ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 279-284

Nonlinear source separation: The Post-Nonlinear Mixtures.

A. TALEB and C. JUTTEN*
INPG, TIRF, 46, ave Félix Viallet, 38031 Grenoble Cedex, France.
GDR-PRC ISIS
e-mail: `taleb` or `chris@tirf.inpg.fr`
www : `http://www-tirf.inpg.fr/`

Abstract. This paper proposes a first approach for separating independent sources in nonlinear mixtures. A brief study of the indeterminacies related to this problem is given. A special case of nonlinear mappings is studied, these mappings have the interesting property of having the same indeterminacies than the linear case. Algorithm and experimental results are brought at the end of this paper.

1. Introduction

The source separation problem, and more generally, independent component analysis, was mainly addressed in the case of linear mixtures. Up to now, extension to nonlinear mixtures has only been sketched by few authors [2, 7, 8]. In [2], Burel proposed a solution for known nonlinear functions. Recently, P. Pajunen *et al.* [7] adressed this problem using self-organizing maps. The idea, although interesting, has a few drawbacks. First, the network stage must be well suited to the input distribution, which may change according to signal power. Secondly, the discrete nature of the network (quantization) implies a limited, and basically poor accuracy, except if we use a huge number of neurons. Finally, this method implicitly assumes source probability density function (pdf) has bounded support. Another technique, introduced by G. Deco *et al.* [4], uses volume conserving nonlinear mappings to obtain statistically independent signals as outputs This condition is generally not acceptable because it imposes very limitative restrictions on the nonlinear mixing function, as we will see in section 2.

2. Nonlinear mixture model and Indeterminacies

Let be a random vector $(e_1(t), \dots, e_n(t))$ which is an unknown instantaneous nonlinear mixture of n independent unknown sources $(s_1(t), \dots, s_n(t))$. The

*C. Jutten is professor in the Institut des Sciences et Techniques de Grenoble (ISTG) of the Université Joseph Fourier (UJF).

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 279-284

relation between sources and observations is in the general case:

$$e_i(t) = \mathcal{F}_i(s_1(t), \dots, s_n(t)) \quad 1 \le i \le n. \tag{1}$$

We suppose that the $\mathcal{F}_i, i = 1, \dots, n$ are bijective mappings of n variables, and that the $\mathcal{F}_i$'s are continious and differentiable (C^1 class).

Without prior information about the type of nonlinearity or about the sources, the indeterminacies are more serious than in the linear case. For linear mixtures it is well known [6], that sources can be recovered only up to a constant scale factor (diagonal matrix $\mathbf{\Lambda}$) and any permutation (permutation matrix $\mathbf{P}$). These two indeterminacies do not imply signal distortion.

On the contrary, for non linear mixtures, indeterminacies may be very strong. In fact, if s_i and s_j are two independent random variables, $f(s_i)$ and $g(s_j)$, where f et g are any mappings, are also statistically independent random variables. Then, statistical independence only insures that the estimated sources $\hat{s}_i(t)$ are any nonlinear function of an original source $s_j(t)$. The original sources can't be retrieved except if more information is provided about either the nonlinear model or the source pdf.

Suppose now that we have a separation structure $\mathcal{G}_i$ able to estimate n independent signals from the mixtures $e_i(t)$. In the general case, the estimated sources are nonlinear functions $K_i(s_i)$ of the original sources [1]:

$$K_i(s_i) = \mathcal{G}_i(e_1, \dots, e_n) \quad 1 \le i \le n \tag{2}$$

Differentiating this function using (2) gives[2]:

$$J_K = \begin{pmatrix} K_1'(s_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_n'(s_n) \end{pmatrix} = J_{\mathcal{G}} \quad J_{\mathcal{F}}, \tag{3}$$

where $J_{\mathcal{G}}$ and $J_{\mathcal{F}}$ are the Jacobian matrices of the $\mathcal{G}_i$'s and the $\mathcal{F}_i$'s respectively. Equation (3) means that the Jacobian matrix of the separating structure must diagonalize the Jacobian of the mixing nonlinear mapping at each point.

