Joint Diagonalization Learning Algorithm for Nonlinear Blind Source Separation

Nuo Zhang, Xiaowei Zhang, Jianming Lu and Takashi Yahagi
Graduate School of Science and Technology
Chiba University, Chiba-shi, 263-8522 Japan
Email: zhang@graduate.chiba-u.jp

Abstract— **Recovering independent source signals from their nonlinear mixtures is a very important issue in many fields. Joint approximate diagonalization of eigenmatrices (JADE) is an efficient method which utilizes fourth-order cumulants of signals. However it cannot deal with nonlinear problem. This paper proposes a robust radial basis function network (RBFN) approach by using higher-order cumulants when observations are suffered from noise and nonlinear distortion. The higher-order cumulants can measure the departure of a random vector from a Gaussian random vector for extracting the non-Gaussian part of a signal. The proposed method can efficiently recover the nonlinearly mixed signals suffered high-level noise simultaneously. The proposed method is divided into two steps. First, the radial basis function helps us to transform the mixed signals to high-dimensional space. Then in the second step, we can linearly separate the mixtures in the high-dimensional space by jointing approximate diagonalization of eigenmatrics. We consider artificial signal and acoustic signal separation and denoising applications. Furthermore, a comparison between the traditional RBF-based method, original JADE and proposed algorithm is produced, from which we can see the proposed algorithm is more suitable and applicable for unsupervised nonlinear signal denoising problem.**

I. Introduction

Estimating the unknown transmission system or recovering unaccessible independent signal from their mixtures are interesting problems which can be encountered in a lot of fields of signal processing. This kind of problems are called blind source separation for source signal are unknown. Blind source separation has received increasing interest and has become an active research area in both statistical signal processing and unsupervised neural learning [1], [2], recently. Many algorithms have been developed for the BSS problem in the literature of telecommunications, medical signal processing, feature extraction, speech separation, time-series analysis and blind deconvolution.

In the last few years, linear problem of blind source separation has become an important research area, with a theoretical basis and a set of methodologies that are becoming progressively more comprehensive. Hence, the main study about blind source separation considers about the instantaneous linear mixture so far [3]. Rather complete coverages of the subject can be found in [4]. However, in general, a nonlinear mixing model is more realistic and accurate than linear model and suitable for many practical situations. Some methods have been developed in traditional independent component analysis manner by researchers [2], in which the difficulty and importance have been shown. Therefore, it is significant to consider blind source separation problem in nonlinear manner in the presence of noise in depth.

Nonlinear blind source separation of signals has been very sparsely studied because of its intrinsic indeterminacy and the unknown distribution of sources as well as the mixing conditions [5]. Several algorithms stem from ICA for nonlinear mixtures are proposed [6]. Note that, Neural networks are valuable tools to deal with a variety of nonlinear problems occurred in many practical domains and have been studied extensively. Examples can be found in the following literature: self-organizing map (SOM), utilized to extract sources from nonlinear mixture (but the computational complexity grows exponentially) [7], information BP algorithm for training of separating system [8], and entropy-based direct algorithm [9]. One of the drawbacks of traditional methods is slow convergence rate because of the highly nonlinear relationship between the output and learning weights of the network.

In contrast, RBF network has several advantages in terms of natural learning manner. An unsupervised approach using RBF network proposed by Cha *et al.* [1] can recover source signals better with fast convergence rate. For applying the advantage of RBF, we can employ a linear blind source separation algorithm to extend RBF then derive a novel nonlinear blind source separation method.

This paper aims to show that an extended RBF-based method can achieve the general nonlinear blind source separation goal and has the ability to reduce noise in signal. Moreover, higher-order cumulants is proofed efficient to reduce noise in a lot of literature [10]. We employ Joint approximate diagonalization of eigenmatrices (JADE),which utilizes the fourth-order cumulants to update the parameters of RBF and obtain a novel nonlinear blind source separation method. Our proposed approach outperforms the original method by adopting higher-order cumulants to derive a robust method with RBF framework, when both nonlinear distortion and noises are present.

