

A NOVEL RECURRENT NETWORK FOR INDEPENDENT COMPONENT ANALYSIS OF POST NONLINEAR CONVOLUTIVE MIXTURES

Daniele Vigliano, Raffaele Parisi and Aurelio Uncini

Dipartimento INFOCOM, Università di Roma “La sapienza” – Italy

Via Eudossiana, 18, 00184 Roma – Italy

daniele.vigliano@poste.it; parisi@infocom.uniroma1.it; aurel@ieee.org;

ABSTRACT

This paper introduces a novel Independent Component Analysis approach to the separation of nonlinear convolutive mixtures. In particular, convolutive mixing of *post nonlinear* mixtures is considered. Source separation is performed by a new efficient recurrent network, which is able to ensure faster training with respect to currently available feedforward architectures, with lower computational costs. The proposed architecture makes proper use of *flexible* spline neurons for *on-line* estimation of the score function. Experimental results are described to demonstrate the effectiveness of the proposed technique.

1. INTRODUCTION

The interest of the scientific community in Blind Source Separation (BSS) performed through Independent Component Analysis (ICA) is growing. Initial studies about ICA aimed only at solving the well-known *cocktail party problem* in a static or slightly reverberant environment. Unfortunately linear mixing models are too unrealistic and unsatisfactory in a lot of real situations. In [4] Hyvarinen and Pajunen gave an important result in the context of *nonlinear* ICA theory, in particular exploring the problem of solution existence. Recently many papers provided solutions to the so-called Post Nonlinear Mixing problem (PNL) [6][2], but only a few works addressed the convolutive post nonlinear problem ([9][10]). In [1] Jutten and Karhunen reviewed most recent advances in BSS of nonlinear mixing models. In [3] a possible approach to source separation in the presence of a convolutive mixing environment was firstly described.

This paper explores the solution of the BSS problem in a convolutive nonlinear mixing environment, which is more general with respect to the ones currently available in literature. The new mixing environment is composed by the cascade of a PNL block, followed by a convolutive mixing channel. Estimation of the score function is an important issue in the presence of non-trivial mixing environments. The ICA approach hereafter proposed is able to perform the score function estimation in an online fashion,

providing a faster and better learning with respect to existing algorithms.

2. THE NONLINEAR PROBLEM

The aim of this section is to introduce the BSS problem in nonlinear environments, underlining the problem of existence and uniqueness of the solution. Let $\mathbf{s}[n]$ be the vector of the N hidden independent sources, while $\mathbf{x}[n] = \mathcal{F}\{\mathbf{s}[n], \dots, \mathbf{s}[n-L]\}$ is the vector of signals received by a N -sensor array after a general nonlinear convolutive mixing (represented by \mathcal{F}). BSS tries to recover the original sources from the observation of $\mathbf{x}[n]$ only, with no particular a-priori assumptions. The recovering performed by ICA is up to some trivial ambiguity¹. The desired solution can be expressed in a closed form as:

$$\mathbf{y}[n] = \mathbf{P}\mathbf{A}\mathbf{D}\mathbf{s}[n] = \mathcal{E} \circ \mathcal{F}\{\mathbf{s}\} \quad (1)$$

where \mathbf{P} is a permutation matrix, \mathbf{A} is a diagonal scaling matrix, \mathbf{D} is a diagonal delay matrix and \mathcal{E} is the nonlinear recovering model.

The issue of separating mixtures with the constraint of the independence of the output signals and no other a priori assumptions is hampered by a strong non uniqueness. It is known in fact that independence preservation alone is not sufficient to recover the original sources from a generic nonlinear mixing environment [2]. If the transform \mathcal{E} has no particular structure, results could be highly unsatisfactory. The main issue for general nonlinear problems is to ensure the presence of conditions (in terms of sources, mixing environment, recovering structure) that grant the existence and uniqueness of the solution. In [4] a constructive way (a Gram-Schmidt like method) to obtain solutions of the separation problem in a nonlinear mixing environment was proposed. Authors demonstrated that the proposed solution was not unique, but also suggested that non uniqueness could be obtained by adding some “soft” constraint to the problem. More specifically, constraints were added to the mixing environment. In [2] a theoretical

¹ Non-uniqueness consists of permutation, scaling and time delaying of the original sources.

demonstration of the existence and uniqueness of the solution was given in the case of PNL with convolutive mixing. It is possible to find in [1] a short description of many other ICA approaches to nonlinear BSS. As a matter of fact, output independence is a weak approach to the problem of source separation in a general nonlinear environment. In the presence of static and convolutive nonlinear mixing it is impossible to recover the desired signals with no other assumptions [1].