Deco's approach imposes volume-conserving mapping $\mathcal{G}$. Then, the determinant of the Jacobian of the separating structure $J_{\mathcal{G}}$ is equal to 1: $det(J_{\mathcal{G}}) = 1$. We clearly see from (3), that this condition requires, for a succesful separation, that $det(J_{\mathcal{F}})$ is factorizable as a product of nonlinear functions:

$$det(J_K) = K_1'(s_1) \cdots K_n'(s_n) = det(J_{\mathcal{G}}).det(J_{\mathcal{F}}) = det(J_{\mathcal{F}}) \tag{4}$$

This constraint does work only for a very limited class of nonlinear mappings. The above model is too general, and in this paper we propose to study a simplified model of mixture, that we call post-nonlinear (PNL) mixtures.

[1] Assuming, without loss of generality, the permutation matrix is the identity.

[2] In the linear case $J_{\mathcal{F}}$ is the mixing matrix $\mathbf{A}$ and $J_{\mathcal{G}}$ is the separating matrix $\mathbf{B}$, J_K then reduces to a diagonal matrix.

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 279-284

3. The Post-Nonlinearity model

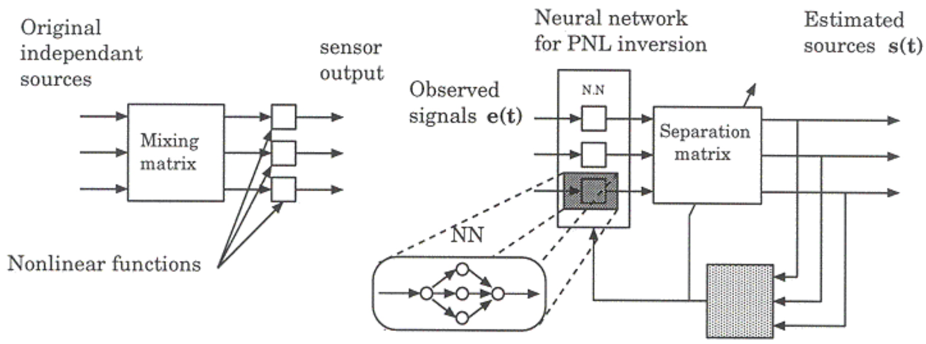

Figure 1: *The post-nonlinearity model and its separating structure.*

We assume that sources are first mixed in a linear memoryless channel, modelled by an invertible mixing matrix $\mathbf{A}$, and then each mixture is distorded by a nonlinear function (see Fig. 1 left). This model is realistic enough in practice, if we consider that the propagation is done via a linear transmission channel, and then the sensors are nonlinear, introducing thus a distortion. Assuming n sources and n sensors, the mixtures in a PNL model are then :

$$e_i(t) = \mathcal{F}_i(\sum_{j=1}^{n} a_{ij} s_j(t)) \quad 1 \leq i \leq n. \tag{5}$$

Using the PNL model assumption, we propose the following separation architecture consisting of two stages (see Fig. 1 right): the first one is a set of n blocks devoted to the inversion of the distortion, the second stage is a simple source separation system (separating matrix $\mathbf{B}$) for instantaneous linear mixtures.

The first result of interest is that the separation of PNL mixtures, under some assumptions about the mixing matrix $\mathbf{A}$, leads to almost the same indeterminacies than the instantaneous linear mixtures: sources are recovered up to an affine mapping *i.e.* $\hat{\mathbf{s}}(t) = \mathbf{\Lambda P s}(t) + \mathbf{v}$, where $\mathbf{v}$ is any vector.

This result holds if the invertible matrix $\mathbf{A}$ is mixing enough. We proved a necessary and sufficient condition is that :

$$\forall a_{ij} \neq 0: \quad \exists k \neq j \ / \ a_{ik} \neq 0, \quad or \quad \exists k \neq i \ / \ a_{kj} \neq 0. \tag{6}$$

For lack of space, we don't give the complete proof, but a simple sketch in the case of 2 mixtures of 2 sources. If $\mathbf{A}$ is diagonal, then outputs of PNL are $\mathcal{F}_1(s_1)$ and $\mathcal{F}_1(s_1)$ wich are already independent ! Inversion of $\mathcal{F}_1$ and $\mathcal{F}_2$ based on output independence is impossible. On the contrary, if the matrix is triangular, the PNL outputs are $\mathcal{F}_1(a_{11}s_1 + a_{12}s2)$ and $\mathcal{F}_2(s_2)$, and it is possible to prove that independence of outputs of the separating structure can only be achieved if $\mathcal{G}_1 = \mathcal{F}_1^-1$ and $\mathcal{G}_2 = \mathcal{F}_2^-1$. The results extend for any regular non diagonal matrix. Generalization to any order is quite evident considering pairwise mixtures, and leads to the above condition (6).