Higher-order cumulants can help us to reduce noise because of several properties of cumulants which are given as follows. If the sets of random variables $(x_1, x_2, ..., x_n)$ and $(y_1, y_2, ..., y_n)$ are independent, then $cum\{x_1 + y_1, x_2 + y_2, \cdots x_n + y_n\} = cum\{x_1, x_2, \cdots x_n\} + cum\{y_1, y_2, \cdots y_n\}$. And if the sets of random variables $(x_1, x_2, ..., x_n)$ are Gaussian

0-7803-6685-9/01/$10.00©2001 IEEE

distributed, then $cum\left\{x_1, x_2, \cdots x_n\right\} = 0$. Therefore, when we introduced JADE into our algorithm to update RBF, it latently utilized the above properties. The maximization of the criteria in JADE can provide good result and outperforms the original method when high-level Gaussian noise is present in nonlinear mixtures. Thus, the proposed approach can also be regarded to remove noise from signals. The theoretical analysis and computer simulation results will show that, the proposed algorithm has better performance than traditional methods when faced to Gaussian noise in nonlinear mixtures.

The rest of the paper is organized as follows. Sect. II shows a brief description of the noisy nonlinear mixing model. Then the separating model of RBF which is updating by JADE algorithm is presented in Sect. III. We will focus on the proposed algorithm in Sect. IV. Simulation results and comparison of performance of the proposed and conventional methods in artificial and acoustic signal separation are presented in Sect. V. Conclusion is drawn in Sect. VI.

II. Nonlinear Mixing and Demixing Model

We consider a general model of blind source separation:

Denoted $x(t) = \left[x_1(t), x_2(t), \cdots x_n(t)\right]^T$ the vector of observed random vector, and $s(t) = \left[s_1(t), s_2(t), \cdots s_n(t)\right]^T$ the independent source vector, we can obtain the nonlinear mixing model as:

$$x(t) = f\left[s(t)\right], \tag{1}$$

where f is an unknown multi-input and multi-output (MIMO) nonlinear mapping function.

Gaussian noise is used in this paper to mimic a noisy nonlinear mixing model, which can be expressed as:

$$x(t) \quad = \quad f\left[s(t) + G\right], \tag{2}$$

where G is Gaussian noise, what is so-called source noise. The noise is assumed independent with the source commponents. For simplicity and convenience, we also assume the dimensions of source signals and mixtures are equal.

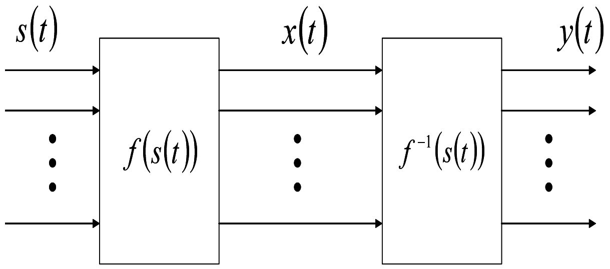

Fig. 1. System model

Figure 1 shows the mixing model expressed in Eq. (1), and blind source separation system. The purpose of the separation system is to recover the original signals and the unmixing matrix (denoted by nonlinear transfer function $f^{-1}(\cdot)$), as well as to remove the Gaussian noise, only from the observations $x(t)$.

Obviously, this problem is difficult to solve without further assumptions. Hence we assume $f(\cdot)$ is componentwise invertible and its inverse $f^{-1}(\cdot) = \left(f_1^{-1}(\cdot), f_2^{-1}(\cdot), \cdots, f_n^{-1}(\cdot)\right)$ exists. We will employ an RBF network to estimate the nonlinear unmixing function, due to its capacity for approximating a function.

Most of the real mixing systems in communications, speech and biomedical areas can be well modeled by the mixing model in Figure 1, in which the nonlinearities introduced by the sensor's nonlinear saturation distortions and nonlinear characteristics occurred in individual receiver channels. Study of the model will provide a solution for a large number of nonlinear separation problems in practice.

III. Radial Basis Function Networks

Neural networks such as multilayer perceptron (MLP) and recurrent neural network (RNN) both have drawback of lower convergence rate. The radial basis function network, initially proposed in [11], is much different from multilayer perceptrons with sigmoidal activation function, in which it utilizes neurons with radial basis function that are locally responsive to input stimulus in the hidden layer.

