Blind Source Separation for Changing Source Number: A Neural Network Approach with a Variable Structure[1]

Shun-Tian Lou, Member, IEEE and Xian-Da Zhang, Senior Member, IEEE
Key Lab for Radar Signal Processing, Xidian University, Xi'an 710071, China

ABSTRACT

Blind source separation (BSS) problems have recently become an active research area in both statistical signal processing and unsupervised neural learning. In most approaches, the number of source signals is typically assumed to be known a priori, but this does not usually hold in practical applications. Although the problem of determining the unknown source number has been studied recently, the BSS problem when the source number is changing dynamically is not yet considered. The main objective of this paper is to study and solve these two problems. Its basic idea is to utilize the correlation coe¢ cients between output components of the neural network (NN) as a mean for determining the unknown source number and/or detecting dynamical change of the source number, and is to develop a neural network with variable structure to perform the corresponding adaptive blind source separation.

1. INTRODUCTION

In recent years, blind source separation (BSS) problems have received increasing interest and have become an active research area in both statistical signal processing and unsupervised neural learning [1]-[12], [16]-[18]. The goal of BSS is to extract statistically independent but unknown source signals from their linear mixtures without knowing the mixing coe¢ cients. The BSS techniques have many potential applications in several areas, such as data communications, speech processing, various biomedical signal processing (MEG/EEG), and array processing in radar and communications.

Many neural learning algorithms have been developed for the BSS problem. Especially, some recently developed algorithms are surprisingly good [1]-[4],[6]-[12], [16]-[18]. However, in their corresponding models and network architectures it is usually assumed that the source number is known a priori and …xed. In most neural approaches to BSS, a typical assumption is that the source number should be equal to the number of sensors (or array elements) and NN outputs. Unfortunately, these assumptions do not necessarily hold in practice.

As Amari and Cichocki [3] pointed out, although recently developed many algorithms are able to successfully separate source signals, there are still many problems to be studied, for example, development of learning algorithms which work: 1) when the number of the source signals is unknown; 2) when the number of source signals is dynamically changing. These two problems are very important for many real-world applications, such as cellular wireless communication systems, array processing in radar, and so on.

The problem of determining the number of source signals has been considered only recently, and two approaches has been proposed [9], [10], [12]. The …rst approach is principal component analysis (PCA) based pre-whitening in noisy conditions. In nature, this is also a subspace approach. The second approach uses a post-processing layer for elimination of redundant signals. To our knowledge, the BSS problem in the case of the source number changing dynamically is not yet studied in available literature.

The main objective of this paper is to study and solve the above two BSS problems. Its basic idea is to utilize the correlation coe¢ cient and the correlation measure of NN output components as a mean for determining the unknown source number and/or detecting dynamical change of the source number, and is to develop a NN with variable structure to perform the corresponding adaptive blind source separation.

The paper is organized as follows. In Section 2 we describe the general BSS problem formula. In Section 3 we …rst present a theoretical analysis to an important observed phenomenon when the NN output number is larger than the source number, and then de…ne correlation coe¢ cients and correlation measures for NN output components, and proposes a new algorithm for determining the unknown source number. In Section 4 we study and solve the BSS problem when the source number is dynamically changing. The neural network approach with a variable structure consists of the three algorithms. The statistical results of independent simulation runs are given in Section 5 to show the e¤ectiveness of the new NN approach. Conclusions are made in Section 6.

2. PROBLEM FORMULATION

Let $s_1(t); ¢¢¢; s_n(t)$ be n zero-mean source signals that are scalar-valued and mutually (spatially) statistically independent (or as independent as possible) at each time instant t. The original source signals $s_i(t)$ are mixed by a $m £ n$ mixing matrix A, and are observed by m sensors or array elements. In real applications, the number of sensors, m, is known and …xed.

[1] This work was supported by the National Natural Science Foundation of China.

Assume $x(t) = [x_1(t); ¢¢¢; x_m(t)]^T$ be the m-dimensional data vector made up of m mixtures. Then, the mixing (data) model can be written in the following vector form:

$$x(t) = As(t) \tag{1}$$

where $s(t) = [s_1(t); ¢¢¢; s_n(t)]^T$ is the source vector consisting of n source signals at t, and A is a full-rank $m £ n$ mixing matrix.