4. A maximum likelihood approach

The likelihood of the observations can be written as a function of the sources, of the mixing matrix and of the nonlinear functions. Adaptation of the linear part of the separating structure can be done by any algorithm for linear instantaneous mixture, theses algorithms are very well known, and we only discuss here the adaptation of nonlinear networks. The maximization of the likelihood of the data will consist in adjusting the estimation of the inverse of the nonlinear functions. The observation log-likelihood is written as:

$$\log p_{E/\mathcal{F}}(\mathbf{e}) = \sum_{i=1}^{n} \log p_{S_i}(s_i) + \log |det(J_{\mathcal{F}^{-1}})|, \tag{7}$$

where $det(J_{\mathcal{F}^{-1}})$ is a function of $\mathbf{s}$. We can expand, using Hermite polynomials, the $\log p_{S_i}(s_i)$ term (Gram-Charlier expansion). This approach has already been used by Gaeta and Lacoume [5], and also by Amari *et al.* [1], in the linear case.

Expanding (7) up to the $4th$ order and cancelling constant terms[3], we get[4]:

$$\log p_{E/\mathcal{F}}(\mathbf{e}) = -\frac{1}{2}\sum_{i=1}^{n} s_i^2 + \frac{1}{6}\sum_{i=1}^{n} \kappa_{3i} H_3(s_i) + \frac{1}{24}\sum_{i=1}^{n} \kappa_{4i} H_4(s_i) + \log |det(J_{\mathcal{F}^{-1}})|, \tag{8}$$

where κ_{3i} and κ_{4i} denote the $3rd$ and $4th$ order normalized autocumulants, and $H_3(u)$ and $H_4(u)$ denotes the $3rd$ and $4th$ univariate Hermite polynomials. By maximizing the mean of this equation, we then have something similar to the contrast function of P. Comon [3], up to a supplementary term $\log |det(J_{\mathcal{F}^{-1}})|$, corresponding to the volume "non-conservation". In the linear case, this term corresponds to the natural logarithm of the determinant of the separating matrix and vanishes when we use a data prewhitening block before the separating block. This equation can be simplified if we know that the sources pdf are symetric (*i.e.* with null skewness, $\kappa_{3i} = 0$).

5. Application to the post-nonlinearity problem

The above expansion (8) is valid in the general case. In this section we apply it in the special case of PNL mixtures. Many algorithms for linear sources separation are available, so the estimation of nonlinear blocks $\mathcal{G}_i, i = 1, \cdots, n$ will only be exposed in this section.

[3] Constant terms are without relevance in the maximization of the likelihood.
[4] In standard mesure.

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 279-284

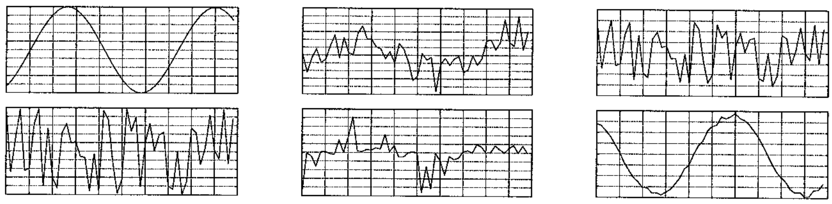

Figure 2: *Original sources, PNL mixtures, Estimated sources.*

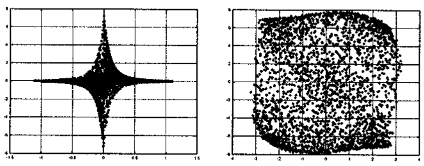

Figure 3: *Joint input and output distribution.*

On each channel, we invert the nonlinear unknown functions $\mathcal{F}_i$, using a simple multilayer perceptron (MLP) with one input $e_i(t)$ and one output $y_i(t)$. Input-output relation is written as:

