SOURCE SEPARATION IN POST NONLINEAR MIXTURES: AN ENTROPY-BASED ALGORITHM

Anisse Taleb, Christian Jutten and Serge Olympieff

INPG-LIS, 46 avenue Félix Viallet
38031 Grenoble Cedex, France
email: taleb or chris@tirf.inpg.fr

ABSTRACT

This paper proposes a new approach for sources separation in special nonlinear mixtures, called post nonlinear mixtures (PNL). We first explain the nice separability properties of these mixtures: solutions have almost the same indeterminacies as in instantaneous linear mixtures. The method proposed in this paper is based on the minimization of the mutual information, which needs the knowledge of source distributions or more exactly of log-derivative of source distributions (the so-called *score functions*). The algorithm consists of three adaptive blocks: one nonlinear block is devoted to adaptive estimation of source score functions, and drives the adaptation of the two other blocks estimating the linear and nonlinear parts of the mixtures. The paper finishes with experimental results which illustrate the efficiency of the algorithm.

1. INTRODUCTION

The problem of source separation has been intensively studied during the last ten years, mainly in the case of linear instantaneous mixtures, and more recently for linear convolutive mixtures. Conversely, source separation in nonlinear mixtures has been very sparsely adressed. This can be easily explained by the following remark. Let u and v be two independent random variables, then $f(u)$ and $g(v)$, where f and g are any nonlinear functions, are also statistically independent. It means that, using the source independence assumption, it will be not possible to estimate the original sources, but only some unknown nonlinear function of the sources. Such distortions on estimated sources are discouraging for addressing this problem.

Anyway, a few authors, despite the difficulty, explored the problem of source separation in nonlinear mixtures. Pajunen *et al.* [7] used Kohonen's maps to interpolate the solutions. Deco [6] studied the very particular (and restricting) case of volume-conserving nonlinear transforms. More recently, Yang *et al.* [12] proposed an algorithm for mixtures with inter-channel nonlinearities.

In this paper, we addressed source separation in particular nonlinear mixtures, the so-called post nonlinear mixtures (PNL). We recall in section 2 the nice separability properties of these mixtures [10]. In section 3, we proposed a new separation algorithm based on a quasi-optimal entropy minimization involving estimation of source score functions. In section 4, experimental results prove the efficacy of the algorithm.

Christian Jutten is professor in Institut des Sciences et Techniques in Université Joseph Fourier of Grenoble

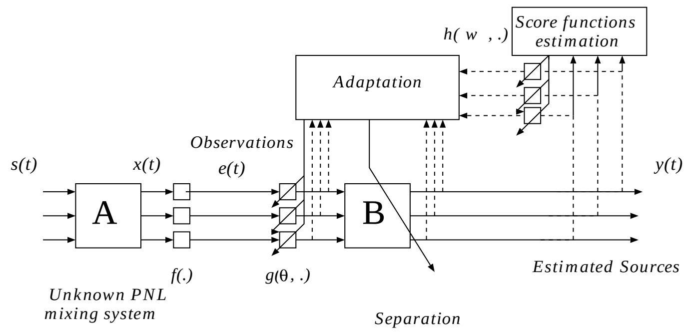

Figure 1: General separation scheme

2. POST NONLINEAR MIXTURES (PNL)

Consider the special nonlinear mixtures of n sources $s_j(t)$ observed by n sensors:

$$\begin{array}{l} x_i(t) = \sum_{i=1}^{n} a_{ij} s_j(t), \\ e_i(t) = f_i(x_i(t)),\ i = 1, \ldots, n, \end{array} \tag{1}$$

where f_i are unknown inversible nonlinear functions, a_{ij} are the unknown real entries of an instantaneous mixing matrix A, and $s_j(t)$ are unknown non Gaussian independent sources. In the following, the mixtures $e_i(t)$ (Fig. 1) will be called post nonlinear mixtures (PNL). Although particular, this model is realistic enough: it corresponds to systems in which the channel transmission is modeled by instantaneous linear mixtures, while sensors with theirs preamplifiers introduce nonlinear mappings.

