BLIND SEPARATING CONVOLUTIVE POST NON-LINEAR MIXTURES

Massoud Babaie-Zadeh[1,2], *Christian Jutten*[1] *and Kambiz Nayebi*[2]

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

[2]Sharif University of Technology, Tehran, Iran

mbzadeh@yahoo.com, Christian.Jutten@inpg.fr, knayebi@sina.sharif.ac.ir

ABSTRACT

This paper addresses blind source separation in convolutive post nonlinear (CPNL) mixtures. In these mixtures, the sources are mixed convolutively, and then measured by nonlinear (*e.g.* saturated) sensors. The algorithm is based on minimizing the mutual information by using multivariate score functions.

1. INTRODUCTION

Blind Source Separation (BSS) is a basic problem in signal processing, which has been considered intensively in the last fifteen years. For linear instantaneous mixtures, the observations are $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{s}$ is the vector of sources, assumed to be statistically independent, $\mathbf{x}$ is the observation vector, and $\mathbf{A}$ is the (constant) mixing matrix. For separating the sources, one tries to estimate a separating system, $\mathbf{B}$, such that the estimated sources are $\mathbf{y} = \mathbf{B}\mathbf{x}$. For linear mixtures, it can be shown that the independence of the components of $\mathbf{y}$, is a necessary and sufficient condition for achieving the separation (up to a scale and a permutation indeterminacy) [1].

Source separation in convolutive mixtures has been addressed by a few authors [2, 3, 4, 5, 6, 7, 8]. In that case, mixing and separating matrices can be modeled by linear time invariant (LTI) filters, *i.e.* the mixing system writes:

$$\mathbf{x}(n) = [\mathbf{A}(z)]\,\mathbf{s}(n) \tag{1}$$

and the separating system:

$$\mathbf{y}(n) = [\mathbf{B}(z)]\,\mathbf{x}(n). \tag{2}$$

For these mixtures too, it has been shown that the output independence is a necessary and sufficient condition for signal separation (up to a filtering and a permutation indeterminacy) [2]. However, it must be noted that for convolutive mixtures, the output independence cannot be deducted from the independence of $y_1(n)$ and $y_2(n)$ for all n, but requires the independence of $y_1(n)$ and $y_2(n-m)$, for all n and all m. Only a few researchers [9, 10, 11, 12, 13, 14, 15, 16] addressed source separation in nonlinear mixtures, whose observations are $\mathbf{x} = f(\mathbf{s})$. The problem consists in restoring the sources by estimating a nonlinear separating system $\mathbf{y} = g(\mathbf{x})$. However, generally, it can be deduced [15] from the Darmois's theorem [17] that the nonlinear mixtures are not separable, *i.e.* the output independence is not sufficient for achieving the separation. Taleb and Jutten [15] have studied a special and realistic case of nonlinear mixtures, called post nonlinear (PNL) mixtures which are separable. As shown in Figure 1, this two-stage system consists of a linear mixing matrix, followed by componentwise nonlinear distortions (due to the sensors). In this paper, we consider the generalization of the PNL model to the case the first stage is a linear convolutive mixtures. We call these mixtures convolutive post nonlinear (CPNL) mixtures. This paper organized as follows. Section 2 contains some preliminary issues about CPNL mixtures, and definitions and properties of multivariate score functions. The estimating equations are developed in Section 3. The separating algorithm and experimental results are presented in Sections 4 and 5, respectively.

