Neural-based Separating Method for Nonlinear Mixtures

Ying Tan

National Laboratory on Machine Perception,
Department of Intelligence Science,
Peking University, Beijing, 100871, China
ytan@pku.edu.cn

Abstract. A neural-based method for source separation in nonlinear mixture is proposed in this paper. A cost function, which consists of the mutual information and partial moments of the outputs of the separation system, is defined to extract the independent signals from their nonlinear mixtures. A learning algorithm for the parametric RBF network is established by using the stochastic gradient descent method. This approach is characterized by high learning convergence rate of weights, modular structure, as well as feasible hardware implementation. Simulation results demonstrated the success of our proposed method in this paper.

Key words: Blind signal separation, nonlinear mixture, RBF networks, statistical independence, cost function.

1 Introduction

Blind source separation in signal processing has received considerable attentions in the last decade[1, 2]. Many blind separation algorithms have been proposed based on different separation models. These algorithms play increasingly important roles in many applications.

Usually, Blind source separation is to recover unobservable independent sources (or "signals") from several observed data masked by linear or nonlinear mixing. Most existing algorithms for linear mixing models stem from the theory of the independent component analysis (ICA) [3]. Therefore, a solution to blind source separation problem exists and this solution is unique up to some trivial indeterminacies (permutation and scaling) according to the basic ICA theory [3]. Even though the nonlinear mixing model is more realistic and practical, most existing blind separation algorithms developed so far are valid for linear models. For nonlinear mixing models, many difficulties occur and both the linear ICA theory and existing linear demixing algorithms are no longer applicable because of the complexity of nonlinear characteristics [4].

Several authors studied the difficult problem of the nonlinear blind source separation and proposed a few efficient demixing algorithms [4–9]. Deco [5] studied a very particular scenario of volume-conserving nonlinear transforms. Pajunen et al.[6] proposed model-free methods which used Kohonen's self-organizing map (SOM) to extract independent sources from nonlinear mixture, but suffers from the exponential growth of network complexity and interpolation error in recovering continuous sources. Burel [7] proposed a nonlinear blind source separation algorithm using two-layer perceptrons by the gradient descent method to minimize the mutual information (measure of dependence). Subsequently, Yang et al. [8] developed an information backpropagation (BP) algorithm for Burel's model by natural gradient method. In their model cross nonlinearities is included. Taleb et al. [9] proposed an entropy-based direct algorithm for blind source separation in post nonlinear mixtures. Recently, authors[12, 13] proposed several algorithms and approaches for separation of nonlinear mixture of sources.

2 Nonlinear mixture model

A generic nonlinear mixture model for blind source separation can be described as $\mathbf{x}(t) = \mathbf{f}[\mathbf{s}(t)]$, where $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_n(t)]^T$ is the vector of observed random variables, superscript T denotes the transposition, $\mathbf{s}(t) = [s_1(t), s_2(t), ..., s_n(t)]^T$ is the vector of the latent variables called the independent source vector, $\mathbf{f}$ is an unknown multiple-input and multiple-output (MIMO) mapping from R^n to R^n called nonlinear mixing transform (NMT). If the mixing function $\mathbf{f}$ is linear, this model reduces to the linear mixing. In order for the mapping to be invertible we assume that the nonlinear mapping $\mathbf{f}$ is monotone.

The separating system $\mathbf{g}(\cdot, \theta)$, is called nonlinear separation transform (NST), is used to recover the original signals from the nonlinear mixture $\mathbf{x}(t)$ without the knowledge of the source signals $\mathbf{s}(t)$ and the mixing nonlinear function $\mathbf{f}(\cdot)$. Obviously, this problem is untractable, in particular for nonlinear mixing system, unless conditions are imposed on the nonlinear function $\mathbf{f}(\cdot)$. At first, the existence of the solution for the NST can be guaranteed. According to related nonlinear ICA theories, the nonlinear ICA problem always has at least one solution. That is, given a random vector $\mathbf{x}$, there is always a function $\mathbf{g}$ so that the components of $\mathbf{y} = [y_1, \cdots, y_n]^T$ given by $\mathbf{y} = \mathbf{g}(\mathbf{x})$ are independent[6, 4].

