}(\mathbf{y})}{\prod_{i=1}^{N} p_{y_i}(y_i)} \mathbf{dy} \qquad (1)$$

This function is always non-negative, and vanishes if and only if the y_i's are independent. Consequently, a blind source separation algorithm can be designed based on minimizing the mutual information of the outputs. It is also shown that this approach is asymptotically a Maximum-Likelihood (ML) estimation of the sources [12].

To do this minimization, knowing an expression for the "gradient" of mutual information is helpful.

2.2. "Gradient" of mutual information

Score Function Difference (SFD) of a random vector has been first introduced in [4]:

Definition 1 (SFD) *The score function difference (SFD) of a random vector* $\mathbf{y}$ *is the difference between its marginal score function* $\boldsymbol{\psi}_{\mathbf{y}}(\mathbf{y})$ *(MSF) and joint score function* $\boldsymbol{\varphi}_{\mathbf{y}}(\mathbf{y})$ *(JSF):*

$$\boldsymbol{\beta}_{\mathbf{y}}(\mathbf{y}) = \boldsymbol{\psi}_{\mathbf{y}}(\mathbf{y}) - \boldsymbol{\varphi}_{\mathbf{y}}(\mathbf{y}) \qquad (2)$$

where the marginal score function is defined by

$$\boldsymbol{\psi}_{\mathbf{y}}(\mathbf{y}) = (\psi_1(y_1), \ldots, \psi_N(y_N))^T \qquad (3)$$

with

$$\psi_i(y_i) = -\frac{d}{dy_i} \ln p_{y_i}(y_i) = -\frac{p'_{y_i}(y_i)}{p_{y_i}(y_i)}. \qquad (4)$$

and the joint score function is defined by

$$\boldsymbol{\varphi}_{\mathbf{y}}(\mathbf{y}) = (\varphi_1(\mathbf{y}), \ldots, \varphi_N(\mathbf{y}))^T \qquad (5)$$

with

$$\varphi_i(\mathbf{y}) = -\frac{\partial}{\partial y_i} \ln p_{\mathbf{y}}(\mathbf{y}) = -\frac{\frac{\partial}{\partial y_i} p_{\mathbf{y}}(\mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})} \qquad (6)$$

The variations of mutual information resulted from a small deviation in its argument (the "differential" of mutual information), is given by the following theorem [11]:

Theorem 1 *Let* $\boldsymbol{\Delta}$ *be a 'small' random vector, with the same dimension than the random vector* $\mathbf{y}$*. Then:*

$$I(\mathbf{y} + \boldsymbol{\Delta}) - I(\mathbf{y}) = E\left\{\boldsymbol{\Delta}^T \boldsymbol{\beta}_{\mathbf{y}}(\mathbf{y})\right\} + o(\boldsymbol{\Delta}) \qquad (7)$$

where $o(\boldsymbol{\Delta})$ *denotes higher order terms in* $\boldsymbol{\Delta}$*, and* $\boldsymbol{\beta}_{\mathbf{y}}$ *is the SFD of* $\mathbf{y}$.

From (7), one may call SFD the *stochastic "gradient"* of mutual information.

2.3. Minimization-Projection (MP) approach

For minimizing $I(\mathbf{y})$ is a parametric model $\mathbf{y} = g(\mathbf{x}, \boldsymbol{\theta})$, one can think about the following steepest descent-like algorithm:

$$\mathbf{y} \leftarrow \mathbf{y} - \mu \boldsymbol{\beta}_{\mathbf{y}}(\mathbf{y}) \qquad (8)$$

As shown in [10], this algorithm converges to a vector with independent components *without trapping in any local minimum*. However, after the convergence, the transformation $\mathbf{x} \mapsto \mathbf{y}$ does not necessarily belong to the parametric family $\mathbf{y} = g(\mathbf{x}, \boldsymbol{\theta})$. To overcome this problem, it is proposed in [10] that at each iteration, the above "minimization" step is followed with a "projection" step, that is, replacing the resulted transformation with its projection on the desired family. In other words, there is two steps at each iteration:

- Minimization:
  1. $\mathbf{y} \leftarrow \mathbf{y} - \mu \boldsymbol{\beta}_{\mathbf{y}}(\mathbf{y})$.
- Projection:
  2. $\boldsymbol{\theta}_0 = \text{argmin}_{\boldsymbol{\theta}} E\{\|\mathbf{y} - \mathcal{G}(\mathbf{x}; \boldsymbol{\theta})\|^2\}$.
  3. $\mathbf{y} = \mathcal{G}(\mathbf{x}, \boldsymbol{\theta}_0)$.