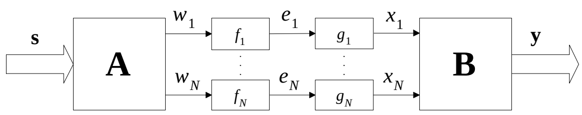

Fig. 1. PNL mixtures: mixing and separating systems

This work has been partly funded by the European project Blind Source Separation and applications (BLISS, IST 1999-14190) and by the French project Statistiques Avancées et Signal (SASI)

2. PRELIMINARY ISSUES

2.1. CPNL mixtures and their separability

In CPNL mixtures, the mixing-separating system is then similar to Fig. 1, but $\mathbf{A}$ and $\mathbf{B}$ are now filter matrices. The separating system is composed of componentwise nonlinear blocks, g_i, such that $x_i = g_i(e_i)$ and of a linear separating filter, $\mathbf{B}(z)$, such that the estimated sources are $\mathbf{y}(n) = [\mathbf{B}(z)]\,\mathbf{x}(n)$. For iid sources and FIR mixing filters, the separability of CPNL mixtures can be directly deduced from the separability of instantaneous PNL mixtures. In fact, denoting $\mathbf{A} = \sum_n \mathbf{A}_n z^{-n}$, and:

$$\mathbf{s} = \left(\dots, \mathbf{s}^T(n-1), \mathbf{s}^T(n), \mathbf{s}^T(n+1), \dots\right)^T \quad (3)$$

$$\mathbf{e} = \left(\dots, \mathbf{e}^T(n-1), \mathbf{e}^T(n), \mathbf{e}^T(n+1), \dots\right)^T \quad (4)$$

we have:

$$\mathbf{e} = \mathbf{f}\left(\bar{\mathbf{A}}\mathbf{s}\right) \quad (5)$$

where $\mathbf{f}$ acts componentwise, and:

$$\bar{\mathbf{A}} = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \mathbf{A}_{n+1} & \mathbf{A}_n & \mathbf{A}_{n-1} & \cdots \\ \cdots & \mathbf{A}_{n+2} & \mathbf{A}_{n+1} & \mathbf{A}_n & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \quad (6)$$

The iid nature of the sources insures the spatial independence of $\mathbf{s}$. Then, the CPNL mixtures can be viewed as a particular PNL mixtures. For FIR mixing matrix $\mathbf{A}(z)$, (5) corresponds to a finite dimension PNL mixture and the separability holds. For more general filter (IIR) matrix, (5) is an infinite dimension PNL mixture, and the separability can be conjectured.

2.2. Mutual information

Random variables $y_1, \dots, y_N$ are independent if and only if:

$$p_{\mathbf{y}}(\mathbf{y}) = \prod_{i=1}^{N} p_{y_i}(y_i). \quad (7)$$

A convenient (scalar) independence measure is the mutual information of the y_i's, denoted by $I(\mathbf{y})$, which is nothing but the Kullback-Leibler divergence between $p_{\mathbf{y}}(\mathbf{y})$ and $\prod_{i=1}^{N} p_{y_i}(y_i)$:

$$\begin{aligned} I(\mathbf{y}) &= D(p_{\mathbf{y}}(\mathbf{y}) \parallel \prod_{i=1}^{N} p_{y_i}(y_i)) \\ &= \int_{\mathbf{y}} p_{\mathbf{y}}(\mathbf{y}) \ln \frac{p_{\mathbf{y}}(\mathbf{y})}{\prod_{i=1}^{N} p_{y_i}(y_i)} \mathbf{dy} \end{aligned} \quad (8)$$

This function is always non negative, and vanishes if and only if the y_i's are independent.

2.3. Independence in convolutive context

As discussed in [8], in convolutive mixtures we have to consider stochastic processes. For sake of simplicity, we restrict the discussion to two random processes. Hence, the independence of $y_1(n)$ and $y_2(n)$ is not sufficient. In fact, $y_1(n)$ and $y_2(n-m)$ have to be independent $\forall n$ and $\forall m$. As a result, $I(y_1(n), y_2(n))$ is not a separation criterion, and we propose [8] the following independence criterion:

