

Nonlinear Blind Source Separation Using Hybrid Neural Networks*

Chun-Hou Zheng^{1,2}, Zhi-Kai Huang^{1,2}, Michael R. Lyu³,
and Tat-Ming Lok⁴

¹ Intelligent Computing Lab, Institute of Intelligent Machines,
Chinese Academy of Sciences, P.O.Box 1130, Hefei, Anhui, China

² Department of Automation,

University of Science and Technology of China

³ Computer Science & Engineering Dept.,

The Chinese University of Hong Kong, Hong Kong

⁴ Information Engineering Dept.,

The Chinese University of Hong Kong, Shatin, Hong Kong

zhengch@iim.ac.cn

Abstract. This paper proposes a novel algorithm based on minimizing mutual information for a special case of nonlinear blind source separation: post-nonlinear blind source separation. A network composed of a set of radial basis function (RBF) networks, a set of multilayer perceptron and a linear network is used as a demixing system to separate sources in post-nonlinear mixtures. The experimental results show that our proposed method is effective, and they also show that the local character of the RBF network's units allows a significant speedup in the training of the system.

1 Introduction

Blind source separation (BSS) in instantaneous and convolutive linear mixture has been intensively studied over the last decade. Most of the blind separation algorithms are based on the theory of the independent component analysis (ICA) when the mixture model is linear [1,2]. However, in general real-world situation, nonlinear mixture of signals is generally more prevalent. For nonlinear demixing [6,7], many difficulties occur and the linear ICA is no longer applicable because of the complexity of nonlinear parameters.

In this paper, we shall in-deep investigate a special but important instance of nonlinear mixtures, i.e., post-nonlinear (PNL) mixtures, and give out a novel algorithm.

2 Post-nonlinear Mixtures

An important special case of the general nonlinear mixing model that consists of so called post-nonlinear mixtures introduced by Taleb and Jutten [5], can be seen as a hybrid of a linear stage followed by a nonlinear stage.

* This work was supported by the National Science Foundation of China (Nos.60472111, 30570368 and 60405002).

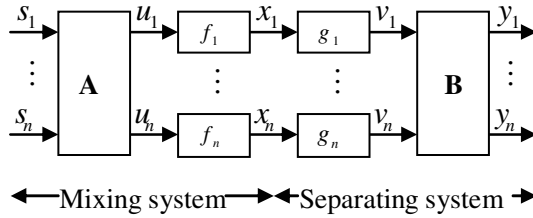


Fig. 1. The mixing –separating system for PNL

In the post-nonlinear mixtures model, the observations $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ have the following specific form (as shown in Fig. 1)

$$x_i = f_i \left(\sum_{j=1}^n a_{ij} s_j \right), i = 1, \dots, n \tag{1}$$

The corresponding vector-matrix form can be written as:

$$\mathbf{x} = f(\mathbf{A}\mathbf{s}) \tag{2}$$

Contrary to general nonlinear mixtures, the PNL mixtures have a favorable separability property. In fact, if the corresponding separating model for post-nonlinear mixtures, as shown in Figure 1, are written as:

$$y_i = \sum_{j=1}^n b_{ij} g_j(x_j) \tag{3}$$

Then it can be demonstrated that [5], under weak conditions on the mixing matrix \mathbf{A} and on the source distribution, the output independence can be obtained if and only if $\forall i = 1, \dots, n, h_i = g_i \circ f_i$ are linear. For more details, please refer to literatures [5].

3 Contrast Function

In this paper, we use Shannon’s mutual information as the measure of mutual dependence. It can be defined as:

$$I(\mathbf{y}) = \sum H(y_i) - H(\mathbf{y}) \tag{4}$$

where $H(y_i) = -\int p(y_i) \log p(y_i) dy_i$ denotes Shannon’s differential entropy.

According to the theory given above, the separating system of PNL we proposed in this paper is shown in Fig.2, where \mathbf{B} and g_i form the unmixing structure for PNL, y_i are the extracted independent components, and ψ_i some nonlinear mappings, which are used only for the optimization of the network.

Assume that each function $\psi_i(\boldsymbol{\varphi}_i, y_i)$ is the cumulative probability function (CPF) of the corresponding component y_i , then z_i are uniformly distributed in $[0, 1]$, Consequently, $H(z_i) = 0$ [7]. Moreover, because $\psi_i(\boldsymbol{\varphi}_i, y_i)$ are all continuous and monotonic increasing transformations (thus also invertible), then it can be easily shown that $I(\mathbf{z}) = I(\mathbf{y})$ [7]. Consequently, we can obtain

$$I(\mathbf{y}) = I(\mathbf{z}) = \sum_i H(z_i) - H(\mathbf{z}) = -H(\mathbf{z}) \tag{5}$$

Therefore, maximizing $H(\mathbf{z})$ is equivalent to minimizing $I(\mathbf{y})$.