2.4. Separation criterion in convolutive mixtures

In convolutive mixtures (of two sources), the instantaneous independence of outputs, that is, the independence of $y_1(n)$ and $y_2(n)$ is not sufficient for source separation [4]. Instead, $y_1(n)$ and $y_2(n-m)$ must be independent for all m. Consequently, in convolutive mixtures, $I(y_1(n), y_2(n))$ cannot be a separation criterion. And as in [13] we use the separation criterion:

$$J = \sum_{m=-M}^{M} I\big(y_1(n), y_2(n-m)\big) \qquad (9)$$

Theoretically, M must be infinity. However, this is not practically possible. Moreover, when using p-order FIR separating filters (which is always possible in separating convolutive mixtures [13]):

$$\mathbf{B}(z) = \mathbf{B}_0 + \mathbf{B}_1 z^{-1} + \cdots + \mathbf{B}_p z^{-p} \qquad (10)$$

then $M = 2p + 1$ is sufficient for separation. In fact, each term of (9) must vanish, which yields to $2M + 1 = 4p + 3$ equations, while (taking into account the scale indeterminacies) there are $4(p + 1) - 2 = 4p + 2$ unknowns in (10).

As in [13], for reducing the complexity of criterion (9), $I\big(y_1(n), y_2(n-m)\big)$ is used as the separation criterion, but at each iteration a different random m is chosen from $\{-M, \ldots, M\}$. Consequently, in average, the same criterion is being minimized, with highly less computations.

3. ESTIMATING EQUATIONS

Here, we use the idea of MP approach (Section 2.3) to design an algorithm for separating convolutive mixtures. The minimization step (8) does not depend on the separating model. The projection step for convolutive mixtures consists in first finding the filter $\mathbf{B}(z)$ which minimizes the error $E\left\{\|\mathbf{y}(n) - [\mathbf{B}(z)]\,\mathbf{x}(n)\|^2\right\}$, and then replacing $\mathbf{y}(n)$ by $[\mathbf{B}(z)]\,\mathbf{x}(n)$. Using the FIR model (10), we are looking for the matrices $\mathbf{B}_0, \mathbf{B}_1, \ldots, \mathbf{B}_p$ which minimize:

$$\mathcal{C} = E\Big\{\|\mathbf{y}(n) - \sum_{k=0}^{p} \mathbf{B}_k \mathbf{x}(n-k)\|^2\Big\} \qquad (11)$$

After doing some calculations, we have:

$$\begin{aligned}\frac{\partial \mathcal{C}}{\partial \mathbf{B}_k} =& 2\sum_{j=0}^{p} \mathbf{B}_j E\left\{\mathbf{x}(n-j)\mathbf{x}^T(n-k)\right\} \\ &- 2E\left\{\mathbf{y}(n)\mathbf{x}^T(n-k)\right\}\end{aligned} \qquad (12)$$

Finally, by letting $\partial\mathcal{C}/\partial\mathbf{B}_k = \mathbf{0}$ for $k = 0, \ldots, p$, and defining:

$$\mathbf{R}_{\mathbf{xx}}(j,k) \triangleq E\left\{\mathbf{x}(n-j)\mathbf{x}^T(n-k)\right\} \qquad (13)$$

$$\mathbf{R}_{\mathbf{yx}}(j,k) \triangleq E\left\{\mathbf{y}(n-j)\mathbf{x}^T(n-k)\right\} \qquad (14)$$

we obtain the following system of linear equations for determining the optimum filter which maps $\mathbf{x}(n)$ to $\mathbf{y}(n)$:

$$\sum_{j=0}^{p} \mathbf{B}_j \mathbf{R}_{\mathbf{xx}}(j,k) = \mathbf{R}_{\mathbf{yx}}(0,k) \quad , \quad k = 1, \ldots, p \qquad (15)$$