The task of blind source separation is to recover the waveforms of the sources $fs_i(t)g$ using the mixtures $x_1(t); ¢¢¢; x_m(t)$ without the knowledge of the mixing matrix A. It is well-known (see, e.g., [5], [15]) that it is impossible to separate several Gaussian source signals from each other. Hence, at most one of the source signals $s_i(t)$ is allowed to have a Gaussian distribution.

To separate n source signals, we use a feed-forward linear NN as a simple linear separating system:

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

where $y(t) = [y_1(t); ¢¢¢; y_n(t)]^T$ is an estimate of s(t); and W(t) is a $n £ m$ separating (or de-mixing) matrix. The key of adaptive blind source separation is to …nd W(t) in a recursive way so that y(t) = W(t)x(t) is an estimate of the source signal vector s(t).

There are three main methods for blind signal separation : the independent component analysis (ICA), the maximum entropy (ME) and nonlinear PCA. The ICA is identical to minimizing the mutual information of the output y(t). The basic idea of minimum mutual information is to choose W that minimizes the dependency among the components of y(t). For non-Gaussian (sub-Gaussian and/or super-Gaussian) source signals, uncorrelatedness is a necessary prerequisite for the stronger independence condition.

To measure the dependency between non-Gaussian output components, we must use higher-order cumulants of the output vector y(t); namely consider the nonlinear transformation of y(t). To this end, let $z_i = g_i(y_i); i = 1; ¢¢¢; n$ be the nonlinear componentwise transformed outputs, and $z = [g_1(y_1); ¢¢¢; g_n(y_n)]^T$. It is shown [16] that subject to maximizing the joint entropy H(z; W) or minimizing mutual information of z; it gives the following learning rule:

$$\frac{dW}{dt} = ´(t)\frac{@H(z;W)}{@W}W^TW = ´(t)\left[I ¡ Á[y(t)]y^T(t)\right]W \tag{3}$$

or

$$W(t+1) = W(t) + ´(t)\left[I ¡ Á[y(t)]y^T(t)\right]W(t) \tag{4}$$

where $´(t) > 0$ is the adaptive learning rate, and I is the $n £ n$ identity matrix. Since $\frac{@H(z;W)}{@W}W^TW$ is called the natural gradient, (4) is referred to as the natural gradient learning rule.

The natural gradient learning rule is very easy to implement, since Á(y) can be computed as follows [16]:

$$\begin{aligned} Á(y) &\overset{¢}{=} [Á(y_1); ¢¢¢; Á(y_n)]^T \\ &= f(\cdot_3; \cdot_4) ± y^2 + g(\cdot_3; \cdot_4) ± y^3 \end{aligned} \tag{5}$$

where ± denotes the Hadamard product of two vectors:

$$f ± x = [f_1x_1; ¢¢¢; f_nx_n]^T; \tag{6}$$

and

$$y^k = \left[(y_1)^k; ¢¢¢; (y_n)^k\right]^T \quad \text{for } k = 2; 3; \tag{7}$$

$$f(\cdot_3; \cdot_4) = \left[f(\cdot_3^1; \cdot_4^1); ¢¢¢; f(\cdot_3^n; \cdot_4^n)\right]^T; \tag{8}$$

$$g(\cdot_3; \cdot_4) = \left[g(\cdot_3^1; \cdot_4^1); ¢¢¢; g(\cdot_3^n; \cdot_4^n)\right]^T: \tag{9}$$

while the functions f(x; y) and g(x; y) are given by

$$f(x;y) = ¡\frac{1}{2}x + \frac{9}{4}xy; \tag{10}$$

$$g(x;y) = ¡\frac{1}{6}y + \frac{3}{2}x^2 + \frac{3}{4}y^2; \tag{11}$$

and the third- and fourth-order cumulants of the i-th output component, $\cdot_3^i$ and $\cdot_4^i$ can be estimated adaptively as:

$$\frac{d\cdot_3^i}{dt} = ¡´\left[\cdot_3^i ¡ (y_i)^3\right]; \quad i = 1; ¢¢¢; n \tag{12}$$

$$\frac{d\cdot_4^i}{dt} = ¡´\left[\cdot_4^i ¡ (y_i)^4 + 3\right]; \quad i = 1; ¢¢¢; n \tag{13}$$

where ´ is the learning rate for updating the cumulants of output components.

