Adaptive Blind source Separation with Dynamical Changing Source Number

Donghui Xu, DongDong Yang, HongGuang Ma, Wenpu Guo, Hui Li

Xi'an Research Inst. Of Hi-Tech Hongqing Town, Xi'an, P. R. 710025, China

e-mail: xdh_45@163.com

Abstract—**Analyzed the reason why the overdetermined blind signal separation of natural gradient algorithm ultimate not stably converge, constructed an adaptive blind separation algorithm suitable for dynamically changing source number. The new algorithm not only overcomes the shortcoming of eventually not stable convergence, but it also need not to delete the redundancies according to the correlation of the output components, which greatly simplifies the computational complexity of the algorithm. Randomly reduce or increase the number of source signals is still able to achieve a better separation effect, simulation results demonstrate the convergence of the algorithm and its effectiveness for blind source separation.**

Keywords- number of source signals, dynamical changes, self-adaptation, blind source separation

I. INTRODUCTION

Blind source separation has been developed for twenty years as an emerging technology discipline, and has made a lot of development in theoretical and practical applications. Now, the blind source separation technology has become the research focus both at home and abroad, and being widely applied in many areas, many scholars dedicated themselves to theoretical and applied research in their respective areas, and more and more towards practical application. It has a wide range of applications in wireless communications, speech recognition, biomedical signal processing, array signal processing, radar signal processing and image denoising and other fields.

Adaptive processing gradually updates the processor parameters according to the accessed data, proceeds to gradually approach the desired results (for blind source separation, the result is each component tends to be independent). Adaptive blind source separation principle is shown in Figure 1:
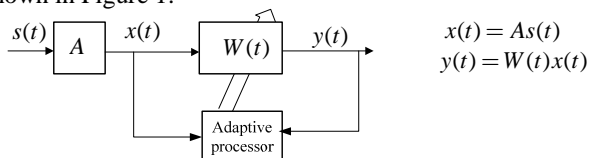

Figure1. Adaptive blind source separation

Adaptive blind separation is adaptively update the separation matrix to make the output

$$y(t) = W(t)x(t) \qquad (1)$$

approximates the source signal vector $s(t)$ as much as possible. If an equivalent changing blind signal separation algorithm becomes adaptive, it will inherit the characteristics with the batch algorithm that it provides "uniform performance". In order to separate the sources, a blind source separation algorithm often use forward linear neural network (NN) as the separation system. The start point of this study is adaptive neural network based independent component analysis. Blind signal separation's neural network architecture is shown in Figure 2:
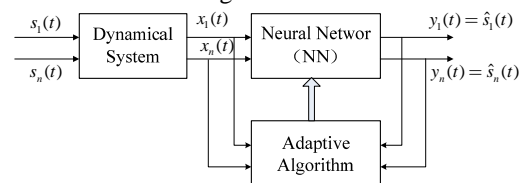

Figure2. Blind signal separation's neural network architecture

Assumed n is known, the separation matrix is $W(t)$, and its adjustment is controlled by the learning algorithm. Linear forward neural network NN with fixed structure is adopted as the separation system, and then the output of the neural network is:

$$y(t) = NN(W(t), x(t)) = W(t)x(t) \qquad (2)$$

Where $W(t)$ is $n \times m$'s separation (or mixing) matrix, in fact it is a linear NN's weight matrix. The key of the blind source separation is to use recursive method to find NN's weight matrix, so that the output $y(t)$ is source vectors' estimation.

II. DETERMINATION OF THE SOURCE NUMBER

Same with most other neural network blind source separation methods, the additional conditions we need is the sensor or array element's number must be greater than or equal to sources' actual number, i.e. $m \geq n$ and $m \leq 2n$, while the number of the output $n \leq l \leq m$. Therefore, the mixing matrix A is a "tall" matrix ($m \times n$), and the neural network weight matrix is a "wide" matrix ($l \times m$). So the separation model is:

$$y(t) = W_{l \times m}(t)x(t) \qquad (3)$$

The solution to the problem of the unknown number sources' blind separation, first proposed by Cichocki et al, using double-layer neural network structure[1]. One way is to use the decomposition method based on the observational data covariance matrix, AIC and MDL criteria etc. probabilistic statistical methods to estimate the number of sources, and to reduce the dimensionality of the data with a pre-whitening layer. The algorithm's disadvantage is that when the mixing matrix morbid or in the presence of a weak