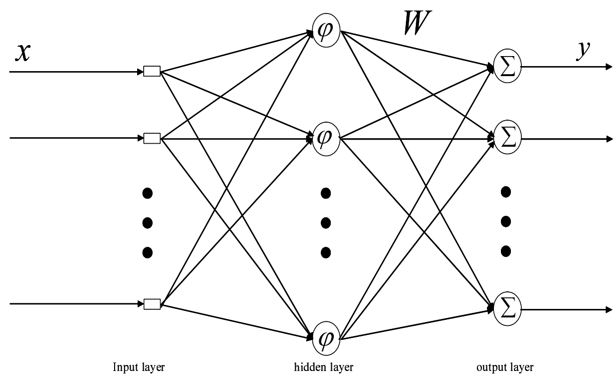

Fig. 2. The RBF network

As shown in Figure 2, an RBF neural network consists of two layers: RBF neuron layer and output layer with linear neurons. An RBF neuron is usually implemented using a Gaussian kernel function, which has two parameters: center and width. The activation function of the RBF neuron is radially symmetirc in the input space, and the output of each RBF neuron depends only on the radial distance between the input vector x and the center parameter for that RBF neuron. The response of each RBF neuron is scaled by its connecting weights to output linear neuron and then summed to produce the overall network output. The whole relationship between input and output of an RBF neural network can be given by:

$$y(x) = \sum_{j=0}^{M} w_j \phi_j(x), \tag{3}$$

and

$$\phi_j(x) = \exp\left(-\frac{1}{\sigma_j^2}(x-\mu_j)^2\right), \qquad (4)$$

where w_j is the connecting weight between the i-th RBF neuron and linear output neuron, $\phi_j(x)$ is the response function of the j-th RBF neuron, μ_j and ϕ_j are the center and width of the j-th RBF neuron, respectively, M indicates the number of RBF neurons in the network.

The performance of an RBF network depends on the number, shape and position of the radial basis functions, and the method used for learning the specific mappping. There are a number of learning strategies for RBF networks such as randomly or employing unsupervised or supervised procedures for selecting the radial basis function centers. In this paper we will use the RBF network to construct the inverse mapping of the nonlinear mixing function shown in Figure 1. The center and width of an RBF neuron are estimated by traditional k-means algorithm. In [12], a mathematical justification for the theory of a nonlinear transformation followed by a linear transformation may be employed in RBF framework. Then the nonlinear observation $x(t)$ transformed into high-dimensional space is more likely to be linearly separable then in a low-dimensional space before transformation. This can be simply verified with XOR problem as in [13]. Accordingly, the weights of output layer are determined by employing JADE algorithm in the high-dimensional space. In the following, we will focus on the learning rules of the parameters in our proposed algorithm.

IV. Joint Diagonalization for RBF Network

Most of the proposed learning algorithms of RBF networks require training signals. They are not suitable to achieve our unsupervised task. As mentioned above, utilizing the characteristic of RBF network which transforms signal in low-dimensional space to high-dimensional space, we then can achieve estimation using a traditional blind source separation algorithm in the high-dimensional space. Motivated by these properties, this paper employs JADE to implement this task since it is a computationally efficient technique. For the properties of higher-order cumulants latently used in JADE, it is also considered robust to Gaussian noise.

This paper will study the nonlinear blind separation problem by adopting the higher-order cumulants for the estimation of RBF's parameters when both nonlinear mixing matrix and high-level noise are present. In some recently developed nonlinear method [14], the higher-moments of sources and estimations are required to be the same. The constraint yields the drawback that all the moment information of sources has to be known, which is too rigorous in real world. Some algorithms based on mutual information require that the pdf of a random variable is known or can at least be estimated accurately. This also lead us to use higher order statistics (HOS). It can be shown that statistical independence of the signals means that the cross-cumulant of all orders should be zero. The proposed higher-order cumulants-based method is not constrained to these severe conditions. The proposed algorithm can separate signals corrupted with high-level noise in general, without prior information of source signals. Thus, it is more flexible and applicable.