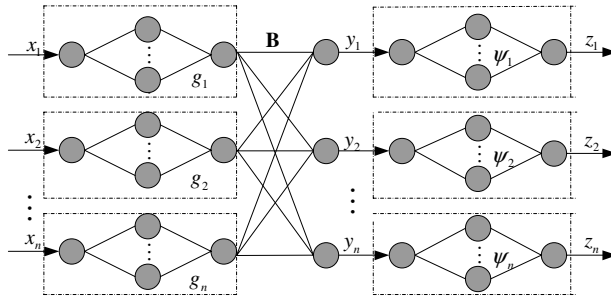


Fig. 2. The particular structure of the unmixing network

It has been proved in the literature [7] that, given the constraints placed on $\psi_i(\boldsymbol{\varphi}_i, y_i)$, then z_i is bounded to $[0, 1]$, and given that $\psi_i(\boldsymbol{\varphi}_i, y_i)$ is also constrained to be a continuous increasing function, then maximizing $H(\mathbf{z})$ will lead $\psi_i(\boldsymbol{\varphi}_i, y_i)$ to become the estimates of the CPFs of y_i . Consequently, y_i should be the duplicate of s_i with just sign and scale ambiguity. Now, the fundamental problem that we should to solve is to optimize the networks (formed by the g_i , \mathbf{B} and ψ_i blocks) by maximizing $H(\mathbf{z})$.

4 Unsupervised Learning of Separating System

With respect to the separation structure of this paper, the joint probabilistic density function (PDF) of the output vector \mathbf{z} can be calculated as:

$$p(\mathbf{z}) = \frac{p(\mathbf{x})}{|\det(\mathbf{B})| \prod_{i=1}^n |g'_i(\boldsymbol{\theta}_i, x_i)| \prod_{i=1}^n |\psi'_i(\boldsymbol{\varphi}_i, y_i)|} \tag{6}$$

which leads to the following expression of the joint entropy:

$$H(\mathbf{z}) = H(\mathbf{x}) + \log|\det(\mathbf{B})| + \sum_{i=1}^n E(\log|g'_i(\boldsymbol{\theta}_i, x_i)|) + \sum_{i=1}^n E(\log|\psi'_i(\boldsymbol{\varphi}_i, y_i)|) \tag{7}$$

The minimization of $I(\mathbf{y})$, which is equal to maximize $H(\mathbf{z})$ here, requires the computation of its gradient with respect to the separation structure parameters \mathbf{B} , $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$.

In this paper, we use RBF [3,4] network to model the nonlinear parametric functions $g_k(\boldsymbol{\theta}_k, x_k)$, and choose Gaussian kernel function as the activation function of the hidden neurons. In order to implement the constraints on the $\boldsymbol{\psi}$ function easy, we use multilayer perceptron to model the nonlinear parametric functions $\boldsymbol{\psi}_k(\boldsymbol{\varphi}_k, y_k)$.

5 Experiment Results

5.1 Extracting Sources From Mixtures of Simulant Signals

In the first experiment, the source signals consist of a sinusoid signal and a funny curve signal [1], i.e. $s(t) = [(\text{rem}(t,27)-13)/9, ((\text{rem}(t,23)-11)/9)^{0.5}]^T$, which are shown in Fig.3 (a). The two source signals are first linearly mixed with the (randomly chosen) mixture matrix:

$$\mathbf{A} = \begin{bmatrix} -0.1389 & 0.3810 \\ 0.4164 & -0.1221 \end{bmatrix} \tag{8}$$

Then, the two nonlinear distortion functions

$$f_1(u) = f_2(u) = \tanh(u) \tag{9}$$

are applied to each mixture for producing a PNL mixture. Fig.3 (b) shows the separated signals.

To compare the performance of our proposed method with other ones, we also use MISEP method [7] to conduct the related experiments based on the same data. The correlations between the two recovered signals separated by two methods and the two original sources are reported in Table.1. Clearly, the separated signals using the method proposed in this paper is more similar to the original signals than the other.

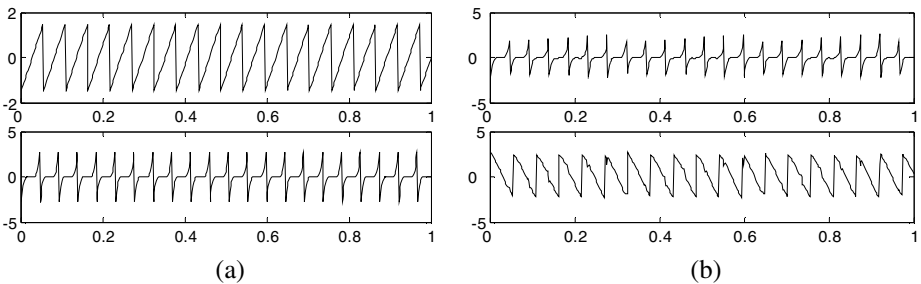


Fig. 3. The two set of signals shown. (a) Source signals. (b) Separated signals.