$$J = \sum_m I(y_1(n), y_2(n-m)) \quad (9)$$

In theory, the value of m must vary from $-\infty$ to ∞, however in practice we restrict ourselves to the set $\{-M, \dots, M\}$, where M depends on the filter lengths. For reducing the computational cost, we consider a stochastic criterion, derived from J by randomly choosing one term $I(y_1(n), y_2(n-m))$ at each iteration (see Section 4). However, since the nonlinear part does not introduce time delay, one can use $I(y_1(n), y_2(n))$ as a criterion for estimating the nonlinear functions g_i. For sake of simplicity, throughout this paper, we note $\mathbf{y}^{(m)}(n) = (y_1(n), y_2(n-m))^T$. In convolutive mixtures, it must be noted that source separation only provides a filtered version of the sources. However, it is possible to obtain the contribution of each source on each sensor by adding a post processing section [7]. Moreover, due to the filtering indeterminacy, if the mixing filter matrix is a rational filter:

$$\mathbf{A}(z) = \begin{bmatrix} \frac{N_{11}(z)}{D_{11}(z)} & \frac{N_{12}(z)}{D_{12}(z)} \\ \frac{N_{21}(z)}{D_{21}(z)} & \frac{N_{22}(z)}{D_{22}(z)} \end{bmatrix} \quad (10)$$

the separating system can be constrained to be a FIR filter:

$$\mathbf{B}(z) = \begin{bmatrix} \frac{N_{22}(z)}{D_{22}(z)} & -\frac{N_{12}(z)}{D_{12}(z)} \\ -\frac{N_{21}(z)}{D_{21}(z)} & \frac{N_{11}(z)}{D_{11}(z)} \end{bmatrix} \prod_{i,j} D_{ij}(z) \quad (11)$$

2.4. Score Functions

In memoryless mixtures, Mutual Information (MI), $I(\mathbf{y}) = \sum_i H(y_i) - H(\mathbf{y})$, can be written $I(\mathbf{y}) = \sum_i H(y_i) - H(\mathbf{y}) - \ln\det \mid \mathbf{B} \mid - \prod_i g_i'(x_i)$. Minimizing this last expression of $I(\mathbf{y})$ with respect to the separating system leads to an equation in which all the statistical knowledge about the signal is contained in (scalar) score functions. However, in convolutive mixtures, such a derivation is not possible due to $\mathbf{B}(z)$ and it is more convenient to derive directly $I(\mathbf{y}) = \sum_i H(y_i) - H(\mathbf{y})$. The derivation then leads to multivariate score functions [8]. In this section, we recall these definitions.
First, the score function of a scalar random variable is defined as follows.

Definition 1 (Score Function) *The score function of a scalar random variable y is the opposite of the log derivative of its density, i.e.:*

$$\psi_y(y) = -\frac{d}{dy}\ln p_y(y) = -\frac{p_y'(y)}{p_y(y)} \quad (12)$$

where $p_y(y)$ denotes the probability density function (PDF) of y.

Let $\mathbf{y} = (y_1, \ldots, y_N)^T$ be a N-dimensional random vector, we can define two different score functions. Let $p_{\mathbf{y}}(\mathbf{y})$ and $p_{y_i}(y_i)$ denote the joint and marginal PDFs, respectively.

Definition 2 (MSF) *The marginal score function (MSF) of $\mathbf{y}$ is the vector whose component i is the score function of the i-th density:*

$$\boldsymbol{\psi}_{\mathbf{y}}(\mathbf{x}) = (\psi_1(y_1), \ldots, \psi_N(y_N))^T \tag{13}$$

where

$$\psi_i(y_i) = -\frac{d}{dy_i} \ln p_{y_i}(y_i) = -\frac{p'_{y_i}(y_i)}{p_{y_i}(y_i)}. \tag{14}$$

Definition 3 (JSF) *The joint score function (JSF) of $\mathbf{y}$ is the gradient of $-\ln p_{\mathbf{y}}(\mathbf{y})$:*

$$\boldsymbol{\varphi}_{\mathbf{y}}(\mathbf{y}) = (\varphi_1(\mathbf{y}), \ldots, \varphi_N(\mathbf{y}))^T \tag{15}$$

where

$$\varphi_i(\mathbf{y}) = -\frac{\partial}{\partial y_i} \ln p_{\mathbf{y}}(\mathbf{y}) = -\frac{\frac{\partial}{\partial y_i} p_{\mathbf{y}}(\mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})} \tag{16}$$

Finally, we introduce the score function difference.