Unfortunately, this kind of mapping is not at all unique. It is shown in [4] that a unique solution subjected to a rotation can be obtained under the assumptions that the problem is a two-dimensional one, mixing function is a conformal mapping, and the densities of the independent components are known and have bounded support. In addition, we add some constraints on the output; i.e., the moment matching between the outputs of the separating system and sources. Accordingly, the output of the nonlinear separating system can be written as

$$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t), \boldsymbol{\theta}) = \mathbf{g}(\mathbf{f}(\mathbf{s}(t)), \boldsymbol{\theta}) = \mathbf{s}(t) \tag{1}$$

where $g(\cdot, \boldsymbol{\theta}) = f^{-1}(\cdot)$ denotes a parametric fitting function class, $\boldsymbol{\theta}$ is a parameter vector to be determined.

3 Nonlinear separation based on an RBF network

3.1 the RBF neural network

An n-input and n-ouput RBF network model consists of three layers; i.e., input layer, hidden layer and output layer. The neurons in hidden layer are of local response to its input and called RBF neurons while the neurons of the output layer only sum their inputs and are called linear neurons. The RBF network is often used to approximate an unknown continuous function $\phi : R^n \to R^n$ which can be described by the affine mapping

$$\mathbf{u}(\mathbf{x}) = \mathbf{B}K(\mathbf{x}, \mathbf{p}) \tag{2}$$

$$K(\mathbf{x}, \mathbf{p}) = [1, \exp(-(\mathbf{x}-\boldsymbol{\mu}_1)^T(\mathbf{x}-\boldsymbol{\mu}_1)/\sigma_1^2), \cdots, \exp(-(\mathbf{x}-\boldsymbol{\mu}_M)^T(\mathbf{x}-\boldsymbol{\mu}_M)/\sigma_M^2)]^T. \tag{3}$$

where $\mathbf{B} = [\alpha_{ij}]$ is a $n \times M$ weight matrix of the output layer, $K(\mathbf{x}, \mathbf{p})$ is Gaussian kernel function vector of the RBF network, which consists of the locally receptive functions. $\mathbf{p} = (\boldsymbol{\mu}_1, \sigma_1, \cdots, \boldsymbol{\mu}_M, \sigma_M)^T$ is the parameter set of the kernel function. Here we let the first component of $K(\mathbf{x}, \mathbf{p})$ be 1 for taking the bias into account.

3.2 Nonlinear separation system based on RBF network

Since the local response power of RBF networks offers great classification and approximation capabilities, the Gaussian RBF network is used as a good function approximator in many modelling applications. If we let $\mathbf{S}$ be a compact subset in R^n and $\mathbf{p}(\mathbf{x})$ be a continuous target vector on $\mathbf{S}$, then for any $\epsilon > 0$ there exist M centroids $\boldsymbol{\mu}_i = [\mu_{i1}, \cdots, \mu_{in}]^T$ and an $n \times M$ constant matrix $\mathbf{B}$ such that $\mathbf{r}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{B} \cdot K(\mathbf{x}, \mathbf{p})$ satisfies $|\mathbf{r}(\mathbf{x}, \boldsymbol{\theta}) - \mathbf{p}(\mathbf{x})| < \epsilon$ for all $\mathbf{x} \in \mathbf{S}$. This approximation ability of RBF networks directly stems from the classic Stone-Weierstrass theorem and is closely related to Parzen's approximation theory. Therefore, the inverse of the nonlinear mixing model can be modeled by using an RBF network. Such architecture is preferred over multilayer perceptrons (MLP) as an RBF network has better capability for functional representation. Since its response is linearly related to its weights, learning in an RBF network is expected to train faster while its local response power offers a good approximation capability. As a result, we can reach

$$\mathbf{y} = \hat{\mathbf{B}}K[\mathbf{f}(\mathbf{s}), \hat{\mathbf{p}}] \propto \mathbf{s} \tag{4}$$

where $\mathbf{g}(\cdot, \hat{\boldsymbol{\theta}}) = \hat{\mathbf{B}}K[., \hat{\mathbf{p}}]$, $\hat{\mathbf{B}}$ and $\hat{\mathbf{p}}$ are the final estimates of parameters $\mathbf{B}$ and $\mathbf{p}$ of the RBF network such that the inverse of $\mathbf{f}$ is well approximated by the RBF network.