A. Joint Approximate Diagonalization of Eigenmatrices

After nonlinear transformation from input to hidden layer of RBF networks, we can likely consider the distortion of nonlinearity is reduced. Therefore it becomes a linear problem:

$$\tilde{x}(t) = Ws(t). \qquad (5)$$

We utilize JADE to estimate the linear mixing matrix W of output layer.

We give a brief introduce of JADE algorithm here, see [15] for more detail.

Joint approximate diagonalization of eigenmatrices exploits the fourth-order cumulants of mixture. Higher-order cumulants have a characteristic that is helpful to measure the dependence among random vector. For example, fourth-order cumulants is a four-dimensional array whose entries are given by the fourth-order cross-cumulants of the data. This can be considered as a four-dimensional matrix, since it has four different indices. The fourth-order cumulants contain all the fourth-order information of the data. The off-diagonal elements of cross-cumulants (all cumulants with $ijkl \neq iiii$) characterize the statistical dependencies among components. If and only if, all components are statistically independent, the off-diagonal elements vanish, and the cumulants (of all orders) are diagonal.

Without any loss of generality, this paper assumes that the source signals have unit variance: $E\left\{|s_i(t)|^2\right\} = 1$ for $1 \leq i \leq n$. For independent sources, we then have

$$C_s = E\left\{s(t)\,s(t)^T\right\} = I_n, \qquad (6)$$

where I_n denotes the $n \times n$ identity matrix, so that

$$C_{\tilde{x}} = WW^T. \qquad (7)$$

We consider exploiting second order information by whitening the signal $\tilde{x}(t)$ of the observation. This is done via a whitening matrix M, i.e. a matrix such that $M\tilde{x}(t)$ is spatially white. The whiteness condition is

$$I_n = MC_{\tilde{x}}M^T = MWW^TM^T \qquad (8)$$

From Eq. 8 we know that MW is a unitary matrix: for any whitening matrix M, it then exists a unitary matrix U such that $MW = U$. Then matrix W can be factored as

$$W = \hat{M}^{-1}U. \qquad (9)$$

Let $N = \{N_r \,|1 \leq r \leq s\}$ be a set of s matrices with common size $n \times n$. A joint diagonalizer of the set N is defined as a unitary maximizer of the criterion

$$C(V,N) = \sum_{r=1,s}\left|diag\left(V^TN_rV\right)\right|^2 \qquad (10)$$

where $|diag(\cdot)|$ is the norm of the vector built from the diagonal of the matrix argument.

On the other hand, JADE algorithm can be considered as minimizing a sum of the squared cross-cumulants of the observed signal [4]: $off\left(\sum cum(y)^2\right)$. Accordingly, from the definition of cumulants, two properties of cumulants can be obtained as follows:

Property 1: If the sets of random variables $(x_1, x_2, ..., x_n)$ *and* $(y_1, y_2, ..., y_n)$ *are independent, then* $cum\{x_1 + y_1, x_2 + y_2, \cdots x_n + y_n\} = cum\{x_1, x_2, \cdots x_n\} + cum\{y_1, y_2, \cdots y_n\}$ [16].

Property 2: If the sets of random variables $(x_1, x_2, ..., x_n)$ *are Gaussian distributed, then* $cum\{x_1, x_2, \cdots x_n\} = 0$.

Obviously, applying the two properties above to the output of the separation system, we can obtain the following relation:

$$\begin{aligned} J &= \left(Cum\,(y+G)^{order}\right)^2 \\ &= \left(Cum\,(y)^{order}\right)^2 + \left(Cum\,(G)^{order}\right)^2 \\ &= \left(Cum\,(y)^{order}\right)^2 \end{aligned} \tag{11}$$

where G is Gaussian noise and B is the weight of output layer. Then, we can conclude that there is no effect on estimations from Gaussian noise, such that better separation results can be obtained.

The proposed algorithm for estimating weights of RBF network contains two steps: first, use traditional k-means algorithm to select the centers and widths of RBF neurons; then update the output layer weights of neural network, by utilizing JADE algorithm.