$$y_i(t) = \sum_{j=1}^{m_i} h_{ij}\sigma(w_{ji}e_i(t) - \theta_{ji}), \; i = 1, \cdots, n. \tag{9}$$

where σ is a sigmoidal function. We use then n simple MLPs, each one is dedicated to the estimation of a nonlinear distortion $\mathcal{G}_i$, $i = 1, \cdots, n$, *i.e.* to the inversion of $\mathcal{F}_i$, $i = 1, \cdots, n$.

The number of neurons in the hidden layer m_i was the same for each MLP. Experimentally in the case of two sources, 5 to 6 neurons are quite sufficient.

The parameters h_{ij},w_{ji} and θ_{ji} are estimated using a gradient ascent of the log-likelihood (note that the MLPs are trained using unsupervised learning algorithm). In the case of two sources and two sensors, we have, $\forall i = 1, 2$:

$$\begin{cases} h_{ij}(t+1) = h_{ij}(t) + \mu_t \frac{\partial}{\partial h_{ij}}(\kappa_{4,1}^2 + \kappa_{4,2}^2 + 24 \log|\frac{dy_i}{de_i}|) & \forall j = 1, .., m \\ w_{ji}(t+1) = w_{ji}(t) + \mu_t \frac{\partial}{\partial w_{ji}}(\kappa_{4,1}^2 + \kappa_{4,2}^2 + 24 \log|\frac{dy_i}{de_i}|) & \forall j = 1, .., m \\ \theta_{ji}(t+1) = \theta_{ji}(t) + \mu_t \frac{\partial}{\partial \theta_{ji}}(\kappa_{4,1}^2 + \kappa_{4,2}^2 + 24 \log|\frac{dy_i}{de_i}|) & \forall j = 1, .., m. \end{cases} \tag{10}$$

Without normalization, the parameters h_{ij} are not bounded and may tend towards infinity. It corresponds to scale indeterminacy of the output. To avoid this, we limit the gain of the output of each MLP, by normalizing the vectors $\mathbf{h}_i = (h_{i1}, \cdots, h_{im})^T$, $i = 1, \cdots, n$.

Simulation results in the case of two sources are shown in Fig. 2. The PNL mixtures are drawn in Fig. 2 (middle), and the estimated sources in Fig. 2 (right) show the independent components analysis provided by the algorithm. In figure 3, the joint distribution of the two outputs are nearly

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 279-284

rectangular which explains good results of the algorithm. In this experiment, $\mathcal{F}_1(x) = x + 5x^3$ and $\mathcal{F}_2(x) = \sinh(3x)$, and x is in the range $[-0.5, 0.5]$.

6. Conclusion

In the general nonlinear case, sources can be estimated only up to any nonlinear function, however we prove that in post-nonlinear mixtures, estimated sources are linear functions of original sources, as for instantaneous linear mixtures.

For nonlinear mixtures, the ML approach leads to a criterion consisting of two terms: the contrast function proposed by Comon for linear mixtures, and an extra term, $\log |det(J_{\mathcal{F}^{-1}})|$ corresponding to local variations of the mixture.

References

[1] S. Amari, A. Cichocki, and H. H. Yang. Recurrent neural networks for blind separation of sources. In *NOLTA 95*, pages 37–42, Las Vegas (USA), December 1995.

[2] G. Burel. Blind separation of sources: a nonlinear neural algorithme. *Neural Networks*, 5:937–947, 1992.

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

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

[5] M. Gaeta and J.-L. Lacoume. Source separation without a priori knowledge: the maximum likelihood solution. In *EUSIPCO 90*, volume 2, pages 621–624, Barcelona (Spain), September 1990.

[6] C. Jutten and J. Hérault. Blind separation of sources, Part I: An adaptive algorithm based on a neuromimetic architecture. *Signal Processing*, 24(1):1–10, 1991.

[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] H. H. Yang, S. Amari, and A. Cichocki. Information back-propagation for blind separation of sources from non-linear mixture, November 23, 1996. Submitted to ICNN 97.