3.3 Cost function

In order to deal with the nonlinear separation problem effectively, we define a cost function, or contrast function, which is the objective function for signal separation, as

$$C(\boldsymbol{\theta}) = I(\mathbf{y}) + \sum_{i_1\cdots i_n} c_{i_1\cdots i_n}[M_{i_1\cdots i_n}(\mathbf{y},\boldsymbol{\theta}) - M_{i_1\cdots i_n}(\mathbf{s})]^2 \tag{5}$$

where $I(\mathbf{y})$ is mutual information of the outputs of the separation system, $M_{i_1\cdots i_n}(\mathbf{y},\boldsymbol{\theta})$ and $M_{i_1\cdots i_n}(\mathbf{s})$ are the $i_1\cdots i_n$-th moments of $\mathbf{y}$ and $\mathbf{s}$, respectively, $c_{i_1\cdots i_n}$ are constants which are used to balance the mutual information and the matching of moments.

According to information theory and related the Kullback-Leibler divergence, mutual information $I(\mathbf{y})$ in Eq. (5) is expressed as

$$I(\mathbf{y}) = \sum_{i=1}^{n} H(y_i) - H(\mathbf{y}) \tag{6}$$

where $H(\mathbf{y}) = -E[\log(p_{\mathbf{y}}(\mathbf{y})]$ is the joint entropy of random vector $\mathbf{y}$, $H(y_i) = -E[\log(p_{y_i}(y_i)]$ is the entropy of random variable y_i, the ith component of $\mathbf{y}$, and $E(\cdot)$ denotes the expectation operator.

The $i_1\cdots i_n$th moment of $\mathbf{y}$ is defined as

$$M_{i_1\cdots i_n}(\mathbf{y}) = E(y_1^{i_1}\cdots y_n^{i_n}) - E(y_1^{i_1})\cdots E(y_n^{i_n}). \tag{7}$$

It can be seen from Eqs. (5)- (7) that the constrast function defined in Eq. (5) is always non-negative, and reaches zero if and only if both mutual information is null and a perfect matching of moments between the outputs of the separation system and original sources is achieved. Therefore, independent outputs with the same moments as that of original sources can be found by minimizing the contrast function by adjusting the parameters of the RBFN separating system, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}}\{I(\mathbf{y}) + \sum_{i_1\cdots i_n} c_{i_1\cdots i_n}[M_{i_1\cdots i_n}(\mathbf{y},\boldsymbol{\theta}) - M_{i_1\cdots i_n}(\mathbf{s})]^2\}. \tag{8}$$

3.4 Learning algorithm of the separating RBF network

In order to derive the unsupervised learning algorithm of all the parameters of the separating RBF network, we employ the gradient descent method. First of all, we compute the gradient of the contrast function of Eq. (5) with respect to the parameter $\boldsymbol{\theta}$ and obtain

$$\frac{\partial C(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial I(\mathbf{y})}{\partial \boldsymbol{\theta}} + \sum_{i_1\cdots i_n} 2c_{i_1\cdots i_n}[M_{i_1\cdots i_n}(\mathbf{y},\boldsymbol{\theta}) - M_{i_1\cdots i_n}(\mathbf{s})]\frac{\partial M_{i_1\cdots i_n}(\mathbf{y},\boldsymbol{\theta})}{\partial \mathbf{y}}\frac{\partial \mathbf{y}}{\partial \boldsymbol{\theta}} \tag{9}$$

where mutual information can be further rewritten as

$$I(\mathbf{y}) = \sum_{i=1}^{n} H(y_i) - E\{\log|\frac{\partial \mathbf{g}(\mathbf{x}, \boldsymbol{\theta})}{\partial \mathbf{x}}|\} - H(\mathbf{x}) \tag{10}$$

where $|\partial \mathbf{g}(\mathbf{x}, \boldsymbol{\theta})/\partial \mathbf{x}|$ is the determinant of the Jacobian matrix of $\mathbf{g}(\mathbf{x}, \boldsymbol{\theta})$ with respect to vector $\mathbf{x}$.