The original k-means algorithm is described as follows, briefly. Let $\mu_m(n)$, $(m = 1, 2, ..., M)$ be the centers of RBF neurons at iteration n, the best-matching center $\hat{\mu}(x)$ at iteration n using the minimum-distance Euclidean criterion can be found as:

$$\hat{\mu}(x) = \arg\min_m \|x - \mu_m(n)\|, \quad m = 1, 2, ..., M \tag{12}$$

The update rule for the locations of the centers are given by

$$\mu_m(n+1) = \begin{cases} \mu_m(n) + \eta\,[x - \mu_m(n)]\,, \mu = \hat{\mu}(x) \\ \mu_m(n)\,, \qquad\qquad\qquad\quad otherwise \end{cases} \tag{13}$$

where η denotes the learning rate.

After estimating the centers and widths of hidden layer, the weights of output layer is updated by maximizing the cost function (10) using JADE.

The next section summaries the learning rule.

B. Description of Learning Rule

In this paper, the simulation results are analysed based on the root-mean-squares error (RMSE), which is given as

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(s-\hat{s})^2}, \tag{14}$$

where $\hat{s}$ is recovered source signals, N denotes the number of signals.

The proposed algorithm can be summarized:

a. Initialize the centers, widths and weights of RBF network by using small random number; and the termination condition of learning by a small positive numbers ε.
b. Update the centers and widths of RBF network's hidden layer by using k-means algorithm.
c. After updating parameters of hidden layer, form the sample covariance $\hat{C}_{\tilde{x}}$ and compute a whitening matrix $\hat{M}$.
d. Form the sample fourth-order cumulants $\hat{Q}_z$ of the whitened process $\hat{z}(t) = \hat{M}\tilde{x}(t)$; compute the n most significant eigenpairs $\left\{\hat{\lambda}_r, \hat{N}_r \,|1 \le r \le n\right\}$.
e. Jointly diagonalize the set $\hat{N}^e = \left\{\hat{\lambda}_r\hat{N}_r \,|1 \le r \le n\right\}$ by a unitary matrix $\hat{U}$.
f. Estimate W by $\hat{W} = \hat{M}^{-1}\hat{U}$.

V. Simulation and Results Analysis

In this section we present and discuss simulation results of the proposed algorithm. We will make a comparision of the separation results between the proposed algorithm and traditional JADE and RBF-based algorithms, together with a performance analysis by using RMSE. In order to make the comparison, we use RBFN with 2 inputs, 6 hidden neurons and 2 outputs, which are the same as traditional method.

A. Example 1

We first apply the proposed algorithm to denoise artificial signal to test its efficiency. The signal used in this experiment is a modulated sinusoid signal and a Gaussian noise with 10dB SNR. The number of iteration is 5000. The mixing process is described as:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = A_2 \begin{bmatrix} (\cdot)^3 \\ (\cdot)^3 \end{bmatrix} A_1 \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}, \tag{15}$$

where $A_1 = \begin{pmatrix} 0.25 & 0.86 \\ -0.86 & 0.25 \end{pmatrix}, A_2 = \begin{pmatrix} 0.5 & 0.9 \\ -0.9 & 0.5 \end{pmatrix}$.

Figure 3 shows two source signals, which are supposed unknown in real situation, mixed by Eq. (15) into two mixtures, as shown in Fig. 4.

Since the mixtures suffered nonlinear distortion and noise simultaneouslyCit is difficult to separate them using conventional JADE method, which is not suitable for nonlinear problem. Also, a nonlinear method using RBF network with matching higher-order moments fails to reduce noise. On the opposite, the proposed algorithm can obtain better separation results by updating RBF network's parameters with higher-order cumulants (JADE) to recover source signal, as illustrated

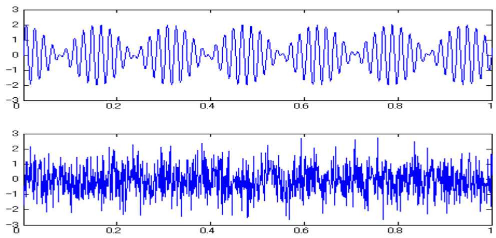

Fig. 3. Source signals.

in Figure 5. Clearly, noise is reduced from modulated sinusoid in a nonlinear mixing manner.