3. THE MIXING-DEMIXING MODEL

This section describes the proposed mixing-demixing structure. The demixing structure is designed on the basis of a-priori information about the mixing model. Specifically, the model of the proposed mixing environment is represented in Fig. 1. In the figure, \mathbf{A} is a $N \times N$ static matrix, $\mathbf{F}[\mathbf{r}(n)] = [f_1[r_1(n)], f_N[r_N(n)]]^T$ is the $N \times 1$ vector of nonlinear distorting functions, one for each channel, and $\mathbf{Z}[k]$ represents a convolutive mixing channel.

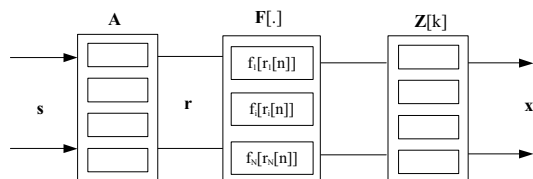


Fig. 1 Block diagram of the proposed convolutive nonlinear mixing model.

The mixing model can be thus represented by

$$\mathbf{x}[n] = \mathcal{F}[\mathbf{s}] = \sum_{k=0}^{L-1} \mathbf{Z}[k] \mathbf{F}[\mathbf{A}\mathbf{s}[n-k]] \quad (2)$$

This model enlarges the set of possible mixing environments from which it is possible to recover separated signals [3].

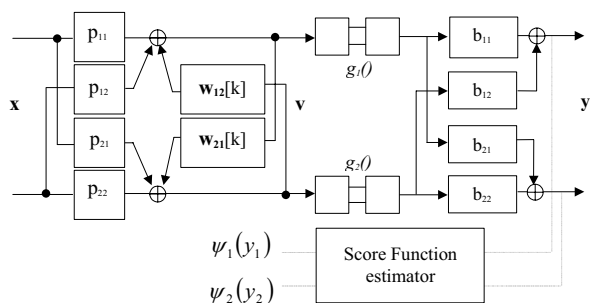


Fig. 2 Proposed recurrent network for nonlinear blind deconvolution and separation.

The recovering structure mirrors the mixing model but the block that performs the deconvolution of the mixing channel is realized in an IIR way. MIMO recurrent neural network approaches to the separation of simply convolved signals has been extensively studied [14][5][12] (see the references in

[13] for a complete overview about this issue). In this work the hybrid architecture described in [12] is adapted to the novel recovering problem here addressed. Fig. 2 shows the recovering structure, with $N=2$ for sake of simplicity. The expression of the i -th output channel is:

$$y_i[n] = \sum_{j=1}^N b_{ij} g_j[v_j[n]] \quad (3)$$

$$v_j[n] = \sum_{h=1}^N p_{jh} x_h[n] + \sum_{\substack{h=1 \\ h \neq j}}^N \sum_{k=1}^{L_i} w_{jh}[k] y_h[n-k]$$

where b_{ij} is an element of the $N \times N$ static matrix, $g_j[v_j[n]]$ is an element of the $N \times 1$ vector of nonlinear compensating functions, one for each channel. Knowledge of the particular kind of mixing model is the key to avoid the strict non uniqueness of the solution. Such an assumption limits the weakness of the output independence condition, reducing the cardinality of all possible independent output solutions. With these additional constraints, the problem of recovering the original sources is not ill-posed any more.

