NONLINEAR ICA SOLUTIONS FOR CONVOLUTIVE MIXING OF PNL MIXTURES

Daniele Vigliano[(*)]; Aurelio Uncini[(*)] ; Raffaele Parisi [(*)]

[(*)]Dipartimento INFOCOM, Università di Roma "La sapienza" – Italy
Via Eudossiana, 18, 00184 Roma – Italy
d.vigliano@inwind.it; aurel@ieee.org; parisi@infocom.uniroma1.it;

ABSTRACT

This paper introduces an ICA approach to a novel nonlinear convolutive BSS problem. The mixing model considered here is an evolution of the Post Nonlinear one: it is the convolutive mixing of a PNL mixture. The main aim of this paper is to enlarge the set of blind sources separation problems that can be approached by Nonlinear ICA with some stricter mixing environments than the one just widely described in literature. The Flexibility of the algorithm is given by the on line estimation of the score function performed by Spline Neurons.

1. INTRODUCTION

The Scientific Community's interest for blind signal processing and in particular for Blind Sources separation (BSS), performed through Independent Component Analysis (ICA) is growing. A large number of problems in biomedical or communication field now can be approached by ICA. The first studies about Independent Component Analysis aim only at resolving the famous Cocktail party problem first in static then in reverberant environments. Recently the so called flexible ICA improves the pdf matching into the neural network processing and it provides a faster learning by the estimation of parameters related with the pdf of the signals. A critical issue about the BSS algorithms is that linear mixing model is too unrealistic and "poor" in a lot of real situations.

In [4] Hyvarinen and Pajunen give an important result in nonlinear ICA theory, exploring the existence of a solution of a nonlinear problem, underlining the hard non uniqueness of the provided solution and also proposing some constraints about the mixing model granting the uniqueness of the solution. Now a large number of papers explore solutions to Post Nonlinear Mixing problem (PNL) [6][3][2]. It is possible to find only few results of the convolutive post nonlinear problem ([9][10]) and of some static nonlinear problems [5] more complex than PNL. In [1] Jutten and Karhunen review the recent advances in BSS of nonlinear mixing models. This paper explores the solution of the BSS problem in a novel convolutive nonlinear mixing environment, stricter than the

one just widely diffused in literature. The new mixing environment is composed by several mixing block: a convolutive mixing channel follows a PNL block.

2. THE NONLINEAR ISSUE

The aim of this section is at introducing BSS problem in nonlinear environment underlining the approaches and exploring the existing algorithms in term of uniqueness and existence of the solution.

Considering an N vector of independent sources $\mathbf{s}[n] = \{s_1[n],..,s_N[n]\}$; considering a vector of signals received by a N-sensor array after a generally nonlinear convolutive mixing $\mathbf{x}[n] = \{x_1[n],..,x_N[n]\}$ of the original sources. The nonlinear convolutive environment was introduced in order to achieve a mixing model closer to the real one than the others just explored in literature. The general formulation of the hidden mixing model is:

$$\mathbf{x}(n) = \mathcal{F}\{\mathbf{s}(n),...,\mathbf{s}(n-L)\} \qquad (1)$$

in which $\mathcal{F}\{.\}$ is a dynamic nonlinear distorting function.

The solution of the BSS problem is expressed as: $\mathbf{y}[n] = \mathcal{G} \circ \mathcal{F}\{\mathbf{s}(n)\}$.

Resolving the blind sources separation problem in this context means to recover, making no particular a-priori assumptions, the original sources from the observation only of $\mathbf{x}[t]$. Into performing the separation, ICA recovers the original sources up to some trivial non-uniqueness[1]. The desired solution can be expressed in a closed form as:

$$\mathbf{y}[n] = \mathbf{P\Lambda D s}[n] \qquad (2)$$

in which $\mathbf{P}$ is a permutation matrix, $\mathbf{\Lambda}$ is a diagonal scaling matrix and $\mathbf{D}$ is a diagonal delay matrix.