The two standard assumption in blind source separation are: 1) the number n of the source signals should be known, 2) the number n of the source signals and output channel number l are equal in the separation network. Generally, both these assumptions are usually not hold in practice.

3. SOURCE NUMBER DETERMINATION

Like in most neural blind source separation approaches, the only additional requirement in this paper is that the number of available sensors or array elements, m; is greater than or equal to the true number n of source signals, that is, $m > n$.

It is well-known that there is indeterminacy or ambiguity in both exact order and amplitude of separated source signals. Since the number m of sensors should be larger than the unknown number n of source signals, the $m £ n$ mixing matrix A is a 'high' matrix. To solve the problem of source number determination, we use a dynamical neural network (DNN). As compared to the general NN with …xed structure, the DNN has a variable structure in which the input number m is …xed, but the output number l is variable. Clearly, $n \le l \le m$. In such a case,

the weight matrix W of the neural network in the learning rule (4) is an $l \times m$ 'wide' matrix. Hence, we should find a separating or de-mixing matrix W adaptively such that

$$W = \Lambda P A^{y} \tag{14}$$

where Λ is a non-singular $l \times l$ diagonal matrix, $A^{y} = (A^{T}A)^{-1}A^{T}$ is the $n \times m$ Moore-Penrose generalized inverse of A, and P is an $l \times n$ 'high' generalized permutation matrix whose each row has one unity;while each column has one or more unity except from zero elements. Using (14), we have $y(t) = Wx(t) = \Lambda P A^{y} \cdot As(t) = \Lambda Ps(t)$; from which it is easy to see that there are n separated source signals and $l - n$ copies (with different scaling factors) of some source signals among the neural network's l outputs. A signal and its copy with different scaling factor are coherent in the sense that their correlation coefficient is theoretically equal to unity. Therefore, the well-known key conclusion of the ICA should be corrected as follows: only when the output number l is equal to the source number n, minimizing the mutual information of the output vector y(t) makes its all components become independent; whereas if $l > n$, minimizing the mutual information of y(t) can only make its n components become independent, while other $l - n$ components will be coherent with some original source signals.

Here, we proposes a new practical approach for determining the source number on-line by detecting correlations between NN output components. As we concluded above, when $l > n$, there are n independent output components and $l - n$ coherent components among l NN outputs. Importantly, coherent signals are easily discriminated. To this end, we define correlation coefficients as

$$r_{ij} = \frac{\mathrm{cov}[y_i(t); y_j(t)]}{\sqrt{\mathrm{cov}[y_i(t)]\mathrm{cov}[y_j(t)]}}; \quad i, j = 1, \cdots, l \text{ but } i \neq j \tag{15}$$

where $\mathrm{cov}[x(t); y(t)] = E\{[x(t) - \bar{m}_x][y(t) - \bar{m}_y]\}$; $\mathrm{cov}[x(t)] = E\{[x(t) - \bar{m}_x]^2\}$; and $\bar{m}_x = E\{x(t)\}$. Note that $0 \leq |r_{ij}| \leq 1$.

On the other hand, the correlation measure of $y_i(t)$ on all of other output components is defined as

$$D_i(t) \triangleq D(y_i(t)) = \sqrt{\frac{1}{l-1} \sum_{j=1, j \neq i}^{*} (r_{ij})^2}; \quad i = 1, \cdots, l. \tag{16}$$

The correlation coefficients r_{ij} and the correlation measures $D_i(t)$ are very important to the BSS with the unknown source number and the changing source number. Clearly, if $|r_{ij}|$ is close to unity for some $y_i(t)$ and $y_j(t)$ after the NN converges to its equilibrium point; then $y_i(t)$ and $y_j(t)$ can be considered to be coherent. On the other hand, if $D_i(t)$ are sufficiently small for all i, then all output components can be considered to be separated from each other. In adaptive blind source separation, r_{ij} and $D_i(t)$ can be computed recursively.

