Towards Blind Separation of Post-Nonlinearly Mixed Sources

By Wai Yie Leong and Danilo P. Mandic

Communications and Signal Processing Group
Department of Electrical and Electronic Engineering
Imperial College London, SW7 2AZ, UK.
w.leong@imperial.ac.uk, d.mandic@imperial.ac.uk

Abstract

A novel approach which extends blind source separation (BSS) of one or group of sources to the case of post-nonlinear mixtures is proposed. This is achieved by an adaptive algorithm in which the cost function jointly estimates the kurtosis and a measure of nonlinearity. Simulation results are presented which illustrate the validity of the proposed approach.

Blind signal separation (BSS) [4], [1] aims at recovering unobservable signals (sources) from their linear or nonlinear mixtures. This technique has recently attracted much interest due to its potentially wide number of applications in diverse fields. Despite recent progress in BSS, the main limitation of existing BSS algorithms is that they have been specifically designed for linear mixtures. To help mitigate these limitations, we set out to extend some existing criteria for blind source separation in order to make them suitable for simultaneous post-nonlinear separation of arbitrary groups of *m* ($1 \leq m \leq n$, where *n* denotes the total number of sources) sources of interest.

The need for modelling of a post-nonlinear system [2], [3], [7], [6] arises in some real world situations, for instance: (1) sensors normally posses nonlinear transfer characteristics; (2) the effects of reflections and interfering signals may introduce nonlinearity into the system.

I. Post-Nonlinear Mixtures

Consider *n* unknown sources $\mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T$ with zero mean. Sources are observed through a nonlinear vector mapping $\mathbf{M}(\cdot)$ and linear mixing matrix $\mathbf{A}$, to give measurements $\mathbf{x}(k)$. This nonlinear mixing problem (from the unknown sources $\mathbf{s}(k)$ to the observation $\mathbf{x}(k)$ can be modelled as a post-nonlinear system. We therefore assume the signals $\mathbf{x}(k)$ are nonlinear memoryless mixtures of n unknown statistically independent sources $\mathbf{s}(k)$, and the observation process can be expressed as

$$\mathbf{x}(k) = \mathbf{M}(\mathbf{A}\mathbf{s}(k)) \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ is an unknown mixing matrix which is assumed to be non-singular.

Our goal is to separate the sources of interest without any prior knowledge of their distributions and the nonlinear mixing mechanism. To that cause, we need to derive a separation structure which involves learning rule for the estimation of the unmixing matrix $\mathbf{W}$, and a way to estimate the nonlinearity within. This unmixing operation can be expressed as

$$\mathbf{y}(k) = \mathbf{W}(\mathbf{x}(k)) \tag{2}$$

where $\mathbf{y}(k)$ denotes the separated output signals.

In order to extract $m \leq n$ sources, the observations (1) will be processed by an $(m \times n)$ separating matrix $\mathbf{W}$, satisfying $\mathbf{W}\mathbf{W}^T = \mathbf{I}$, which yields the output vector (or estimated sources) $\mathbf{y}(k)$. The matrix $\mathbf{g} = \mathbf{W}\mathbf{A}$ will denote an $m \times n$ global matrix from the sources to the outputs.

II. The proposed Separation Algorithm

A. Nonlinear Separation Algortihm

For the separation of post-nonlinear mixtures, we propose the following ”mixed norm” criterion:

$$\boldsymbol{J}(\mathbf{y}(k)) = \sum_{i=1}^{n} |cum[y_i^4(k)]| - E\{log \sum_{i=1}^{n} [f_i(y_i)(k)|\} \tag{3}$$

where $f_i(\cdot)$ is the nonlinearity. The left hand side part of (3) is responsible for standard BSS, whereas the right hand part of (3) estimates the nonlinearity within the mixing process. It is important to note that (3) holds only if the functions $f_i(\cdot)$ are invertible, a restriction that must be taken into account in the development of learning algorithms.

B. Nonlinearity: The Activation Function

Notice that the first term in (3) represents the standard demixing criterion, whereas the second term in (3) is related to the estimation of the nonlinearity effects within the mixtures. This way, criterion (3) represents a joint constrained optimisation problem.