Definition 4 (SFD) *The score function difference (SFD) of $\mathbf{y}$ is the difference between its JSF and MSF:*

$$\boldsymbol{\beta}_{\mathbf{y}}(\mathbf{y}) = \boldsymbol{\psi}_{\mathbf{y}}(\mathbf{y}) - \boldsymbol{\varphi}_{\mathbf{y}}(\mathbf{y}) \tag{17}$$

The following theorem relates the independence of the components of a random vector $\mathbf{y}$ to its SFD [8].

Theorem 1 *The components of the random vector $\mathbf{y}$ are independent, if and only if, its SFD is zero,* i.e.

$$\boldsymbol{\varphi}_{\mathbf{y}}(\mathbf{y}) = \boldsymbol{\psi}_{\mathbf{y}}(\mathbf{y}) \tag{18}$$

2.5. Gradient of mutual information

For designing the separation algorithm, we use the output mutual information, $I(\mathbf{y})$, as the independence criterion. Hence, the parameters of the separating system are computed so that $I(\mathbf{y})$ reaches its minimum (zero), following a gradient descent algorithm. The gradient of $I(\mathbf{y})$, with respect to the parameters of the separating system, is then deduced from the following theorem [18].

Theorem 2 *Let $\boldsymbol{\Delta}$ be a 'small' random vector, with the same dimension than the random vector $\mathbf{x}$. Then:*

$$I(\mathbf{x} + \boldsymbol{\Delta}) - I(\mathbf{x}) = E\left\{\boldsymbol{\Delta}^T \boldsymbol{\beta}_{\mathbf{x}}(\mathbf{x})\right\} + o(\boldsymbol{\Delta}) \tag{19}$$

where $o(\boldsymbol{\Delta})$ denotes higher order terms in $\boldsymbol{\Delta}$.

Note that for any multivariate differentiable function $f(\mathbf{x})$, we have:

$$f(\mathbf{x} + \boldsymbol{\Delta}) - f(\mathbf{x}) = \boldsymbol{\Delta}^T \nabla f(\mathbf{x}) + o(\boldsymbol{\Delta}) \tag{20}$$

A comparison between (19) and (20) shows that SFD is nothing but the *stochastic gradient* of the mutual information.

3. ESTIMATION EQUATIONS

Solving the estimation equations, $E\left\{\boldsymbol{\Delta}^T \boldsymbol{\beta}_{\mathbf{x}}(\mathbf{x})\right\} = 0$, requires the SFD. Because joint as well as marginal densities are unknown, it is necessary to estimate these densities. Then, deriving the gradient of $I(\mathbf{y})$ with respect to the separating system, leads to the practical estimation equations.

3.1. Estimating Score Functions

In this subsection, we suppose that $\mathbf{y}_1, \ldots, \mathbf{y}_T$ denote T observed samples of random vector $\mathbf{y}$.

3.1.1. Estimating JSF

For estimating JSF of a random vector, we used the kernel estimator method [19] where $K(\mathbf{x})$ denote a multivariate Gaussian kernel function, with zero mean and identity covariance matrix. Then, the estimated joint PDF of the random vector $\mathbf{y}$ from the observations $\mathbf{y}_1, \ldots, \mathbf{y}_T$ will be:

$$\hat{p}_{\mathbf{y}}(\mathbf{y}) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{h} K\left(\frac{\mathbf{y} - \mathbf{y}_t}{h}\right) \tag{21}$$

where the "bandwidth" h is the smoothing parameter. Using this estimator, the i-th component of JSF will be estimated by:

$$\hat{\varphi}_i(\mathbf{y}) = \frac{\sum_{t=1}^{T} \frac{\partial K}{\partial y_i}\left(\frac{\mathbf{y} - \mathbf{y}_t}{h}\right)}{\sum_{t=1}^{T} K\left(\frac{\mathbf{y} - \mathbf{y}_t}{h}\right)} \tag{22}$$

3.1.2. Estimating MSF and SFD

The components of MSF can be estimated by kernel estimators, too. However, this estimation does not lead to good separation results. In fact, the gradient algorithm stops if the SFD (the difference between MSF and JSF) is equal to zero. However, if we independently estimate MSF and JSF, the estimation errors are independent, too, and SDF is not equal to zero when outputs becomes statistically independent. Practically, for avoiding this problem, we propose to estimate MSF from the JSF. Theoretically, this idea is based on the following result [20].