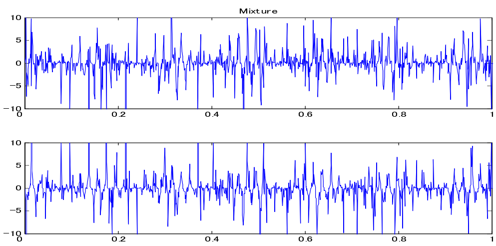

Fig. 4. Mixtures.

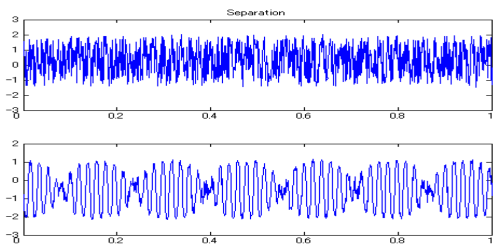

Fig. 5. Separation results.

Here we only show the results of proposed algorithm as the other two traditional algorithms cannot recover the sources. We employ the RMSE to compare the performance of the proposed and original methods, as illustrated in Figure 6. From performance comparison, we can see that even the two RBF-based methods have almost the same convergence rates, the proposed algorithm can reduce noise in nonlinear mixtures with smaller RMSE. On the other hand, the JADE method cannot reduce noise from nonlinear mixture.

B. Example 2

We also performed a simulation with a human voice and a Gaussian noise with 10dB SNR. Figure 7 shows the source signals, which are mixed by Eq. (15) into mixtures, illustrated in Figure 8. The separation results are presented in Figure 9.

From the separation result illustrated in Figure 9, we can see that the proposed algorithm provides good estimations when

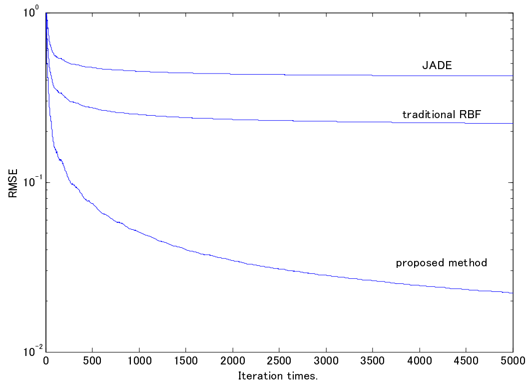

Fig. 6. The RMS error comparison of the proposed method and the orginal algorithms.

apply to acoustic signal denoising problem. Thus it can be used as a denoising method.

The comparison of proposed algorithm and traditional methods is shown by Figure 10. We can see, our algorithm is more suitable when applies to blind signal separation problem with nonlinear distortion present, which cannot be achieved by using traditional algorithms.

VI. Conclusion

An unsupervised nonlinear learning structure, by using JADE method, for improving traditional RBF is proposed in this paper. As it has been shown, the proposed algorithm provides robust and accurate estimation for noisy mixtures, such that it is more suitable for unsupervised signal denoising. The simulation result also showed that the proposed algorithm is applicable for acoustic signal denoising and separation problem.

The traditional RBF and JADE algorithms can hardly estimate source signals, when the prior information of them is unknown. In our algorithm, the prior knowledge is not required, thus it is more suitable for various situations.

Our proposed algorithm showed the flexibility when applied to signal, therefore, it can be regarded as a generalization for nonlinear unsupervised learning.

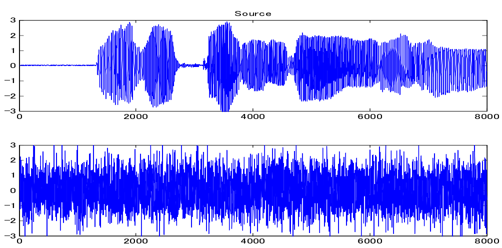

Fig. 7. Source signals.