The transfer matrix of the convolutive block of the architecture shown in fig.2 is:

$$\frac{\mathbf{v}[z]}{\mathbf{x}[z]} = \frac{1}{1 - w_{12}w_{21}} \begin{bmatrix} p_{11} + p_{21}w_{12}[z] & p_{12} + p_{22}w_{12}[z] \\ p_{21} + p_{11}w_{21}[z] & p_{22} + p_{21}w_{21}[z] \end{bmatrix} \quad (4)$$

Spline estimation neurons perform the estimation of both nonlinear compensating function $g_j[v_j[n]]$ and score functions. More specifically, spline estimation neurons provide a model $h(x) \cong \hat{y} = \hat{h}(u(x), i(x))$, which is a function of two parameters (u, i), directly derived from the independent one. The spline curve results as a composition of $N-1$ spans, being N the number of control points. Thus evaluation of (u, i) from x , allows to compute $\hat{y}(x) = \hat{h}(u, i) = 0.5 \mathbf{T}_u \mathbf{M} \mathbf{Q}_i$, where \mathbf{M} is a matrix of coefficients, $\mathbf{T}_u = [u^3 \ u^2 \ u^1 \ 1]$ is the vector of the local abscissa and $\mathbf{Q}_i = [Q_i \ Q_{i+1} \ Q_{i+2} \ Q_{i+3}]^T$ is the vector of control points. The local learning capability is the most attractive property of Spline Neurons: 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. See [6][7][8] for more details about Spline approximation theory and about the efficiency of this kind of network.

4. THE DEMIXING ALGORITHM

This section explores the blind demixing algorithm. Consider N observations $[x_1(k), \dots, x_N(k)]$ over a $(M+1)$ -point time block and the corresponding N outputs over the same time block. This defines the vectors $\mathcal{X} = [\mathbf{x}(0), \dots, \mathbf{x}(M)]^T$ and $\mathcal{Y} = [\mathbf{y}(0), \dots, \mathbf{y}(M)]^T$. In this paper the Kullback-

Leibler (KL) divergence between $p_{\mathbf{r}}[\mathbf{y}]$ and $\tilde{p}_{\mathbf{r}}[\mathbf{y}] = \prod_{i=1, j=1}^{N, M} p_{y_i}[y_i[j]]$ quantifies the output independence. Considering a specific demixing model with parameters Φ , the cost function of the algorithm is:

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

The cost function (5) depends on both output signals and model parameters. KL divergence is minimized with respect to Φ by both stochastic and natural gradient approaches. In deriving learning rules, the following terms requires particular attention:

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

where $\psi_i(y_i) = \dot{p}_i(y_i)/p_i(y_i)$ is the score function². The learning of the spline neurons that provides the score function estimation is performed in a direct way (see [2]), by minimization of

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

where $\tilde{\psi}_j$ is the spline model of the score function. The resulting learning rule for the spline neuron is

$$\frac{\partial \varepsilon}{\partial \mathbf{Q}_i^\psi} = \left[\frac{1}{4} (\mathbf{T}_u \mathbf{M}) \mathbf{T}_u \mathbf{M} \mathbf{Q}_i^{\psi_i} + \frac{1}{\Delta y} \dot{\mathbf{T}}_u \mathbf{M} \right] \quad (8)$$

being \mathbf{M} a matrix of coefficients, \mathbf{T}_u the local abscissa vector and $\dot{\mathbf{T}}_u = [3u^2 \quad 2u \quad 1 \quad 0]$.

Considering the set of the learning parameters $\Phi = \{b_{ij}, w_{pq}[k], \mathbf{Q}^g, \mathbf{Q}^\psi\}$, where \mathbf{Q}^g and \mathbf{Q}^ψ are the parameters of the spline networks, from (5) it is possible to obtain

$$\frac{\partial \mathfrak{J}\{\mathbf{y}, \Phi\}}{\partial \Phi} = \frac{\partial}{\partial \Phi} \sum_{n=0}^M \left[-\log |\det \mathbf{B}| - \log |\det \mathbf{P}| + \right. \\ \left. -\log \prod_{i=1}^N g_i[v_i[n]] - \sum_{i=1}^N \log p_{y_i}(y_i) \right] \quad (9)$$