Regarding different concrete parameters $\mathbf{B}$, $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ of the parameter set $\boldsymbol{\theta}$ of the RBF network, we have the following gradient equations of the separated signal $\mathbf{y}$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{B}} = K(\mathbf{x}, \mathbf{t}), \tag{11}$$

$$\frac{\partial \mathbf{y}}{\partial \boldsymbol{\mu}} = \mathbf{B} \cdot diag[\mathbf{v_1} \circ K(\mathbf{x}, \mathbf{t})], \tag{12}$$

$$\frac{\partial \mathbf{y}}{\partial \sigma} = \mathbf{B} \cdot diag[\mathbf{v_2} \circ K(\mathbf{x}, \mathbf{t})]. \tag{13}$$

where $\mathbf{v_1} = [2(\mathbf{x}-\boldsymbol{\mu}_1)/\sigma_1^2, \cdots, 2(\mathbf{x}-\boldsymbol{\mu}_M)/\sigma_M^2]^T$, $\mathbf{v_2} = [2\|\mathbf{x}-\boldsymbol{\mu}_1\|^2/\sigma_1^3, \cdots, 2\|\mathbf{x}-\boldsymbol{\mu}_M\|^2/\sigma_M^3]^T$, function $diag[\cdot]$ denotes diagonal matrix, symbol $\circ$ denotes Hadamard product which is the multiplication of corresponding pairs of elements between two vectors.

Finally, from Eqs. (9), (11)- (13), we can easily calculate the gradients of the constrast function with respect to each parameter of parameter set $\boldsymbol{\theta}$.

3.5 Performance Index and Algorithm Description

From Eq. (10), by omitting the unknown $H(\mathbf{x})$, an index to measure the independence of the outputs of the separation system is defined as

$$J_i = \sum_{i=1}^{n} H(y_i) - E\{\log|\frac{\partial \mathbf{g}(\mathbf{x}, \boldsymbol{\theta})}{\partial \mathbf{x}}|\} \tag{14}$$

Even though the index J_i may be negative, the lower the value of J_i is, the more independent the outputs of the separating system is. The smallest negative value of J_i is just equal to the reciprocal of $H(\mathbf{x})$. In a similar manner, according to Eq. (10), a performance index measuring moment match up to the k-th order between the outputs of the separation system and original sources can also be directly defined as

$$J_m^k = \sum_{i_1 \cdots i_n \le k} [M_{i_1 \cdots i_n}(\mathbf{y}, \boldsymbol{\theta}) - M_{i_1 \cdots i_n}(\mathbf{s})]^2 \tag{15}$$

The maximum value of k is chosen such that the inverse of the mixing nonlinear transform can be uniquely approximated by an RBF network through the minimization of the cost function. In actual implementation, usually only up to forth-order moment is enough for this purpose by experiments. We expect both J_i and J_m^k are at their minimia simultaneously, so the two indices can be combined into one overall index as follows

$$J = J_i + \alpha J_m^k \tag{16}$$

where α is a proportionality constant weighting the two quantities.

4 Simulation Results

Consider a two-channel nonlinear mixture with a cubic nonlinearity:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{A_2} \begin{bmatrix} (\cdot)^3 \\ (\cdot)^3 \end{bmatrix} \mathbf{A_1} \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} \tag{17}$$

where mixing matrices $\mathbf{A_1}$ and $\mathbf{A_2}$ are nonsingular.

The source vector $\mathbf{s}(t)$ consists of a sinusoidal signal and an amplitude-modulated signal; i.e., $\mathbf{s}(t) = [0.5 * [1 + \sin(6\pi t)] \cos(100\pi t), \sin(20\pi t)]^T$.

In this experiment we use an RBF network with six hidden neurons with Gaussian kernel.

An example of the evolution curves for the learning algorithm is shown in Fig. 1. The learning curve is smooth and it converges after 500 iterations.

The value of the performance index after convergence of the learning algorithm is very small so that the separated signals obtained by our model are seen to be mutually independent.

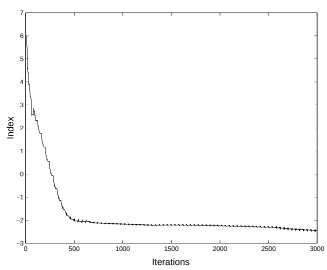

Fig. 1. Learning curve of the proposed algorithm

Fig. 2 shows the two source signals $\mathbf{s}(t)$ and the input signals $\mathbf{x}(t)$ of the separating system , i.e., the mixture of the sources. Figs. 3 show the signals separated by the proposed approach.