Assuming PNL mixtures, the separation architecture is a 2-stage structure (Fig. 1): the first stage is a set of n nonlinear blocks suited for inverting the nonlinear mappings f_i , and the second stage is a separating matrix **B** suited for instantaneous linear mixtures. Each nonlinear block provides a parametric estimation $g_i(\theta_i, x)$, where θ_i is a parameter vector. The aim of this function is to cancel the nonlinear distortion f_i, which should be achieved if $g_i(\theta_i, f_i(x)) \propto x$. PNL mixtures have a very interesting property (according to general nonlinear mixtures), summarized in the following lemma [10]:

Lemma 1 *Let be n PNL mixtures of n sources. If the signals $x_i(t)$ are statistically dependent, then the outputs $y_i(t)$ of the separating structure are pairwise independent if and only if $y_i(t) = \alpha_i s_{\sigma(i)}(t) + \beta_i$, where α_i and β_i are real constants and $\sigma(i)$ is a permutation on $\{1, \ldots, n\}$.*

If there is at most one Gaussian source, the independence of x_i's corresponds to a condition on the mixing matrix A. In fact, the source separation without distortion can be achieved if and only if the mixing matrix has at most 2 nonzero entries per row or per column. Then, under this simple condition generally satisfied, the above lemma claims that, in PNL mixtures, source separation is possible with the same (scale and permutation) indeterminacies as in instantaneous linear mixtures, plus a translation, but without nonlinear distortions.

2.1. Source estimation

The separating structure is tuned so that outputs (estimated sources) $y_i(t)$, $i = 1, \ldots, n$, become statistically independent, that is if $\prod_i p_{Y_i}(y_i) = p_Y(y)$. As many researchers have already proposed [5],[11]..., the statistical independence can be measured using the Kullback-Leibler (KL) divergence between the product of marginal densities $\prod_i p_{Y_i}(y_i)$ and the joint density $p_Y(y)$. Independence is achieved if and only if the KL divergence is equal to zero, or which is equivalent, if the mutual information is $I(y)$ minimal :

$$I(y) = \sum_{i=1}^{n} H(y_i) - H(y) \tag{2}$$

The minimization of (2) is difficult because the entropies $H(y_i)$ $(i = 1, \ldots, n)$ explicitely require to know the densities p_{Y_i}. To overcome this problem, it is possible to approximate the densities with Gram-Charlier (GC) expansion. In [10], we explored this idea and approximated the unknown source densities by a 4-th order Gram-Charlier expansion. It leads to a 2-term criterion: the first term is the Comon's contrast function [5] for linear instantaneous mixtures (sum of 4-order squared cumulants), the second term is directly related to the nonlinear part of the mixture. With soft nonlinear mixtures, this method leads to satisfying experimental results. Conversely, with hard nonlinearities, results are disappointing. This is a direct effect of the truncation done in the GC expansion.

In the following, to avoid drawback of GC based algorithm, we developp a new algorithm based on a direct minimization of $I(y)$.

Considering the separating structure: $y = B[g(e)]$, the mutual information (2) is:

$$\begin{aligned} I(y) &= \sum_{i=1}^{n} H(y_i) - H(e) \\ &\quad - \sum_{i=1}^{n} E[\ln \mid g_i'(\theta_i, e_i) \mid] \\ &\quad - \ln \mid \det(B) \mid \end{aligned} \tag{3}$$

2.2. Estimation of the linear part of the separating structure

Deriving the criterion (3) with respect to the separation matrix B directly leads to the algorithm:

$$B(t+1) = B(t) + \mu_t K(y) B^{-T}(t) \tag{4}$$

where μ_t is the stepsize, and $K(y)$ is a square $n \times n$ matrix. This matrix depends on components y_i, $i = 1, \ldots, n$ of output vector y:

$$k_{ij}(y) = \begin{cases} 0 & \text{if } i = j \\ \psi_{Y_i}(y_i) y_i & \text{if } i \neq j \end{cases} \tag{5}$$

where the functions $\psi_{Y_i}(u) = \frac{d}{du} \ln p_{Y_i}(u)$, the so-called *score functions*, are unknown. Introducing the concept of relative gradient [3] or natural gradient [1], a few authors suggested to multiply the gradient by BB^{-T} which leads to the *equivariant*[1] algorithm:

$$B(t+1) = (I + \mu_t K(y)) B(t) \tag{6}$$

Anyway, optimal implementation of (6) also requires to know the score functions $\psi_{Y_i}(y_i)$ or the densities p_{Y_i}.