Theorem 3 *Let* $\mathbf{y}$ *and* $p_{\mathbf{y}}(\mathbf{y})$ *be a random vector and its differentiable density, respectively. Then:*

$$\psi_i(u) = E\{\varphi_i(\mathbf{y}) \mid y_i = u\} \tag{23}$$

where ψ_i *and* φ_i *are the* i*-th component of MSF and JSF of* $\mathbf{y}$*, respectively.*

Since φ_i is a function of y_i, this theorem claims that the SFD is the opposite of the variation of φ_i around its mean (ψ_i). When this variation is zero, y_i is then independent of the other components. For estimating the expected value in (23) we use the spline smoothing technique. Hence, the estimated SFD is obtained according to the following steps:

1. Estimate JSF by using kernel estimators,
2. Calculate the smoothed version of JSF by using spline smoothing method,
3. Finally, SFD is the difference between smoothed JSF and JSF.

3.2. Computing the gradients

Let the separating filter $\mathbf{B}(z)$ be:

$$\mathbf{B}(z) = \mathbf{B}_0 + \mathbf{B}_1 z^{-1} + \cdots + \mathbf{B}_p z^{-p} \tag{24}$$

We must compute the derivative of $I\left(\mathbf{y}^{(m)}(n)\right)$ with respect to each $\mathbf{B}_k$ and to each function g_i. Consider:

$$\tilde{g}_i = g_i + \epsilon_i \circ g_i \Rightarrow \tilde{x}_i = x_i + \epsilon_i(x_i) \tag{25}$$

and:

$$\tilde{\mathbf{B}}(z) = \mathbf{B}(z) + \mathcal{E}(z) \tag{26}$$

where $\epsilon_i(x_i)$ is a 'small' variation of $g_i(x)$ and $\mathcal{E}(z)$ is a 'small' variation of $\mathbf{B}(z)$, such that:

$$\mathcal{E}(z) = \mathcal{E}_0 + \mathcal{E}_1 z^{-1} + \cdots + \mathcal{E}_p z^{-p}. \tag{27}$$

Now, up to first order terms, we can write:

$$\begin{aligned}\tilde{\mathbf{y}}(n) &= \left[\tilde{\mathbf{B}}(z)\right]\tilde{\mathbf{x}}(n) \\ &= \mathbf{y}(n) + [\mathcal{E}(z)]\,\mathbf{x}(n) + [\mathbf{B}(z)]\,\boldsymbol{\eta}(n)\end{aligned} \tag{28}$$

where the i-th component of $\boldsymbol{\eta}(n)$ is $\eta_i(n) = \epsilon_i(x_i(n))$. Using the following notations:

$$\begin{aligned}I &= I\left(\mathbf{y}^{(m)}(n)\right) \\ \tilde{I} &= I\left(\tilde{\mathbf{y}}^{(m)}(n)\right) \\ \boldsymbol{\delta}(n) &= [\mathcal{E}(z)]\,\mathbf{x}(n) \\ \boldsymbol{\gamma}(n) &= \boldsymbol{\beta}_{\mathbf{y}^{(m)}(n)}(\mathbf{y}^{(m)}(n)) \\ \boldsymbol{\beta}(n) &= \boldsymbol{\gamma}^{(-m)}(n) \\ \boldsymbol{\xi}(n) &= [\mathbf{B}(z)]\,\boldsymbol{\eta}(n)\end{aligned}$$

ans assuming the sources are stationary, we apply Theorem 2 and obtain (up to first order terms):

$$\tilde{I} - I = E\left\{\boldsymbol{\beta}^T(n)\boldsymbol{\delta}(n)\right\} + E\left\{\boldsymbol{\beta}^T(n)\boldsymbol{\xi}(n)\right\} \tag{29}$$