In order to derive a learning algorithm corresponding to (3), let us consider separately the minimisation either parts of cost function (3). Let J_K correspond to the first term in (3) (kurtosis) and J_N to the second term (nonlinearity). Observe that

$$\begin{aligned} J_N(\mathbf{W}(k), \mathbf{y}(k)) &= \frac{\partial \sum_{i=1}^{n} log f_i(y_i(k))}{\partial \mathbf{W}(k)} \\ &= \frac{\partial \sum_{i=1}^{n} log f_i(y_i(k))}{\partial \mathbf{y}(k)} \frac{\partial \mathbf{y}(k)}{\partial \mathbf{W}(k)} \end{aligned} \tag{4}$$

where $\mathbf{f}(\mathbf{y}(k)) = [f_1(y_1(k)), f_2(y_2(k)), \ldots, f_n(y_n(k))]^T$ is the column vector whose *i*th component is

$$\begin{aligned} f_i(y_i(k)) &= -\frac{\partial log q_i(y_i(k))}{\partial y_i(k)} \\ &= -\frac{\partial q_i(y_i(k))/\partial y_i(k)}{q_i(y_i(k))} \\ &= -\frac{q_i'(y_i(k))}{q_i(y_i(k))} \end{aligned} \tag{5}$$

where $q_i(y_i(k))$, $i = 1, \ldots, n$, are true probability density functions of the source signals. In fact, minimising the above cost function leads to the minimisation of the mutual information [5]. The differential (infinitesimal increment) of the cost function can be evaluated as

$$dJ_N = \mathbf{f}(\mathbf{y}(k)) d\mathbf{y}(k) \tag{6}$$

where k denotes the iteration number. Notice also that

$$d\mathbf{y}(k) = -d\mathbf{x}(k)\mathbf{y}(k) \tag{7}$$

Applying the natural gradient concept by introducing a new differential

$$d\mathbf{x}(k) = [\mathbf{I} + \mathbf{W}(k)]^{-1} d\mathbf{W}(k) \tag{8}$$

Assuming that small changes of the output signals are only affected by a small variation of the ”synaptic” weights $\{w_{ij}(k)\} \in \mathbf{W}(k)$, we can approximate (7) by

$$d\mathbf{y}(k) = -[\mathbf{I} + \mathbf{W}(k)]^{-1} d\mathbf{W}(k)\mathbf{y}(k) \tag{9}$$

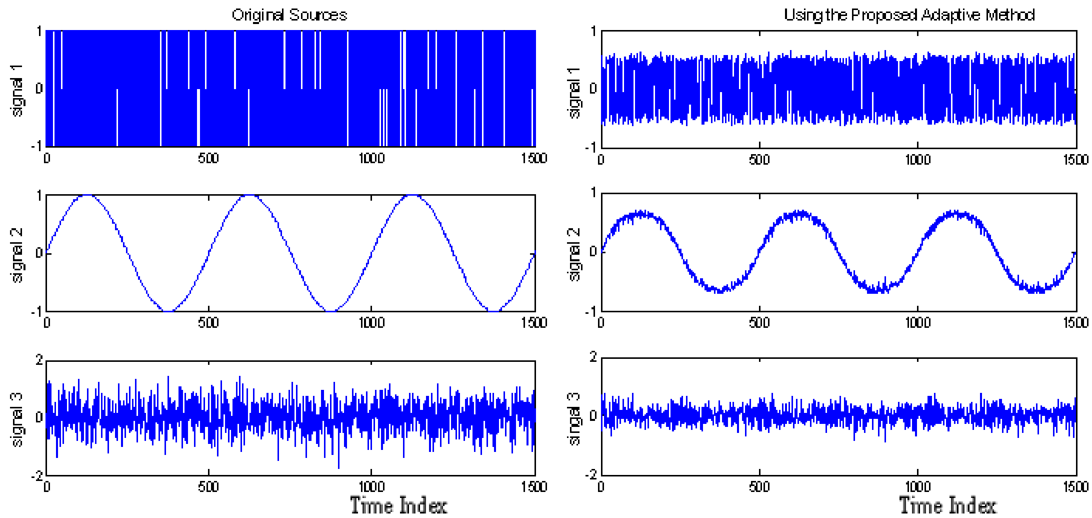

Fig. 1. Source signals used in simulations: (a) The original source signals, s_1 with binary distribution, s_2 a sine wave and s_3 with Gaussian distribution; (b) The separated output signals based on the proposed nonlinear predictor, y_1 with binary distribution, y_2 a sine waveform and y_3 with Gaussian distribution.

Hence, on the basis of the standard gradient descent, we obtain an approximate learning rule, given by

$$\begin{aligned} \triangle\mathbf{W}(k) &= -\eta_0(k)\frac{\partial J_N}{\partial\mathbf{W}(k)} \\ &= \eta_0(k)[\mathbf{I}+\mathbf{W}(k)]\mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k) \end{aligned} \tag{10}$$

which finally yields a sequential update in the form of

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta_0(k)[\mathbf{I}+\mathbf{W}(k)]\mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k) \tag{11}$$

Applying standard gradient descent to (3), an update which minimises the ”mixed norm” cost function (3) is obtained as

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta_0(k)\mathbf{f}(\mathbf{y}(k))[\mathbf{x}(k) - (\mathbf{I}+\mathbf{W}(k))\mathbf{y}^T(k)] \tag{12}$$

where the separated outputs, $\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k)$.