2.3. Estimation of the nonlinear part of the separating structure

Deriving now the criterion (3) with respect to the parameters θ_i of the nonlinear blocks directly leads to the gradient algorithm:

$$\theta_i(t+1) = \theta_i(t) + \lambda_t \Delta_i(t) \tag{7}$$

where the adaptive increment writes :

$$\Delta_i(t) = \nabla_{\theta_i} \ln \mid g_i'(\theta_i, e_i) \mid + \nabla_{\theta_i} g_i(\theta_i, e_i) \sum_{j=1}^{n} \psi_{Y_j}(y_j) b_{ji} \tag{8}$$

This algorithm can be improved using a second-order optimization method. Such a method consists in replacing the scalar stepsize λ_t by a matrix stepsize G_i, generally equal to the inverse of the Hessian. Then the algorithm (7) becomes :

$$\theta_i(t+1) = \theta_i(t) + G_i \Delta_i(t) \tag{9}$$

To avoid the computation cost of the inverse of the Hessian, usual approximations can be used. It can be proved that the natural gradient corresponds to the matrix :

$$G_i = \lambda_t E \left[\nabla_{\theta_i} \ln \mid g_i'(\theta_i, e_i) \mid \nabla_{\theta_i}^T \ln \mid g_i'(\theta_i, e_i) \mid \right]^{-1} \tag{10}$$

The two algorithms (6) and (9) do not take into account the solution indeterminacies (see Lemma 1 in Section 2). To avoid this problem, a simple and efficient method consists in using penalty terms so that the outputs of nonlinear and linear blocks are zero mean with unit variances. For instance, concerning the linear part, we propose to enforce the diagonal terms $k_{ii}(y)$ of the matrix $K(y)$ (equal to zero in (5)) to :

$$k_{ii}(y) = 1 - y_i^2. \tag{11}$$

[1] it means that the performance does not depend on the mixing matrix A

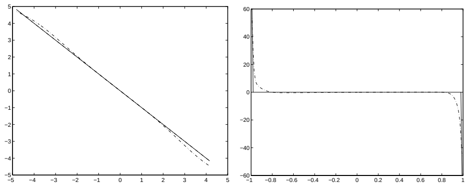

Figure 2: Estimation of score functions : Gaussian (left), Uniform (right), (–) theoric, (- -) estimated

2.4. Estimation of score functions

Considering algorithms (6) and (9) derived from the criterion (3), it appears that the score functions are the quantities of interest. They contain all the necessary information on source distributions for minimizing the criterion but, like the sources densities, they are unfortunately unknown. The interest of score functions has been already pointed out by a few authors: Pham *et al.* [8] proved that maximum likelihood estimation of sources can be performed by zeroing $E[y_i\psi_{Y_j}(y_j)]$ for $i \neq j$; Cardoso and Laheld [3] proved that the equivariant algorithm EASI achieves the best performance when using nonlinear functions proportional to the sources score functions. This result can also be generalized for convolutive mixtures [4]. In [4] and [8], authors perform a linear parametric estimation by projecting the score functions on a subspace spanned by 3 nonlinear functions (which must be carefully chosen).

To avoid the choise of these nonlinear functions, which seems tricky in the case of nonlinear mixtures, we propose to estimate directly the score functions, using a "universal" nonlinear model. We chose to use n multilayer perceptrons (MLP), one for estimating each score function ψ_{Y_i} (Fig. 1). Each network provides an estimation $h_i(w_i, u)$, parameters w_i of which are tuned in order to minimize the mean square error:

$$\mathcal{E}_i(w_i) = E[(h_i(w_i, u) - \psi_{Y_i}(u))^2]. \tag{12}$$

Deriving (12) with respect to the parameters w_i yields:

$$\begin{aligned} \nabla_{w_i}\mathcal{E}_i(w_i) = \quad & 2E[h_i(w_i,u)\nabla_{w_i}h_i(w_i,u) \\ & +\nabla_{w_i}\tfrac{\partial h_i(w_i,u)}{\partial u}]. \end{aligned} \tag{13}$$

Surprinsingly[2], this last equation no longer depends on the target function ψ_{Y_i}. This property, due to properties of ψ_{Y_i}, points out that the mean square error estimation of score functions leads to an unsupervised learning of the MLP, although the criterion (12) is clearly supervised.