For 4×4 matrices, this system contains $4(p+1)$ unknowns and $4(p+1)$ linear equations. The solution of this equation is:

$$\mathbf{B} = \mathbf{R}_{\mathbf{yx}}\mathbf{R}_{\mathbf{xx}}^{-1} \qquad (16)$$

where:

$$\mathbf{B} \triangleq [\mathbf{B}_0 \quad \mathbf{B}_1 \quad \cdots \quad \mathbf{B}_p] \qquad (17)$$

$$\mathbf{R}_{\mathbf{xx}} \triangleq \begin{bmatrix} \mathbf{R}_{\mathbf{xx}}(0,0) & \hat{\mathbf{R}}_{\mathbf{xx}}(0,1) & \cdots & \hat{\mathbf{R}}_{\mathbf{xx}}(0,p) \\ \mathbf{R}_{\mathbf{xx}}(1,0) & \hat{\mathbf{R}}_{\mathbf{xx}}(1,1) & \cdots & \hat{\mathbf{R}}_{\mathbf{xx}}(1,p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{\mathbf{xx}}(p,0) & \hat{\mathbf{R}}_{\mathbf{xx}}(p,1) & \cdots & \hat{\mathbf{R}}_{\mathbf{xx}}(p,p) \end{bmatrix} \qquad (18)$$

$$\mathbf{R}_{\mathbf{yx}} \triangleq \left[\ \mathbf{R}_{\mathbf{yx}}(0,0) \quad \mathbf{R}_{\mathbf{yx}}(0,1) \quad \cdots \quad \mathbf{R}_{\mathbf{yx}}(0,p)\ \right] \qquad (19)$$

Remark. If we assume that the signals are stationary, then $\mathbf{R}_{\mathbf{xx}}$ and $\mathbf{R}_{\mathbf{yx}}$, as defined in equations (13) and (14), will be the auto-correlation and cross-correlation matrices $\mathbf{R}_{\mathbf{xx}}(k-j)$ and $\mathbf{R}_{\mathbf{yx}}(k-j)$, where:

$$\mathbf{R}_{\mathbf{xx}}(k) \triangleq E\left\{\mathbf{x}(n)\mathbf{x}^T(n-k)\right\} \qquad (20)$$

$$\mathbf{R}_{\mathbf{yx}}(k) \triangleq E\left\{\mathbf{y}(n)\mathbf{x}^T(n-k)\right\} \qquad (21)$$

And the equation system (15) will be written as:

$$\sum_{j=0}^{p} \mathbf{B}_j \mathbf{R}_{\mathbf{xx}}(k-j) = \mathbf{R}_{\mathbf{yx}}(k) \quad , \quad k = 1, \ldots, p \qquad (22)$$

which is very similar to Yule-Walker equations in Auto-Regressive (AR) data modeling. However, here the dimension is higher, and each component of the above equation stands for a matrix, not a scalar. This similarity comes from the similarity of (10) to AR data modeling. However, (15), which is the counterpart of the "covariance method" in AR data modelling, is preferred, because no stationarity is assumed in developing it and it better coincides with a finite number of data points.

- Initialization: $\mathbf{y}(n) = \mathbf{x}(n)$.
- Loop:
  1. Choose a random m from the set $\{-M, \ldots, +M\}$.
  2. Estimate $\boldsymbol{\beta}_{\mathbf{y}^{(m)}}$, the SFD of $\big(y_1(n), y_2(n-m)\big)^T$.
  3. Update the outputs by:
     $$\mathbf{y}^{(m)} \leftarrow \mathbf{y}^{(m)} - \mu\boldsymbol{\beta}_{\mathbf{y}^{(m)}}(\mathbf{y}^{(m)})$$
  4. Remove the DC of each output, and normalize its energy.
  5. Compute $\mathbf{B}_k$, $k = 0, \ldots, p$, from (15).
  6. Let $\mathbf{y}(n) = [\mathbf{B}(z)]\,\mathbf{x}(n)$.
- Repeat until convergence.

Fig. 2. The separating algorithm.

4. THE ALGORITHM

Taking into account the classical scale indeterminacy of BSS, the final separating algorithm for convolutive mixtures will be obtained as sketched in Fig. 2.