In (9) the expected value of the signals has been replaced by the instantaneous value. The learning rule for the elements of the matrix \mathbf{B} is:

$$\frac{\partial}{\partial \Phi} \mathfrak{J}\{\Phi, \mathbf{y}\} = -(I - \Psi_y \mathbf{y}^T) \mathbf{B} \quad (10)$$

The learning rule for the elements of the static matrix \mathbf{P} is:

$$\frac{\partial \mathfrak{J}}{\partial \mathbf{P}} = -\mathbf{P} + \left(\begin{array}{c} \left[\begin{array}{c} \ddot{g}_1(v_1) \\ \dot{g}_1(v_1) \\ \ddot{g}_N(v_N) \\ \dot{g}_N(v_N) \end{array} \right] + \left[\begin{array}{cc} \dot{g}_1(v_1) & 0 \\ 0 & \dot{g}_N(v_N) \end{array} \right] \Psi_y^T \mathbf{B} \end{array} \right) \mathbf{v}^T [n] \mathbf{P} \quad (11)$$

The learning rule for the IIR filter taps \mathbf{W} is:

$$\left\{ \begin{array}{l} \frac{\partial \mathfrak{J}}{\partial \mathbf{w}_{12}[k]} = - \left(\begin{array}{c} \ddot{g}_1(v_1) \\ \dot{g}_1(v_1) \end{array} + \Psi_y^T(\mathbf{B})_1 \dot{g}_1(v_1) \right) v_2[n-k-1] \\ \frac{\partial \mathfrak{J}}{\partial \mathbf{w}_{21}[k]} = - \left(\begin{array}{c} \ddot{g}_2(v_2) \\ \dot{g}_2(v_2) \end{array} + \Psi_y^T(\mathbf{B})_2 \dot{g}_2(v_2) \right) v_1[n-k-1] \end{array} \right. \quad k=0..L-1 \quad (12)$$

The learning rules for the control points \mathbf{Q}^g of the Spline neurons that compensate for the nonlinear distorting functions are:

$$\frac{\partial \mathfrak{J}\{\mathbf{y}, \Phi\}}{\partial \mathbf{Q}_i^g} = - \left[\frac{\dot{\mathbf{T}}_u \mathbf{M}}{\dot{\mathbf{T}}_u \mathbf{M} \mathbf{Q}_i^{g_i}} + \Psi_y^T(\mathbf{B})_j \mathbf{T}_u \mathbf{M} \right] \quad (13)$$

where \mathbf{M} and \mathbf{T} have the same meaning as in (8).

The use of a recurrent network to separate convolutive mixtures solves one of the main problems occurring when using feedforward networks, that is the high number of taps needed in separating FIR filters. The same problem can be solved by a recurrent network with a lower number of parameters if compared with a feed forward one; a reduced number of parameters mean a faster learning.

4. EXPERIMENTAL RESULTS

This section collects the experimental result of the proposed architecture. Although the algorithm is able to perform the separation of n-channel mixtures, in this test the mixture of a male voice (speaking: ‘‘Le donne i cavalier l’arme’’) a female voice (speaking ‘‘Una donna indegna’’) was considered. Fig. 3 shows the resulting signals after a training of 900 epochs with the learning rate set to $\mu=7*10^{-7}$; the recovering network had 103 control points spline neurons and 11 tap filters for the recurrent network. The mixing environment applied was $\mathbf{A}=[0.7, -0.3; 0.5, 0.7]$ and

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

The nonlinear distortions applied in this test were:

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

The Signal Interference Ratio (SIR dB) introduced in [11] measures the performance of the proposed algorithm. The SIR index trend during the learning shows the performance of the algorithm in a quantitative way.

Fig. 3 and Fig. 4 show how the algorithm is successful in performing the separation of the output signals. In other experiments the same mixing environment was approached with a FIR-based architecture. The separation

² Sometimes in literature score functions are defined by considering (6) with opposite sign.

performance was quite the same, but while the FIR matrix required 31-tap filters, the recurrent structure required 11-tap filters. This led to a significant gain in terms of computational effort. In this experiment, signals are represented within the range $[-1, 1]$, note that to produce a sensible distortion input s are normalized so that v_1 and v_2 spread the range $[-0.8, 0.8]$.