With simple networks (MLP with one hidden layer containing 5 to 6 neurons), this method provides very good results even for hard nonlinear score functions, as shown in Fig. 2. Moreover, we would like to emphasize on the generalization of this approach : it can be used in any algorithm driven by entropy minimization (or maximization). In fact, these algorithms require to know the score functions, or more generally the gradient of $\ln p_Y$ in the multivariate case. In [9], this method is applied in linear source separation and obtains performance clearly better (the crosstalk is less than -60dB) than others. The method can also be used for complex data (usual in narrowband telecommunications): in [2], we show that the performance obtained with this algorithm is clearly better than those obtained with the EASI algorithm. Especially, the algorithm EASI requires a condition on the source kurtosis to be stable. This condition is cancelled by using our approach.

[2] In fact, it is not so surprising, because density estimation is achieved by histograms or kernel estimation which are basically unsupervised methods, and score functions are nothing but the log-derivative of the densities.

3. COMPUTER EXPERIMENTS

In this section we illustrate the efficiency of the proposed approach by a computer simulation. The sources are a sinusoid and a uniform white noise (Fig. 3). The mixing matrix (randomly chosen) is :

$$A = \begin{bmatrix} \texttt{-2.29} & \texttt{0.49} \\ \texttt{1.84} & \texttt{0.41} \end{bmatrix}$$

The mixtures $e(t)$ (Fig. 3, bottom) are obtained by applying a different and unknown nonlinear distortion on each channel (linear mixture), in this example the two nonlinear distorsions were :

$$\begin{aligned} f_1(u) &= \tfrac{1}{10}(x + x^3) \\ f_2(u) &= \tfrac{3}{10}x + \tanh 3x \end{aligned} \tag{14}$$

The joint distribution of the mixtures (Fig. 4, left) indicates the existence of a nonlinear dependence (with linear mixtures, joint distribution is contained in a simple parallelogram when the distributions have bounded supports).

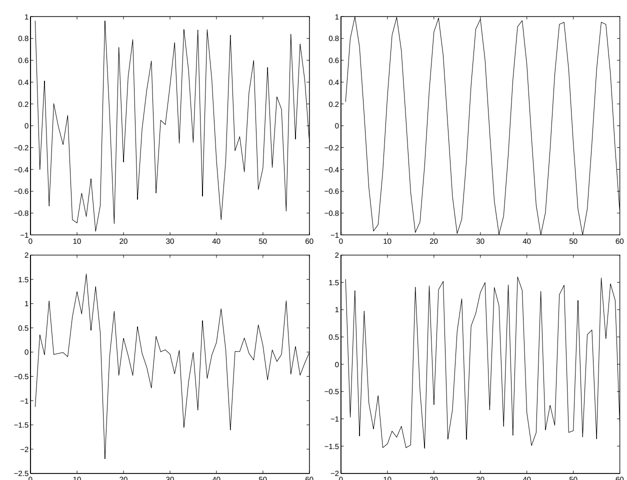

Figure 3: Waveforms : Sources (top), PNL mixtures (bottom)

After convergence of the algorithm, joint distribution of the nonlinear block outputs (after compensation by g_i) shows that the effect of the distortions has been succesfully cancelled (Fig. 4, right). The linear stage performs then a linear source separation on this mixture. The estimated sources are shown in Fig. 6. Their joint distribution (Fig.