The issue of separating mixture from models (1) with the only constraint of independent output signals and no other a priori assumption is affected by a strong non uniqueness, it can be shown with the following well known example. Considering two random variables s_1, s_2,

---

[1] Non-uniqueness consists into permutation, scaling factor and time delay of the original sources.

with the joint pdf $p_{s_1,s_2}(s_1,s_2) = \begin{cases} s_1 e^{-s_1^2} & s_1 \in \mathbb{R}^+ \\ 1/2\pi & s_2 \in [0, 2\pi[\end{cases}$

processed by the nonlinear transform: $[y_1, y_2] = \mathcal{H}(\mathbf{s}) = [s_1 \cos(s_2), s_1 \sin(s_2)]$. As is well known in the literature the resulting random variables y_1, y_2, are independent too and have gaussian distribution. This shows that independence conservation constraint alone is not strong enough to recover original sources from a generic nonlinear mixing environment [2]. If the transform $\mathcal{H}(.)$ has no particular structure or there are no more assumptions about the mixing (and demixing) model, the provided results could not be the desired solution. The main issue for generic nonlinear problems is to ensure the presence of conditions (in term of sources, mixing environment, recovering structure) granting at least theoretically the possibility to achieve the desired solution; that is the same as investigating the existence and uniqueness of the solution to the given problem. In [4] authors proposed a constructive way (a Gram-Schmidt like method) to obtain solutions of the separation problem in a static nonlinear mixing environment. In order to grant the uniqueness of the solutions it was applied some constraints to the mixing environment. The idea introduced, is general: adding some "soft" constraint to the problem produces the uniqueness of the solution.

In [2] a theoretical demonstration of the existence and uniqueness of the solution is given when the problem is PNL with convolutive mixing. It is possible to find in [1] a short explanation of many other ICA approach to nonlinear BSS. In [5] was performed the solution in the case of PNL static model followed by static mixing, stricter static nonlinear model that adds static mixing after the PNL mixing model. Thus the output independence is a weak approach to the problem of sources separation in a general nonlinear environment. In static and convolutive nonlinear mixing it is impossible to perform the recovering of the desired signals making no other assumption [1].

3. THE MIXING-DEMIXING STRUCTURE

This section explores the recovery of original sources from nonlinear convolutive mixing assuming a-priori information about the mixing model; the information assumed for the mixing model are used to design the recovering network. The mixing environment modelled in this paper is represented in Fig. 1.

In which $\mathbf{A}$ is a NxN static matrix, $\mathbf{F}[\mathbf{r}(n)] = [f_1[r_1(n)], \quad f_N[r_N(n)]]^T$ is the Nx1 vector of nonlinear distorting functions, one for each channel, and $\mathbf{Z}[k]$ is a FIR matrix with L taps filters.
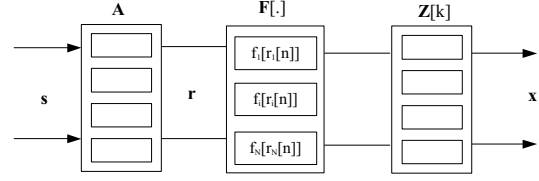

Fig. 1 The Block diagram of the convolutive nonlinear mixing model.

The mixing model written in a closed form is:

$$\mathbf{x}[n] = \mathcal{F}[\mathbf{s}] = \sum_{k=0}^{L-1} \mathbf{Z}[k]\mathbf{F}[\mathbf{As}[n-k]] \qquad (3)$$

This mixing model enlarges the set of possible mixing environments from which is possible to recover separated signals. The recovering structure mirrors the mixing model.
The close form for the recovered outputs is:

$$\mathbf{y}[n] = \mathcal{G}[\mathbf{x}] = \mathbf{BG}\left[\sum_{k=0}^{L-1} \mathbf{W}[k]\mathbf{x}[n-k]\right] \qquad (4)$$

In which $\mathbf{B}$ is a NxN static matrix, $\mathbf{G}\left[\mathbf{W}[n]*\mathbf{x}[n]\right]$ is the Nx1 vector of nonlinear compensating functions, one for each channel, and $\mathbf{W}[k]$ is a FIR matrix with K taps filters. Introducing the knowledge about the particular kind of mixing model is the key to avoid the strict nonuniqueness of the solution; such assumption limits the weakness of the output independence condition reducing the cardinality of all possible independent output solutions; with this constraint the problem of recovery the original sources isn't ill posed any more.

4. THE DEMIXING ALGORITHM AND NETWORK MODEL

This section explores the blind demixing algorithm, the adaptive network and the network used to compensate the nonlinear distortion. The blind algorithm performs an on-line adaptive learning of the network parameters on the base of the output independence estimation. Considering N observations $[x_1(k),...,x_N(k)]$ over a (M+1)-point time block and the corresponding N outputs over the same time block, defining the vectors $\mathcal{X} = [\mathbf{x}(0),...,\mathbf{x}(M)]^T$, $\mathcal{Y} = [\mathbf{y}(0),...,\mathbf{y}(M)]^T$. In this work the Kullback-Leibler divergence between $p_{\mathcal{Y}}[\mathbf{y}]$ and $\tilde{p}_{\mathcal{Y}}[\mathbf{y}] = \prod_{i=1;j=1}^{N,M} p_{y_i}[y_i[j]]$ is used as a way to quantify the output independence. Considering a specific demixing model with parameters $\mathbf{\Phi}$, the cost function of the algorithm is:

$$\Im\{\mathbf{y}[n],\mathbf{\Phi}\} = KL\left(p_{\mathbf{y}},\prod_{i=1;j}^{N} p_{y_i}\left[y_i[j]\right]\right) = \tag{5}$$

$$= \int_{\Im} p_{\mathbf{y}}(\mathbf{y}) \ \log\left(\frac{p_{\mathbf{y}}(\mathbf{y})}{\prod_{i,j} p_{y_i}\left[y_i[j]\right]}\right) d\mathbf{y}$$