A variable learning rate scheme is proposed with the following diagonal form

$$\Gamma(t) = \mathrm{diag}\left[\gamma_1(t) \quad \gamma_2(t) \quad \cdots \quad \gamma_l(t)\right] \tag{17}$$

The $\gamma_i(t)$ is only the learning rate of the output component $y_i(t)$. If $D_i(t)$ is sufficient small, the output signal $y_i(t)$ is considered to be uncorrelated with all of other output components. Otherwise, $y_i(t)$ is correlated with other outputs, and the larger $D_i(t)$, the stronger the correlatedness. A natural choice of $\gamma_i(t)$ is that the smaller $D_i(t)$ is, the lower the learning rate $\gamma_i(t)$ should be, vice versa. More specifically, we choose $\gamma_i(t)$ as

$$\gamma_i(t) = \eta(D_i(t)); \qquad i = 1, \cdots, l \tag{18}$$

where $\eta(\cdot) > 0$ is a nonlinear function, for example it can be piecewise linear for simplicity.

Our approach to determining the unknown number of source signal on-line are as follows:

Algorithm 1 (Source Number Determination)

Step 1: In the initialization stage of running learning algorithm (4), we take $l = m$. Since there are the larger bias in the estimates of both the correlation coefficient r_{ij} between $y_i(t)$ and $y_j(t)$ and the correlation measure $D_i(t)$ of $y_i(t)$ computed recursively from a shorter set of samples, we take $\Gamma(t) = \mathrm{diag}(\alpha, \cdots, \alpha)$ until 300 samples; where α is a larger constant.

Step 2: After 300 samples, we use $\gamma_i(t) = \eta(D_i(t))$ to run the learning algorithm (4).

Step 3: If all of the correlation measures $D_i(t); i = 1, \cdots, m$ are sufficiently small (for example $D_i(t) < 0.05$), we consider all of source signals to be separated from each other.

Step 4: Detect a pair of output components $(y_i; y_j)$ with sufficiently large correlation coefficient r_{ij} (for example $r_{ij} > 0.8$), and delete either $y_i(t)$ or $y_j(t)$. Repeat this deleting process until all of the redundant copied signals are deleted.

We refer to Step 1 as the initialization stage, Steps 2 and 3 as tracking stage, and Step 4 as source number determination stage. If applying the learning algorithm (4) with fixed learning rate, Step 2 is omitted.

In the tracking phase, the redundant output signal (such as $y_i(t)$) will be a copy of another signal (such as $y_j(t)$). Without loss of generality, we promise that the output $y_i(t)$ will be deleted. To do this, the ith row of W is deleted and its size is become $(l - 1) \times m$. In a similar way, the ith row and ith column in matrices $\Gamma(t)$ and Λ are deleted. In this way, the NN has still m inputs, but the number of outputs becomes $l - 1$ from l. It is easy to verify that the NN output vector is $y(t) = [y_1(t), \cdots, y_{i-1}(t), y_{i+1}(t), \cdots, y_l(t)]^T$; i.e., $y_i(t)$ has been deleted, while all others are unchanged. Repeating the above process, we finally have n separated source signal outputs.

4. SOURCE NUMBER CHANGING

In standard adaptive source separation approaches, it is assumed that the source number is unchanged. However,

this assumption is not true in many practical applications. A typical example is cellular wireless communication systems within which the number of users is changing dynamically. This dynamical change includes both source number decreasing and increasing. The case of source number decreasing is discussed in detail in this paper.

Once the number n of source signals are determined, the NN has m inputs and n outputs. Let us consider the case when the signal $s_k(t)$ separated in the output channel y_i disappears suddenly at some time. When the number of output channels is decreased from n to $n - 1$, there appears the copy of another source signal (say $s_d(t)$) in output channel y_i. For convenience, let the separated signal $s_d(t)$ output at the channel y_j: Computing recursively the correlation coe¢cient r_{pq} for each pair of di¤erent outputs $(y_p; y_q)$ with the same time segment, and detecting whether r_{pq} is su¢ciently large, we can …nd the output pair $(y_i; y_j)$ such that $y_i(t)$ is a scaled copy of $y_j(t)$: Note that unlike the case of determining the unknown number of source signals, we should delete $y_i(t)$ rather than $y_j(t)$ here: How to determine which one is $y_i(t)$? For this purpose, we de…ne the following self-similar degree of a signal for two time segments:

$$ss_k(\tau) = \frac{1}{L}\sum_{t=0}^{L-1} y_k(t+T)y_k(t+T-\tau); \qquad k = i; j \quad (19)$$

where L is the length of a segment of output signal $y_k(t)$, T is its starting point of time; $T + L - 1$ is its stopping point of time, and τ is the distance between the two segments of $y_k(t)$: Clearly, $T + L - 1$ should be equal to the current time. Since the self-similar degree describes the correlation degree of two di¤erent time segments of the same signal, it is easy to see that by selecting appropriate τ; the output with a small self-similar degree is $y_i(t)$ to be deleted, while another output has a large self-similar degree and should remain. Then, we can use a similar way to construct the $(n-1)\times(n-1)$ diagonal matrix $\Gamma(t)$, the $(n-1)\times(n-1)$ matrix Λ and the $(n-1)\times m$ de-mixing matrix W: After implementing the learning algorithm (4), the NN with m inputs and $n - 1$ outputs can track the remaining $n - 1$ signals.

The above discussion can be summarized into the following algorithm.

Algorithm 2 (Source Number Decreasing)

Step 1: Find an output pair $(y_i; y_j)$ with su¢ciently large correlation coe¢cient r_{ij} (for example $r_{ij} > 0:8$).

Step 2: Determine y_i or y_j should be deleted by using (19) to inspect their self-similar degrees.

Step 3: Delete the redundant signal.

The case of source number increasing is somewhat complex. To make the NN work well under this new environment, it should have the following abilities: 1) Detect whether one or more new source signals add, 2) Use a variable structure to track new source signal(s) together with the original n source signals.

The …rst task can be completed in very short time by detection the correlation measure. Before a new source signal adds, the NN lies at its equilibrium point, meaning the output components are independent each other. That is to say, all correlation coe¢cients r_{ij} are small enough, and hence the correlation measure $D_i(t)$ of $y_i(t); i = 1; ¢¢¢; n$ on all of other output components are su¢ciently small as well. When a new source signal adds, the equilibrium point of the NN is broken, which may result to a large change of correlation measures $D_i(t)$ for all output components: By detecting such a change, we know whether a new source signal adds. In general, $D_i(t) < 0:05$ when the NN lies at its equilibrium point, while $D_i(t) > 0:3$ soon after a new source signal adds.

But for second task, we have not any good method except for re-initialization NN weight. Our future investigations will focus on the new method for this case.

5. SIMULATIONS

In order to show the e¤ectiveness of the NN with a variable structure for adaptive blind source separation, we apply it to separate source signals from the observed mixtures. For comparison, we run the learning algorithm (4) with …xed learning rate ($\eta = 55$) and with the diagonal variable learning rate matrix $\Gamma(t)$ to each simulation. For convenience of statement, we call them Method 1 and Method 2, respectively. When applying Method 2, we use the following formula to compute the learning rates:

$$\eta_i(t) = \begin{cases} K_1 D_i(t) - 10; & D_i(t) > 0:05 \\ K_2 D_i^2(t); & D_i(t) \le 0:05 \end{cases} \quad (20)$$

$$\eta_i(t) = 250 \quad \text{if } \eta_i(t) > 250 \quad (21)$$

where $K_1 = 1200$ and $K_2 = 20000$ were taken.

Example 1: This is an example for determining the unknown number of source signals. The …ve source signals are as follows:

$$s(t) = \begin{bmatrix} \text{sign}(\cos(2\pi 155t)) \\ \sin(2\pi 25t)\sin(2\pi 800t) \\ \sin(2\pi 300t + 6\cos(2\pi 60t)) \\ \sin(2\pi 90t) \\ n(t) \end{bmatrix} \quad (22)$$

The sign(¢) is a sign function, and n(t) is a noise source uniformly distributed in $[-1; +1]$.

In simulation, we take $l = m = 10$ in Methods 1 and 2. The elements of the 10×5 mixing matrix A are uniformly distributed random numbers in $[-1; +1]$.