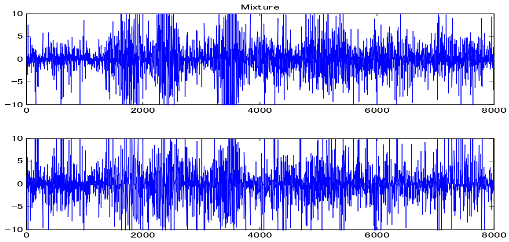

Fig. 8. Mixtures.

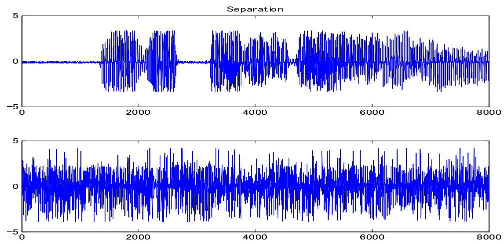

Fig. 9. Separation results.

References

[1] I. Cha and S. Kassam, "Channel equalization using adaptive complex radial basis function networks," *IEEE J. Sel. Areas*, vol. 13, pp. 122–131, Jan 1995.

[2] A. Hyvärinen, "Independent component analysis in the presence of gaussian noise by maximizing joint likelihood," *Neurocomputing*, vol. 22, pp. 49–67, 1998.

[3] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–20, 1991.

[4] E. O. A. Hyvärinen, J. Karhunen, *Independent Component Analysis*. Wiley, New York, 2001.

[5] P. P. A. Hyvärinen, "Nonlinear independent component analysis: Existence and uniqueness results," *Neural Networks*, vol. 12, pp. 429–439, 1999.

[6] A. H. P. Pajunen and J. Karhunen, "Nonlinear blind source separation by self-organizing maps," *Proc. ICONIP'96*, vol. 2, pp. 1207–1210, New York, 1996.

[7] M. Herrmann and H. H. Yang, "Perspectives and limitations of self-organizing maps in blind separation of source signals," *Progress in Neural Information Processing: Proceedings of ICONIP'96*, pp. 1211–1216, 1996.

[8] S. A. H. H. Yang and A. Cichocki, "Information backpropagation for blind separation of sources from nonlinear mixture," *Proc. IEEE ICNN, TX*, pp. 2141–2146, 1997.

[9] C. J. A. Taleb and S. Olympieff, "Source separation in post nonlinear mixtures: an entropy-based algorithm," *Proc. ESANN'98*, pp. 2089–2092, 1998.

[10] N. C. R. S. Bhattacharya and S. Sinha, "2-D signal modelling and reconstruction using third-order cumulants," *Signal Processing*, vol. 62, no. 1, pp. 61–72, Oct 1997.

[11] D. L. B.S. Broomhead, "Multi-variable functioinal interpolation and adaptive networks," *Complex Systems*, vol. 2, pp. 321–355, 1990.

[12] T. Cover, "Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition." *IEEE Transactions on Electronic Computers EC-14*, pp. 326–334, 1965.

[13] S. Haykin, *Neural Networks*. Englewood Cliffs, NJ.: Macmillan College Publishing Company, Inc., 1994.

[14] J. Ying Tan; Jun Wang; Zurada, "Nonlinear blind source separation using a radial basis function network," *Neural Networks, IEEE Transactions on*, vol. 12, no. 1, pp. 124–134, Jan 2001.

[15] S. A. Cardoso, J.F., "Blind beamforming for non gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, Dec 1993.

[16] A. P. C.L. Nikias, *Higher-Order Spectra Analysis: a Nonlinear Signal Processing Framework*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

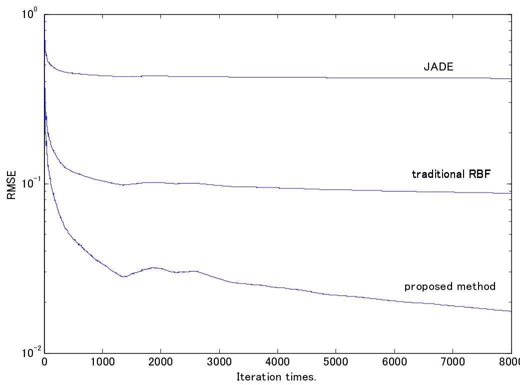

Fig. 10. The RMS error comparison of the proposed method and the orginal algorithms.