The (5) is function of both output signals and model parameters $\mathbf{\Phi}$. KL divergence is minimized with respect of the model's parameters using both steepest descendent (stochastic gradient) and Natural gradient. Minimizing the cost function (5) with respect of the algorithm's parameters shows some terms like:

$$\frac{\partial}{\partial\mathbf{\Phi}}\log\left[p_{y_i}(y_i)\right] = \frac{\dot{p}_{y_i}(y_i)}{p_{y_i}(y_i)}\frac{\partial y_i}{\partial\mathbf{\Phi}} = \psi_i(y_i)\frac{\partial y_i}{\partial\mathbf{\Phi}} \tag{6}$$

in which $\psi_i(y_i) = \dot{p}_i(y_i)/p_i(y_i)$ are the so called Score Functions[2]. In this paper, the Spline Neurons are used to perform the on-line estimation of both Score function and nonlinear compensating functions (for a detail about the Spline Neurons see [5][6][7][8]). The most suitable property of the Spline Neurons, as function estimator, is theirs local learning: for each learning step only the four control points nearest to the training input are considered; no matter how many control points the Spline curve has. In this paper the learning of the score function is performed in a direct way (see [2]) minimizing, with respect to the spline control points $\mathbf{Q}$, the:

$$\varepsilon_j = \frac{1}{2}E\left\{\left[\tilde{\psi}_j(y_j,\mathbf{\Phi}) - \frac{\dot{p}_{y_j}(y_j)}{p_{y_j}(y_j)}\right]^2\right\} \quad j=1..N \tag{7}$$

$\tilde{\psi}_j$ is the Spline model of the score function. The resulting rule for the Spline Neuron is:

$$\frac{\partial\varepsilon}{\partial\mathbf{Q}_{i,(i+m)}^{\psi}} = \left[\frac{1}{4}\mathbf{T_u M}_m \mathbf{T_u M Q}_i^{g_j} + \frac{1}{\Delta y}\dot{\mathbf{T}}_u \mathbf{M}_m\right] m=0..3 \tag{8}$$

in which $\mathbf{M}_m$ is the m-th column of the matrix $\mathbf{M}$ (matrix of coefficients), $\mathbf{T}$ is the local abscissa vector.
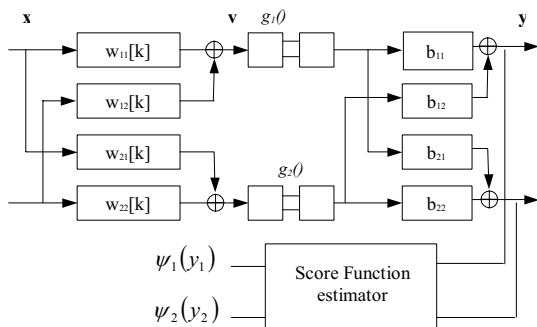

Fig. 3 Feed Forward proposed network for the nonlinear blind deconvolution and separation.

---

[2] Sometimes in literature Score Function is defined considering the opposite of the function (6).

Fig 3 shows the network used to perform the separation; this network is a cascade of blocks, each of one is well described in literature and previously used to resolve more simple problems. The output of the channel i-th is expressed in close form as:

$$y_i[n] = \sum_{j=1}^{N} b_{ij}g_j\left[\sum_{h=1}^{N}\sum_{k=0}^{L-1} w_{jh}[k]x_h[n-k]\right] \quad i=1..N \tag{9}$$

Considering the set of the learning parameters $\mathbf{\Phi} = \left\{b_{ij}, w_{pq}[k], \mathbf{Q}^g, \mathbf{Q}^\Psi\right\}$ in which $\mathbf{Q}^g$ and $\mathbf{Q}^\Psi$ are the parameters of the Spline networks. Deriving the cost function (5) with respect the learning parameter results:

$$\frac{\partial\Im\{\mathbf{y}[n],\mathbf{\Phi}\}}{\partial\mathbf{\Phi}} = \frac{\partial}{\partial\mathbf{\Phi}}\sum_{n=0}^{M}\left[-\log|\det\mathbf{B}| + \right. \tag{10}$$

$$\left. -\log\prod_{i=1}^{N}g_i\left[v_i[n]\right] - \log|\mathbf{W}(0)| - \sum_{i=1}^{N}\log p_{y_i}(y_i)\right]$$