Fig.1 shows the separation process how to determine the unknown number of source signals obtained by using Method 2. To measure the performance of the algorithms, we use the following cross-talking error [17]:

$$E = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|b_{ij}|}{\max_k |b_{ik}|} - 1\right) + \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{|b_{ij}|}{\max_k |b_{kj}|} - 1\right) \quad (23)$$

where $B = \{b_{ij}\} = WA$ is a $l\times n$ matrix,

$\max_k |b_{ik}| = \max\{|b_{i1}|, ¢¢¢, |b_{in}|\}$; and $\max_k |b_{kj}| = \max\{|b_{1j}|, ¢¢¢, |b_{lj}|\}$. Fig.2 shows the statistical result (the mean and the standard deviation) of the cross-talking errors for 100 independent runs of Methods 1 and 2:

Example 2: Here we consider the case when a source signal disappears suddenly. The source signals are still given by (22) until the 2000 th sample, but the source signal sign(cos(2¼155t)) disappears since the 2001st sample. In start learning phase, we take l = m = 6.

Fig. 3 is the separation processes provided by Methods 2. The mean and the standard deviation of the cross-talking errors for 100 independent runs of Methods 1 and 2 are given in Fig.4.

From statistical results in Figs.2 and 4, we see that Method 2 (with variable learning rate) has the smaller cross-talking error, faster convergence rate and better reconstructed signal waveforms, as compared with Method 1 with a …xed learning rate.

6. CONCLUSIONS

In standard neural and adaptive BSS approaches, the source number is assumed to known and/or …xed. However, these assumptions may not hold in practice. In this paper, we study the BSS problems when the source number is unknown and is changing dynamically, and propose a NN approach with a variable structure to solve the two problems.

The main results of this paper can be summarized as follows:

1. A theoretical analysis is given to demonstrate why the NN with m inputs and l outputs has n independent output components and l ¡ n (scaled) copies of some source signals appear at the output when the source number n is less than the NN output number l.

2. Using the correlation coe¢cients and the correlation measure, we present the method to determine the number of redundant signals and to detect the change of source number.

3. By changing the structure of the neural network, we propose the two adaptive source separation algorithms for source number determination and source number decreasing, respectively .

The simulation results show the e¤ectiveness of the algorithms.