Table 1. Correlations between two original sources and the two recovered signals

Experiment	simulant signals		speech signals		
	y_1	y_2	y_1	y_2	
MISEP	S_1	0.9805	0.0275	0.9114	0.0280
	S_2	0.0130	0.9879	0.0829	0.9639
Method in this paper	S_1	0.9935	0.0184	0.9957	0.0713
	S_2	0.0083	0.9905	0.0712	0.9711

5.2 Extracting Sources from Mixtures of Speech Signals

To test the validity of the algorithm proposed in this paper ulteriorly, we also have experimentalized using real-life speech signals. In this experiment two speech signals (with 3000 samples, sampling rate 8kHz, obtained from <http://www.ece.mcmaster.ca/~reilly/kamran/id18.htm>) are post-nonlinearly mixed by:

$$A = \begin{bmatrix} -0.1412 & 0.4513 \\ 0.5864 & -0.2015 \end{bmatrix} \tag{10}$$

$$f_1(u) = \frac{1}{2}(u + \frac{1}{6}u^3), \quad f_2(u) = \frac{3}{5}u + \tanh(u) \tag{11}$$

The experimental results are shown in Fig.4 and Table.1, which conforms the conclusion drawn from the first experiment.

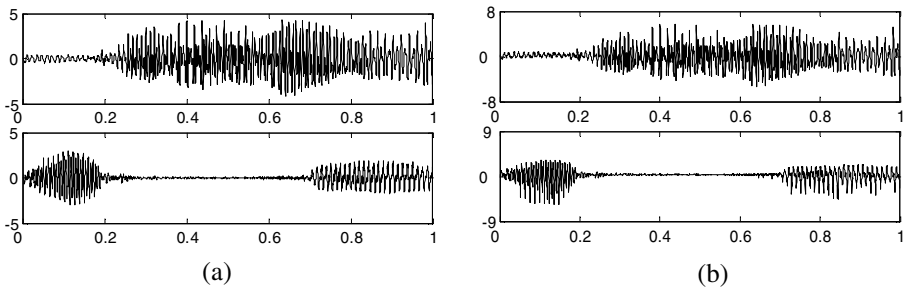


Fig. 4. The two set of speech signals shown. (a) Source signals. (b) Separated signals.

5.3 Training Speed

We also performed tests in which we compared, on the same post-nonlinear BSS problems, networks in which the \mathbf{g} blocks had MLP structures. Table 2 shows the means and standard deviations of epochs required to reach the stop criterion, which was based on the value of the objective function $H(\mathbf{z})$, for MLP-based networks and RBF-based networks.

Table 2. Comparison of training speeds between MLP-based and RBF-based networks

	Two superg		Superg. and subg.	
	RBF	MLP	RBF	MLP
Mean	315	508	369	618
St. dev	141	255	180	308

From the two tables we can see that the separating results of the two methods are very similar, but the RBF-based implementations trained faster and show a smaller oscillation of training times (One epoch took approximately the same time in both kinds of network). This mainly caused by the local character of RBF networks.

6 Conclusions

We proposed in this paper a novel algorithm for post-nonlinear blind source separation. This new method works by optimizing a network with a specialized architecture, using the output entropy as the objective function, which is equivalent to the mutual information criterion but needs not to calculate the marginal entropy of the output. Finally, the experimental results showed that this method is competitive to other existing ones.

References

1. Hyvärinen, A., Karhunen, J., Oja, E.: Independent Component Analysis. J. Wiley, New York (2001)
2. Hyvärinen, A., Pajunen, P.: Nonlinear Independent Component Analysis: Existence and Uniqueness Results. *Neural Networks*, 12(3) (1999) 429–439
3. Huang, D.S.: Systematic Theory of Neural Networks for Pattern Recognition. Publishing House of Electronic Industry of China, Beijing (1996)
4. Huang, D.S.: The United Adaptive Learning Algorithm for the Link Weights and the Shape Parameters in RBFN for Pattern Recognition. *International Journal of Pattern Recognition and Artificial Intelligence*.11(6) (1997) 873-888
5. Taleb, A., Jutten, C.: Source Separation in Post- nonlinear Mixtures. *IEEE Trans. Signal Processing*, 47 (1999) 2807–2820
6. Martinez, Bray, D. A.: Nonlinear Blind Source Separation Using Kernels. *IEEE Trans. Neural Networks*, 14(1) (2003) 228–235
7. Almeida, L. B.: MISEP – Linear and Nonlinear ICA Based on Mutual Information. *Journal of Machine Learning Research*.4(2) (2003) 1297-1318