In (10) the expected value of the signals has been replaced by the instantaneous value. In order to minimize $\Im\{\mathbf{\Phi},\mathbf{y}\}$ the general rule provided by the steepest descendent method is:

$$\mathbf{\Phi}(k+1) = \mathbf{\Phi}(k) - \eta_\mathbf{\Phi}(k)\frac{\partial}{\partial\mathbf{\Phi}}\Im\{\mathbf{\Phi},\mathbf{y}\} \tag{11}$$

The learning rule for the elements of the matrix $\mathbf{B}$ is:

$$\frac{\partial}{\partial\mathbf{\Phi}}\Im\{\mathbf{\Phi}(k),\mathbf{y}\} = -\left(I - \mathbf{\Psi}_\mathbf{y}\mathbf{y}^T\right)\mathbf{B} \tag{12}$$

The learning rule for the elements of the FIR matrix $\mathbf{W}$ is:

$$\frac{\partial\Im}{\partial\mathbf{W}[k]} = -\mathbf{W}[0]^{-T}\delta_k - \left\{\left[\frac{\ddot{g}_1(v_1)}{\dot{g}_1(v_1)}\cdots\frac{\ddot{g}_N(v_N)}{\dot{g}_N(v_N)}\right]^T + \right. \tag{13}$$

$$\left. + \begin{bmatrix}\mathbf{\Psi}_\mathbf{y}^T(\mathbf{B})_1 & 0 \\ 0 & \mathbf{\Psi}_\mathbf{y}^T(\mathbf{B})_N\end{bmatrix}\begin{bmatrix}\dot{g}_1(v_1) \\ \dot{g}_N(v_N)\end{bmatrix}\right\}\mathbf{x}^T(n-k)$$

in the last equation $(\mathbf{B})_m$ is used as the m-th column of the matrix $\mathbf{B}$. The learning rules for the control points $\mathbf{Q}^g$ of the Spline neurons that compensate the nonlinear distorting functions are:

$$\frac{\partial\Im\{\mathbf{\Phi}(k),\mathbf{y}\}}{\partial\mathbf{Q}_{i,(i+m)}^g} = -\left[\frac{\dot{\mathbf{T}}_u\mathbf{M}_m}{\dot{\mathbf{T}}_u\mathbf{M Q}_i^{g_j}} + \mathbf{\Psi}_\mathbf{y}^T(\mathbf{B})_j\mathbf{T_u M}_m\right] m=0,..,3 \tag{14}$$

in which $\mathbf{M}_m$ and $\mathbf{T}$ have the same meaning as in (8). The learning rule for the Spline SG neurons dedicated to Score Functions is (8). The use of FIR allows the inverting also quite non-minimum phase systems (with zeros near to the unit circle) that could have convergence problem in other situations. One of the main problem using FIR is the length of filters: to resolve real problems or simply some non trivial ones a large number of filter taps is required; must be noted that learning time grow in a nonlinear way with the FIR length.

4. EXPERIMENTAL RESULTS

This section collects the experimental result of the proposed architectures. Although the algorithm is able to perform the separation of n-channel mixture in this test has been considered the mixture of a male voice (speaking: "Le donne i cavalier l'arme") and white noise.
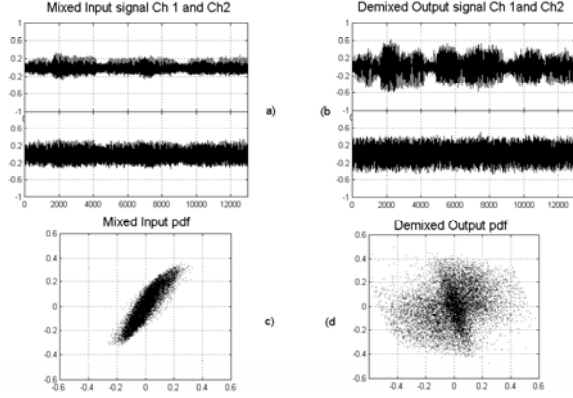

Fig.4   a) Input mixture; b) Output demixed signals; c) Input mixd Pdf; d) Output demixed Pdf.