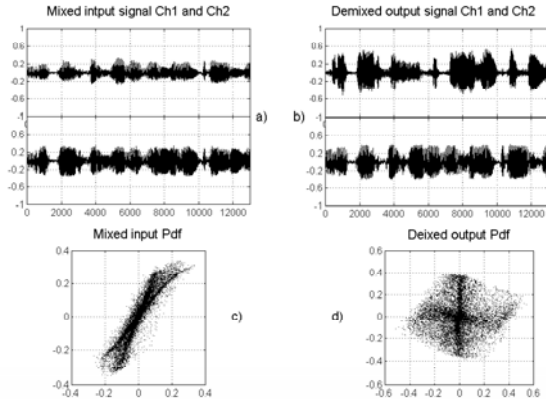


Fig.3 a) Input mixture; b) Output demixed signals; c) Input mixed Pdf; d) Output demixed Pdf.

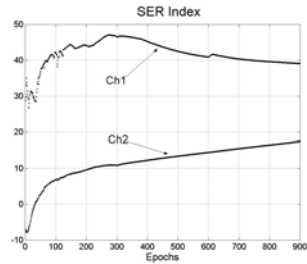


Fig. 4 Signal interference ratio index during training

Fig. 5 shows the performance of the spline neurons in score function estimation (fig.5-c,d) and nonlinear compensating functions (fig. 5-a,b).

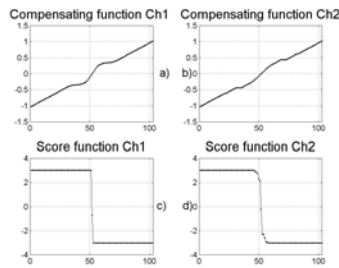


Fig. 5 a-b) Non linear Compensating function; c-d) score functions

5. CONCLUSION

This paper introduces a novel demixing recurrent algorithm for the separation of the convolutive mixing of

PNL mixtures. The proposed recurrent network is able to perform online estimation of the score function by use of spline neurons. Spline neurons perform also the nonlinear compensating function estimation. Use of recurrent networks yields a significant reduction in the number of parameters, with a considerable gain in terms of computational costs.

REFERENCES

- [1] C. Jutten, J. Karhunen, "Advances in Nonlinear Blind Sources Separation", 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan, April 2003.
- [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] D. Vigliano, A. Uncini, "Flexible ICA solution for a novel nonlinear blind source separation problem", to be published on IEE Electronic Letter on Signal Processing.
- [4] A. Hyvarinen, P. Pajunen, "Nonlinear Independent Component Analysis: Existence and Uniqueness Results", Neural Networks 12(2): 429-439, 1999.
- [5] K. Torkkola, "Blind deconvolution, information maximization, and recursive filters", In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, pag. 3301-3304, Munich, Germany, April 21-24, 1997.
- [6] M. Solazzi, R. Parisi, A. Uncini, "Blind Source separation in nonlinear 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 Processing", 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. Najeb, "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), pp. 138-143, 2001.
- [11] D. Shobben, K. Torkkola, P. Smaragdis, "Evaluation of blind signal separation methods", in proceeding of ICA and BSS, Aussois, France, January 11-15, 1999.
- [12] S. Choi, A. Cichocki, "Adaptive Blind Separation of speech signals: Cocktail party problem", ICSP97, Seoul, Korea, 26-28, pp. 617-622, August 1997.
- [13] K. Torkkola, "Blind separation of audio signals: Are we there yet?", In Proceedings of ICA and BSS, Aussois, France, pag. 239-244, January 11-15, 1999.
- [14] S. Choi, H. Hong, H. Glotin, F. Bertommier, "Multichannel Signal Separation for Cocktail Party Speech Recognition: A dynamic Recurrent Network", In Proc. ICSLP, Beijing, China, 2000.