The first term of (29) can be simplified as:

$$E\left\{\boldsymbol{\beta}^T(n)\boldsymbol{\delta}(n)\right\} = \sum_{k=0}^{M} E\left\{\boldsymbol{\beta}^T(n)\mathcal{E}_k\mathbf{x}(n-k)\right\} \tag{30}$$

From this equation, it is obvious that:

$$\frac{\partial}{\partial \mathbf{B}_k} I\left(\mathbf{y}^{(m)}(n)\right) = E\left\{\boldsymbol{\beta}(n)\mathbf{x}^T(n-k)\right\} \tag{31}$$

After a few simple steps, the second term of (29) becomes:

$$\begin{aligned}E\left\{\boldsymbol{\beta}^T(n)\boldsymbol{\xi}(n)\right\} &= \sum_{k=0}^{M} E\left\{\boldsymbol{\beta}^T(n)\mathbf{B}_k\boldsymbol{\eta}(n-k)\right\} \\ &= E\left\{\boldsymbol{\eta}^T(n)\sum_{k=0}^{M}\mathbf{B}_k^T\boldsymbol{\beta}(n+k)\right\} \\ &= E\left\{\boldsymbol{\eta}^T(n)\boldsymbol{\alpha}(n)\right\}\end{aligned} \tag{32}$$

where:

$$\boldsymbol{\alpha}(n) = \sum_{k=0}^{M}\mathbf{B}_k^T\boldsymbol{\beta}(n+k) = \left[\mathbf{B}^T(z)\right]\boldsymbol{\beta}(-n) \tag{33}$$

Finally, developing $\boldsymbol{\eta}^T(n)$ and $\boldsymbol{\alpha}(n)$, we have:

$$\begin{aligned}E\left\{\boldsymbol{\eta}^T(n)\boldsymbol{\alpha}(n)\right\} &= \sum_{i=1}^{N} E\left\{\epsilon_i(x_i(n))\alpha_i(n)\right\} \\ &= \sum_{i=1}^{N} E\left\{\epsilon_i(x_i(n))E\left\{\alpha_i(n) \mid x_i(n)\right\}\right\} \\ &= \sum_{i=1}^{N}\int_{-\infty}^{+\infty}\epsilon_i(x)E\left\{\alpha_i(n) \mid x_i(n) = x\right\}p_{x_i}(x)dx\end{aligned} \tag{34}$$

We then deduce that the *relative gradient* of $I(\mathbf{y})$ with respect to the function g_i is the function

$$(\nabla_{g_i} I)(x) = E\left\{\alpha_i(n) \mid x_i(n) = x\right\} \tag{35}$$

Hence, choosing $\epsilon_i(x) = -(\nabla_{g_i} I)(x)$ leads to decreasing of I. For estimating the expected value (35), we used the smoothing spline method, *i.e.* the function $(\nabla_{g_i} I)(x)$ is the smoothing spline which fits the data $(x_i(n), \alpha_i(n))$.

4. THE ALGORITHM

The matrices $\mathbf{B}_k$ and $x_i = g_i(e_i)$ are estimated according to:

$$\mathbf{B}_k = \mathbf{B}_k - \mu_1\frac{\partial I\left(\mathbf{y}^{(m)}(n)\right)}{\partial \mathbf{B}_k} \tag{36}$$

$$x_i = x_i - \mu_2(\nabla_{g_i} I)(x_i) \tag{37}$$

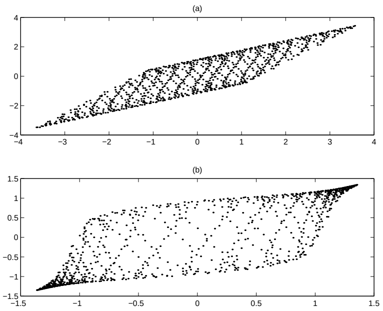

Fig. 2. Joint distribution a) (w_1, w_2) after the convolutive mixtures b) (e_1, e_2), the observations just after the nonlinearities.