Fig. 4 shows the resulting signals after a training of 900 epochs with the learning rat e set $\mu=7*10^{-7}$; the recovering network has 103 control points Spline neurons and a 31 taps Fir matrix. The mixing environment applied, with respect to fig.1, is: **A**=[0.7, -0.3; 0.5, 0.7];

$$\mathbf{Z}\left[z^{-1}\right]=\begin{bmatrix} 0.7-0.3z^{-1}+0.6z^{-2} & 0.3-0.2z^{-1}-0.06z^{-2} \\ -0.3+0.3z^{-1}+0.11z^{-2} & 0.7+0.3z^{-1}-0.06z^{-2} \end{bmatrix}.$$

The nonlinear distortions applied in this test are:

$$F\left[f_1\left(r_1\right),f_2\left(r_2\right)\right]=\left[r_1+2r_1^{3},0.5r_2+\tanh\left(5r_2\right)\right].$$

The Signal Interface Ratio (SIR dB) introduced in [11] measures the performance of the proposed algorithm.
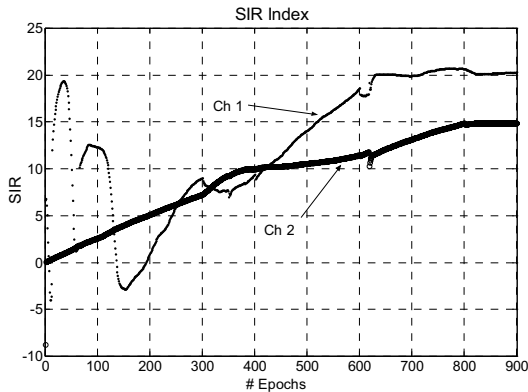

Fig. 5   Signal interference ratio index during the training

The SIR index trend during the learning shows the performance of the algorithm in a quantitative way.

Fig 5 shows how, after a first period, the algorithm perform the separation of the output signals. The reason of the starting transient has been the bad starting condition and the number of block each of one separately have to converge to the optimum values.

5. CONCLUSION

This paper introduces a novel mixing model for which the BSS performed by ICA is granted. The algorithm widely explored in this paper is able to separate the convolutive mixing of a PNL mixture.   The FIR recovering network performs the on line estimation of the score function by the Spline Neurons. Spline Neurons perform also the nonlinear compensating function estimation.

REFERENCES

[1]   C. Jutten, J. Karhunen, "Advances in Nonlinear Blind Sources Separation", 4th International Symposium on Independent Component Analysis and B lind Signal Separation (ICA2003), April 2003, Nara, Japan.
[2]   A. Taleb, "A Generic Framework for Blind Sources Separation in Structured Nonlinear Models", in IEEE transaction on signal processing, vol. 50. no 8 August 2002.
[3]   A. Taleb, C. Jutten, "Sources Separation in post nonlinear mixtures", in transaction on signal processing, vol. 47. no 10 August 1999.
[4]   A. Hyvarinen, P. Pajunen, "Nonlinear Independent Component Analysis: Existence and Uniqueness Results", Neural Networks 12(2): 429-439, 1999.
[5]   M. Solazzi, F. Piazza, A. Uncini, "Nonlinear Blind Source Separation by Spline neural Network", ICASSP 2001, Salt Lake city, USA May 8-11, 2001.
[6]   M. Solazzi, R. Parisi, A. Uncini, "Blind Source separation in nonliner mixtures by adaptive spline neural network", ICA 2001 in Proc. of the 3rd Workshop on Independent Component Analysis and Signal Separation (ICA2001), San Diego (California, USA), 2001.
[7]   A. Uncini, L. Vecci, F. Piazza, "Learning and approximation capabilities of adaptive Spline activation function neural network", NN, Vol. 11, no. 2, pag. 259-270 March 1998.
[8]   M.Solazzi, F. Piazza, A. Uncini, "An adaptive Spline Nonlinear Function for Blind Signal Proessing", proc. of IEEE Whorkshop on neural networks for signal Processing X, pp396-404, December, 2000.
[9]   F. Milani, M. Solazzi, A. Uncini, "Blind Source Separation of convolutive nonlinear mixtures by flexible spline nonlinear functions", proc. of IEEE Int. Conference on Acoustic Speech and Signal Processing, ICASSP'02, Orlando, USA, May, 2002.
[10] M. B. Zade, C. Jutten, K. Najeby, "Blind Separating, Convolutive Post nonlinear Mixture", ICA 2001 in Proc. of the 3rd Workshop on Independent Component Analysis and Signal Separation (ICA2001), San Diego (California, USA), 2001, pp. 138–143.
[11] D. Shobben, K. Torkkola, and P. Smaragdis, "Evaluation of blind signal separation methods", in proceeding of ICA and BSS, Aussois, France, January 11-15, 1999.