5 Concluding Remarks

A neural-based separating approach is established to separate nonlinearly mixed sources in terms of a novel cost function which consists of mutual information

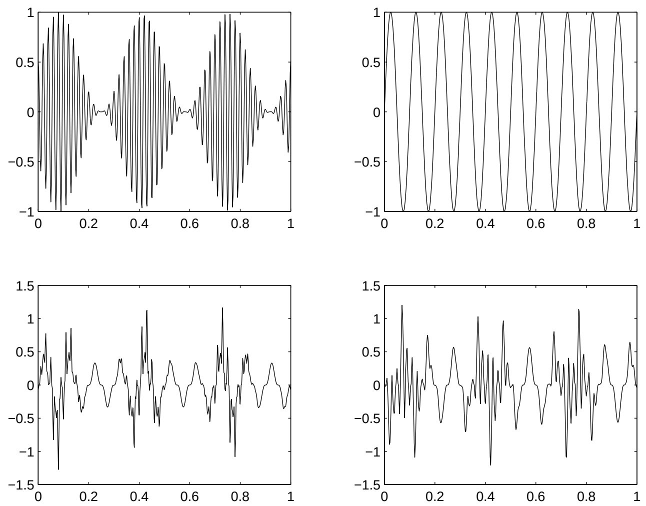

Fig. 2. Two source signals (above) and their nonlinear mixtures (below)

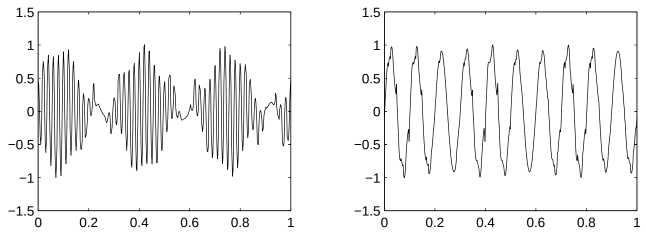

Fig. 3. Separated signals of our proposed method

and cumulants' matching. Because of the local response of RBF networks, the proposed method is of fast learning convergence rate of weights, natural unsupervised learning characteristics, modular network structure as well as suitable hardware implementation. All of these properties make it be an effective candidate for real-time multi-channel separation of nonlinear mixtures of sources. Extensive simulation results verified the validation of our methods.

Acknowledgments. Project 60673020 supported by National Natural Science Foundation of China.

References

1. Papadias, C.B., Paulraj, A.: A constant modulus algorithm for multi-user signal separation in presence of delay spread using antenna arrays. IEEE Signal Processing Letters, vol. 4, (1997) 178-181.
2. Herault, J., Jutten, C.: Space or time adaptive signal processing by neural network models. Proc. AIP Conf., Snowbird, UT, 1986, in Neural Networks for Computing, J. S. Denker, Ed. New York: Amer. Inst. Phys., (1986) 206-211.
3. Jutten, C., Herault, J.: Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture. Signal Processing, 24, (1991) 1-20.
4. Hyvarinen, A., Pajunen, P.: Nonlinear independent component analysis: Existence and uniqueness results. Neural Networks, 12, (1999) 429-439.
5. Deco, G., Brauer, W.: Nonlinear higher-order statistical decorrelation by volume-conserving neural architectures. Neural Networks, 8, (1995) 525-535.
6. Pajunen, P., Hyvarinen, A., Karhunen, J.: Nonliner blind source separation by self-organizing maps. Progress in Neural Information Processing: Proceedings of ICONIP'96, 2, Springer, (1996) 1207-1210.
7. Burel, G.: Blind separation of sources: a nonlinear neural algorithm. Neural Networks, 5, (1992) 937-947.
8. Yang, H. H., Amari, S., Cichocki, A.: Information back-propagation for blind separation of sources from non-linear mixture. Proc. IEEE IJCNN, Houston, USA, (1997) 2141-2146.
9. Taleb, A., Jutten, C., Olympieff, S.: Source separation in post nonlinear mixtures: an entropy-based algorithm. Proc. ESANN'98, (1998) 2089-2092.
10. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separtion. Advances in Neural Information Processing Systems, 8, eds. David S. Touretzky, Michael C. Mozer and Michael E. Hasselmo, MIT Press: Cambridge, MA, (1996) 757-763.
11. Cichocki, A., Unbehauen, R.: Robust neural networks with on-line learning for blind identification and blind separation of sources. IEEE Trans. Circuits and Systems-I, 43, (1996) 894-906.
12. Tan, Y., Wang, J., Zurada, J.M.: Nonlinear Blind Source Separation Using Radial Basis Function Networks. IEEE Transaction on Neural Networks, 12(1), (2001) 124-134.
13. Tan, Y., Wang, J.: Nonlinear Blind Separation Using Higher-Order Statistics and A Genetic Algorithm. IEEE Transaction on Evolutionary Computation, 5(6), (2001) 600-612.