III. Experimental Results

In the experiments, the simulations were based on three source signals: s_1 with binary distribution, s_2 a sine waveform and s_3 with Gaussian distribution, as shown in Fig.1. The input for all signals was scaled to range [-2, 2], with positive kurtosis, and learning rates $\eta_0(k)$. Monte Carlo simulations with 1500 iterations of independent trials were performed. The initial values of the predictor weights and the demixing matrix vector $\mathbf{W}(k)$ were randomly generated for each run. The simulations were conducted without prewhitening. In theory, by minimisation of the normalised kurtosis of the extracted signal, we will recover the first source, since it has the smallest kurtosis value (binary signal). A 3×3 mixing matrix was randomly generated, and is given by

$$\mathbf{A} = \begin{vmatrix} -0.8623 & -0.5502 & -0.0542 \\ 0.1812 & 0.3532 & -0.2561 \\ -0.5511 & -0.4358 & 0.9759 \end{vmatrix} \tag{13}$$

The proposed adaptive algorithm which does not require any prepocessing (prewhitening) is particularly suitable for blind source separation with post-nonlinear mixing matrices. The scatter plot (Figure 2) has

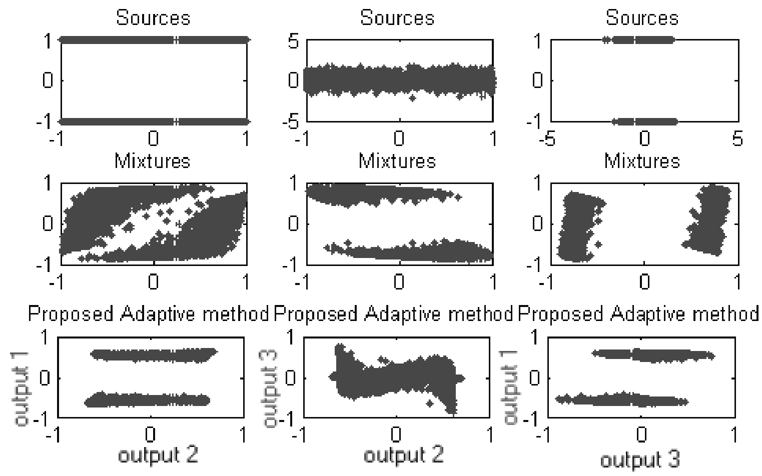

Fig. 2. Scatter plot comparing the independence of the output signals; Column 1: signal 1 and 2; Column 2: signal 2 and 3; Column 3: signal 1 and 3.

confirmed the validity of the theoretical results and demonstrated the performance of the algorithm. The output signals are closely matched with the original unknown sources (Fig.1).

IV. Conclusions

We have proposed an approach for post-nonlinear blind source separation. The proposed method and conditions have been discussed. A neural-network model and its associated adaptive learning rule for post-nonlinear mixtures have been introduced. The proposed adaptive algorithm does not require any prepocessing (prewhitening), and due to the design of the contrast function, it is particularly suitable for blind source separation with post-nonlinear mixing matrices. Simulation results have confirmed the validity of the theoretical results and demonstrated the performance of the algorithm.

References

[1] A.Hyvarinen and E.Oja. Independent component analysis: A tutorial. Technical report, Helsinki University of Technology, April 1999.

[2] A.Taleb and C.Jutten. Source separation in post-nonlinear mixtures. *IEEE transactions on Signal Processing*, 47(10):2807–2820, 1999.

[3] H.Valpola, X.Giannakopoulos, A.Honkela, and J.Karhunen. Nonlinear independent component analysis using ensemble learning: Experiments and discussion. Technical report, Helsinki University of Technology, Neural Networks Research Centre, 2000.

[4] P.Comon. Independent component analysis, a new concept? *Special Issue on Higher-Order Statistics, Signal Processing*, 36(3):287–314, April 1994.

[5] S.Amari. Natural gradient works efficiently in learning. In *Neural Computation*, volume 10, pages 251–276, Jan 1998.

[6] W.Y.Leong and D.P.Mandic. Adaptive blind extraction for post-nonlinearly mixed signals. In *2006 IEEE International Workshop on Machine Learning for Signal Processing Formerly the IEEE Workshop on Neural Networks for Signal Processing, Maynooth, Ireland*, pages 91–96, September 6-8, 2006.

[7] W.Y.Leong and J.Homer. EKENS: A learning on nonlinear blindly mixed signals. In *The 30th International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*, volume 4, pages 81–84, Pennsylvania Convention Center, US, March 19-23, 2005.