where μ_1 and μ_2 are two small positive constants, and the gradients are obtained from (31) and (35). In (36), m is chosen randomly at each iteration in the set $\{-M, \ldots, M\}$. Hence, on the average, the cost function (9) will be minimized. Conversely, in (37), we always used $m = 0$. Practically, after each iteration, one estimates the smoothing function which fits to the data (e_i, x_i). This is for preventing these functions, g_i, to become too fluctuating (and hence noninvertible). Without this step, the estimation of each sample x_i may be very noisy. We used smoothing splines with λ close to 1 (say $\lambda = 0.99999$). Moreover, for overcoming the scale indeterminacies in estimating the g_i's and $\mathbf{B}$, we normalize x_i and y_i at each iteration.

5. EXPERIMENTAL RESULTS

The algorithm has been experimentaly tested for a mixture of a sine and a triangle waves, with the mixing matrix

$$\mathbf{A}(z) = \begin{bmatrix} 1 + 0.2z^{-1} + 0.2z^{-2} & 0.5 + 0.3z^{-1} + 0.1z^{-2} \\ 0.5 + 0.3z^{-1} + 0.1z^{-2} & 1 + 0.2z^{-1} + 0.2z^{-2} \end{bmatrix} \tag{38}$$

and the nonlinearities

$$f_1(x) = f_2(x) = \tanh(x) + 0.1x \tag{39}$$

Figure 2 shows the joint distribution of the convolutively mixed signals (before the nonlinearities) and of the observations (after the nonlinearities). We used sample of size $T = 1000$. We choose second order filters in the separating system. The step sizes are $\mu_1 = \mu_2 = 0.02$. For estimating the JSF, we use Gaussian kernels with a bandwidth $h = 0.3$. The MSF are estimated from JSF with a smoothing spline with $\lambda = 0.8$. For estimating the conditional mean in (35), another smoothing spline with $\lambda = 0.1$ is used. Finally, after each iteration, another smoothing spline with smoothing parameter $\lambda = 0.99999$ is used for smoothing g_i. The

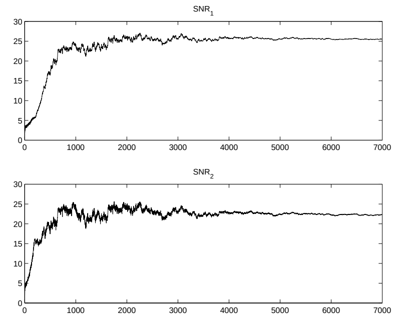

Fig. 3. Output SIR's

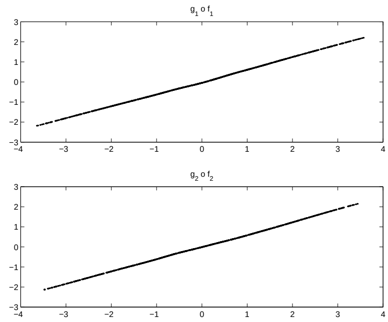

Fig. 4. Composition of the nonlinearities $(g_i \circ f_i)$ (over $i = 1$, below $i = 2$)

separation performance is measured with the Signal to Interference Ratio (SIR), defined as follows (assuming no permutation):

$$\mathrm{SIR}_i = 10 \log_{10} \frac{E\left\{y_i^2(n)\right\}}{E\left\{\left. y_i^2(n)\right|_{s_i(n)=0}\right\}} \tag{40}$$

where $\left. y_i(n)\right|_{s_i(n)=0}$ is the output $y_i(n)$ when the source $s_i(n)$ is zero. Figures 3, 4 and 5 show the separation results. Figure 3 is the output SIR's and points out the ability of the algorithm to separate CPNL sources. Figure 4 shows that the nonlinear distortions of the sensors have been com-

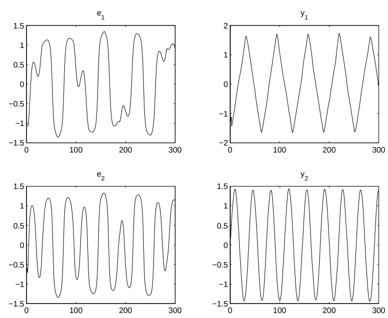

Fig. 5. Observed and output signals

pensated. Finally, Fig. 5 shows the observations (after filtering and nonlinear distortions) and the estimated sources (after separation).