5. EXPERIMENTAL RESULTS

To show the ability of the algorithm in separating convolutive mixtures, we mix two uniformly distributed random sources with zeros means and unit variances. The mixing matrix $\mathbf{A}(z)$ is:

$$\begin{bmatrix} 1 + 0.2z^{-1} + 0.1z^{-2} & 0.5 + 0.3z^{-1} + 0.1z^{-2} \\ 0.5 + 0.3z^{-1} + 0.1z^{-2} & 1 + 0.2z^{-1} + 0.1z^{-2} \end{bmatrix}$$

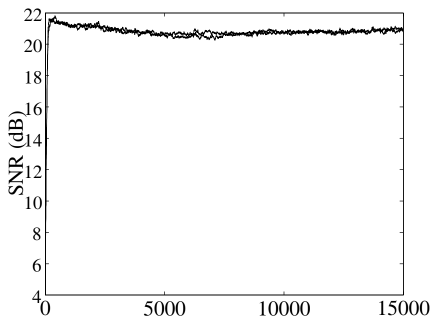

Fig. 3. The output SNR's versus iterations.

The parameters of the separation algorithm are: second order filters ($p = 2$), $M = 2p + 1 = 5$, adaptation rate $\mu = 0.1$, and a 500 sample data block. For estimating SFD, an approach proposed by Pham [14] has been used.

The separation quality is evaluated by output Signal to Noise Ratio (SNR), defined by (assuming there is no permutation):

$$\text{SNR}_i(\text{in dB}) = 10 \log_{10} \frac{E\left\{y_i^2\right\}}{E\left\{y_i^2|_{s_i=0}\right\}} \tag{23}$$

where $y_i|_{s_i=0}$ stands for what is at the i-th output, where the i-th input is zero. The averaged SNR's (taken over 100 runs of the algorithm), is shown at Fig. 3. This figure points out the ability of the MP approach in separating convolutive mixtures.

6. CONCLUSION

Extending a recently proposed [10] general approach for mutual information minimization in a parametric model, called Minimization-Projection (MP) approach, a new algorithm for separating convolutive mixtures has been proposed in this paper.

The advantage of the MP approach, is that its minimization step (Equation (8)) converges without trapping in any local minimum [10]. Consequently, it can be conjectured that the algorithms based on MP approach have better convergence behaviour than the traditional approach of applying a steepest descent algorithm on each parameter of the separating model [4].

The main drawback of the method, based on SFD, is the necessity of estimating a multivariate score function (or density), which becomes tricky and requires large samples, when the dimension (*i.e.* number of sources) grows. Practically, these approaches are limited to 3 or 4 sources.

7. REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiely & Sons, 2001.

[2] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[3] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation," *IEEE Trans. Signal Processing*, pp. 2158–2168, August 1994.

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

[5] U.A. Lindgren and H. Broman, "Source separation using a criterion based on second-order statistics," *IEEE Trans. Signal Processing*, pp. 1837–1850, July 1998.

[6] H.L. Nguyen Thi and C. Jutten, "Blind sources separation for convolutive mixtures," *Signal Processing*, vol. 45, pp. 209–229, 1995.

[7] N. Charkani and Y. Deville, "Self-adaptive separation of convolutively mixed signals with a recursive structure. part 1: Stability analysis and optimization of asymptotic behaviour," *Signal Processing*, vol. 73, pp. 225–254, 1999.

[8] N. Charkani and Y. Deville, "Self-adaptive separation of convolutively mixed signals with a recursive structure. part 2: Theoretical extension and application to synthetic and real signals," *Signal Processing*, vol. 75, pp. 117–140, 1999.

[9] Ph. Loubaton C. Simon and C. Jutten, "Separation of a class of convolutive mixtures : a contrast function approach," *Signal Processing*, vol. 81, no. 4, pp. 883–888, April 2001.

[10] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Minimization-projection (MP) approach for blind source separation in different mixing models," in *ICA2003*, April 2003, pp. 1083–1088.

[11] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Differential of mutual information function," *IEEE Signal Processing Letters*, January 2004, to be appered.

[12] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of IEEE*, vol. 9, pp. 2009–2025, 1998.

[13] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Blind separating Convolutive Post-Nonlinear mixtures," in *ICA2001*, San Diego, December 2001, pp. 138–143.

[14] D. T. Pham, "Fast algorithm for estimating mutual information, entropies and score functions," in *ICA2003*, Nara, Japan, April 2003, pp. 17–22.