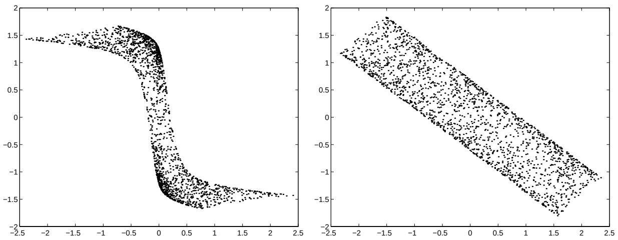

Figure 4: Distributions : PNL mixtures (left), Output of nonlinear functions g_i (right)

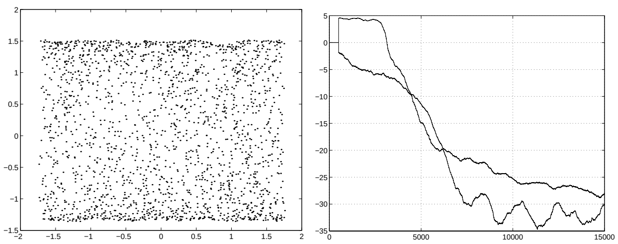

Figure 5: Distribution of estimated sources (left), Residual Crosstalk decrease (right)

5, left) indicates that the independence has been reached. Figure 5 (left), shows the decrease of the residual crosstalk during the run of the algorithm. The residual crosstalk is mesured by :

$$C_i = 10 \log_{10} E[(\hat{y}_i - \hat{s}_i)^2] \tag{15}$$

where $\hat{u} = \frac{u}{\sigma_u}$ is a normalized version of u.

4. CONCLUSION

In this paper, we presented an algorithm for the separation of PNL mixtures. Theses mixtures are realistic and correspond to a lot of real world applications. Moreover, they enjoy nice separability properties without distortions. The method of separation is based on mutual information, exact minimization of which needs the knowledge of the estimated sources score functions. These functions are estimated using an unsupervised algorithm, and allows to estimate with very high accuracy the parameters of the nonlinear separation architecture.

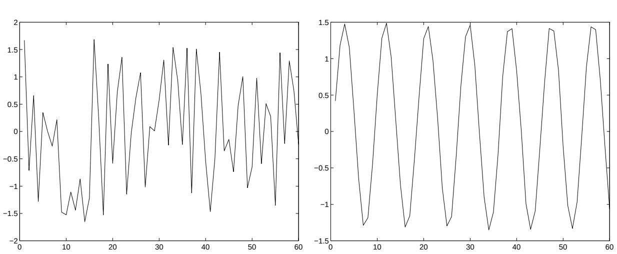

Figure 6: Waveforms of the estimated sources

5. REFERENCES

[1] S-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 1997. To appear.

[2] S. Bozinoski, A. Taleb, J-C. Guizzo, and C. Jutten. Séparation de sources, application à la séparation de signaux et de brouilleurs dans un satellite de télécommunications. In *XVIéme colloque GRETSI*, Grenoble, Septembre 1997.

[3] J-F. Cardoso and B. Laheld. Equivariant adaptive source separation. *IEEE Trans. on S.P.*, 44(12):3017–3030, December 1996.

[4] N. Charkani and Y. Deville. Optimization of the asymptotic performance of time-domain convolutive source separation algorithms. In *ESANN'97*, pages 273–278, Bruges, Belgium, April 1997.

[5] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287 – 314, April 1994.

[6] G. Deco and W. Brauer. Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures. *Neural Networks*, 8(4):525–535, 1995.

[7] P. Pajunen, A. Hyvarinen, and J. Karhunen. Nonlinear blind source separation by self-organizing maps. In *ICONIP 96*, volume 2, pages 1207–1210, Hong-Kong, September 1996.

[8] D. T. Pham, P. Garat, and C. Jutten. Separation of a mixture of independent sources through a maximum likelihood approach. In *Signal Processing VI, Theories and Applications*, pages 771–774, Brussels, Belgium, August 1992. Elsevier.

[9] A. Taleb and C. Jutten. Entropy optimization, application to blind source separation. In *ICANN'97*, Lausanne, October 1997. Submitted.

[10] A. Taleb and C. Jutten. Nonlinear source separation : The post-nonlinear mixtures. In *ESANN'97*, pages 279–284, Bruges, April 1997.

[11] L. Xu, C. C. Cheung, and S. Amari. Nonlinearity ans separation capability: Further justification for the ica algorithm with mixtures of densities. In *ESANN 97*, pages 291–296, Bruges, April 1997.

[12] H. H. Yang, S. Amari, and A. Cichocki. Information back-propagation for blind separation of sources from non-linear mixture. In *ICNN'97*, Houston.