Results

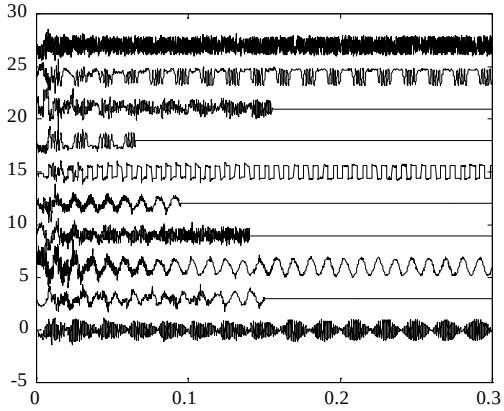

Fig.1 The separation process using Method 2 for source number determination

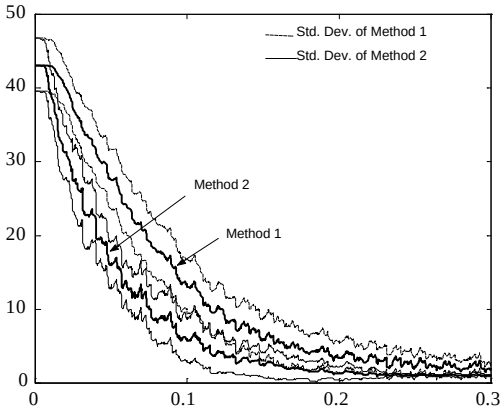

Fig.2 Statistical results of the two methods for the source number determination

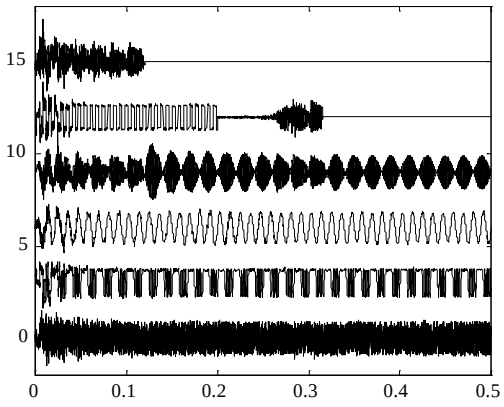

Fig.3 The separation process using Method 2 for source number decreasing

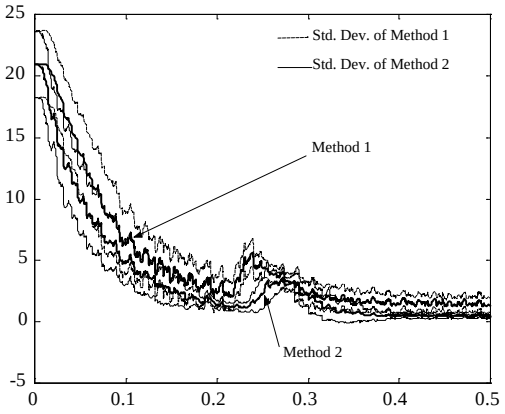

Fig.4 Statistical results of the two methods for the source number decreasing

7. REFERENCES

References

[1] S. Amari, A. Cichocki, and H -H. Yang, "A new learning algorithm for blind signal separation," in Advanced in Neural Information Processing Systems, vol.8, Boston, MA: MIT Press, 1996, pp.752-763.

[2] S. Amari, "Natural gradient works e¢ ciently in learning," Neural Comput., vol.10, pp.251-276, 1998.

[3] S. Amari and A. Cichocki, "Adaptive blind signal processing – – Neural network approaches,' Proc. IEEE, vol. 86„ pp. 2026-2048, 1998.

[4] A. J. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," Neural Comput., Vol. 7, pp. 1129-1159, 1995.

[5] X.-R. Cao and R.-W. Liu, "A general approach to blind source separation," IEEE Trans. Signal Processing, vol.44, 562-571, 1996.

[6] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," IEEE Trans. Signal Processing, vol.44, pp.3017-3030, 1996.

[7] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust learning algorithm for blind separation of signals," Electr. Letters, vol.30, no.17, 1386-1387, 1994.

[8] A. Cichocki and R. Unbehauen, "Robust neural networks with on-line learning for blind identi…cation and blind separation of sources," IEEE Trans. Circuits and Systems I: Fundamentals Theory and Applications, vol.43, 894-906, 1996.

[9] A. Cichocki and W. Kasprzak, "Local adaptive learning algorithms for blind separation of natural images," Neural Network World 6, 1996, IDG Co., Prague, 515-523.

[10] A. Cichocki, J. Karhunen, W. Kasprzak, and R. Vigario, "Neural networks for blind separation with unknown number of sources," Neurocomputing, vol.24, 55-93, 1999.

[11] P. Comon, "Independent component analysis, a new concept?" Signal Processing, vol.36, pp.287-314, 1994.

[12] J. Karhunen, A. Cichocki, W. Kasprzak, and P. Pajunen, "On neural blind separation with noise suppression and redundancy reduction," Int. J. Neural Systems, vol.8, 219-237, 1997.

[13] R. Linsker, An application of the principle of maximum information preservation to linear systems, in Advances in Neural Information Processing Systems 1, Touretzky D.S. (ed), Morgan-Kau¤man, 1989.

[14] A. Stuart and J. K. Ord, Kendall's Advanced Theory of Statistics. Edward Arnold, 1994.

[15] L. Tong, R. -W. Liu, V. C. Soon, and Y.-F. Huang,"Indeterminacy and identi…ability of blind identi…cation," IEEE Trans. Circuits and Systems, vol.38, 499-509, 1991.

[16] H.-H. Yang and S. Amari, Adaptive on-line learning algorithms for blind separation – – Maximum entropy and minimum mutual information, Neural Comput., vol.9, pp.1457-1482, 1997.

[17] H. -H. Yang, S. Amari, and A. Cichocki, Information-theoretic approach to blind separation of sources in non-linear mixture, Signal Processing, vol.64, pp.291-300, 1998.

[18] H. -H. Yang, "Series updating rule for blind separation derived from the method of scoring," IEEE Trans. Signal Processing, vol.47, pp.2279-2285, 1999.