6. CONCLUSION

In this paper, We address the problem of source separation in Convolutive Post NonLinear mixtures (CPNL), which are genralization of instantaneous PNL mixtures. We first show that the separability of CPNL can be deduced from the separability of PNL, provided than the filters $\mathbf{A}(z)$ are FIR filters. Secondly, we propose an algorithm based on the minimization of mutual information, whose efficiency is pointed out by experimental results. The main drawback of the algorithm is its computation cost, especially because it requires the estimation of multivariate densities. Consequently, the algorithm would not be tractable for separating sources in mixtures of a large number of sources (more than 3 or 4), because it would require too large sample. Current work address practical issues for simplifying the algorithm in order to overcome these problems.

7. REFERENCES

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

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

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

[4] A. Gorokov and Ph. Loubaton, "Second-order blind identification of convolutive mixtures with temporally correlated sources: a subspace based approach," in *Proc. European Signal Processing Conf. EUSIPCO 96*, Trieste, Italia, September 1996, pp. 2093–2096.

[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] A. Mansour, C. Jutten, and Ph. Loubaton, "Adaptive subspace algorithm for blind separation of independent sources in convolutive mixture," *IEEE Trans. on Signal Processing*, vol. 48, pp. 583–586, 2000.

[7] C. Simon, *Séparation aveugle des sources en mélange convolutif*, Ph.D. thesis, l'université de Marne la Vallée, Novembre 1999, (In French).

[8] 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.

[9] G. Burel, "Blind separation of sources: a nonlinear neural algorithm," *Neural Networks*, vol. 5, pp. 937–947, 1992.

[10] L. Parra, G. Deco, and S. Miesbach, "Statistical independance and novelty detection with information preserving nonlinear maps," *Neural Computation*, vol. 8, pp. 260–269, 1996.

[11] P. Pajunen, A. Hyvarinen, and J. Karhunen, "Non linear source separation by self-organizing maps," in *ICONIP 96*, Hong-Kong, September 1996.

[12] A. Taleb and C. Jutten, "Non-linear source separation: the post non-linear mixtures," in *ESANN'97*, Bruges, Belgium, April 1997, pp. 279–284.

[13] T.-W. Lee, B.-U.. Koehler, and R. Orglmeister, "Blind source separation in nonlinear mixing models," in *Neural Networks for signal Processing*, 1997.

[14] H.-H. Yang, S.I. Amari, and A. Cichocki, "Information-theoretic approach to blind separation of sources in non-linear mixtures," *Signal Processing*, pp. 291–300, February 1998.

[15] A. Taleb and C. Jutten, "Source separation in post nonlinear mixtures," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2807–2820, 1999.

[16] G. C. Marques and L. B. Almeida, "Separation of nonlinear mixtures using pattern repulsion," in *ICA 99*, Aussois (FRANCE), 1999, pp. 277–282.

[17] G. Darmois, "Analyse générale des liaisons stochastiques," *Rev. Inst. Intern. Stat.*, vol. 21, pp. 2–8, 1953, (in French).

[18] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Differential of mutual information function," *IEEE Signal Processing Letters*, 2001, submitted.

[19] W. Härdle, *Smoothing techniques with implementation in S*, Springer-Verlag, 1991.

[20] M. Babaie-Zadeh, C. Jutten, and K. Nayebi, "Using joint score functions in separating post non-linear mixtures," in *Scientia-Iranica*, 2001, submitted.