signal, the separation effect will be bad. Likewise, if the compressed data in the separation layer, using $n \times n$ orthogonal matrix to reconstruct source signal, there will be the same problems. Another method is natural gradient algorithm using ordinary definite blind separation [2], the different is that separation matrix is no longer $n \times m$, but $m \times m$, so the separation model [1] is:

$$y(t) = W_{m \times m}(t)x(t) \qquad (4)$$

Therefore it doesn't need to know the number of source signals. We would illustrate this problem through the following lemmas and theorems.

Lemma 1: neural network $y_t = Wx_t$, W is a non-singular separation matrix, network's output mutual information $I(y_t, W)$ can be used as the cost function of the unknown source number BSS problem, if and only if $y_t = G\xi_t$, then the function has the minimum value. (G is $m \times m$ dimensional permutation matrix, $\xi_t \in \Omega$)

Theorem 1: If the mixing matrix A is full column rank ($m \geq n$), if and only if the separation matrix $W = G[(A^+), N(A^T) + \bar{W}^T]^T$, separation system's output mutual information will obtain the minimum value. A^+ is the pseudo-inverse of A, $N(A^T)$ is a $m \times (m-n)$ dimensional matrix consist of the $(m-n) \times m$ dimensional base vectors of the null space of A, $\bar{W}$ is a $(m-n) \times m$ dimensional matrix, constructed by rows of A^+'s arbitrary $m-n$ lines.

For a $m \times n$ dimensional nonsingular "high" mixing matrix A, it should be able to find a $l \times m$ "wide" separation matrix, making

$$W = \Lambda P A^+ \qquad (5)$$

Where, Λ is a nonsingular diagonal matrix ($l \times l$), $A^+ = (A^T A)^{-1} A^T$ is A's Moore-Penrose generalized inverse ($n \times m$), P is a generalized permutation matrix ($l \times n$), each of its row has a "1" element, each of its column has more "1" elements. Through formula (4), we can obtain:

$$y(t) = Wx(t) = \Lambda P A^+ \cdot As(t) = \Lambda Ps(t) \qquad (6)$$

So in the output, there are n separated source signals, and $l - n$ copies of it. Signals and its copies with different amplitude factor are coherent signals, in theory their correlation coefficient is equal to 1, therefore, ICA separation theorem needs to be amended.

ICA separation theorem: only when the number of output l equal to the number of sources n, the output vector would have minimum mutual information, and it's all components are independent; At the time $l > n$, while output vector has n independent components, other $l - n$ components would be coherent with some source signals.

Therefore, to get n input source signals, a related layer must be attached behind to the separation layer, after deletion of unwanted redundant components; residuals are the signals we want to extract. Shun-tian Lou et al through inspection of output components' coherence to detect and delete redundant signals which achieved a good separation effect.

III. DIVERGENCE ANALYSIS OF THE OVERDETERMINED BLIND SIGNAL SEPARATION ALGORITHM

Present in the overdetermined blind signal separation algorithm, especially when the number of output channels satisfy $n \leq l \leq m$, though the separation model we could achieve the object of signal separation, it eventually unavoidable divergence. Amari et al proved it and analyzed the reasons for the divergence: when the algorithm convergence, some (≥ 2) row vectors corresponding to the same sources' copies in the separation matrix W will influence each other. Through theorem 1, we know there exist a base vector difference in $Null(A^T)$ (mixing matrix A^T's null space) between these row vectors, so in the adaptive blind separation process, separation matrix will iterate in one equivalent class, define the equivalence class as follows:

$$C_w = \left\{ \tilde{W} \mid \tilde{W} = G[(A^+), N(A^T) + \bar{W}^T]^T \right\}$$

Where, G is a $m \times m$ permutation matrix, A^+ is A's pseudo-inverse, $N(A^T)$ represents a $m \times (m-n)$ dimensional matrix constructed by the base vectors of A^T's null space, $\bar{W}$ is a $(m-n) \times m$ dimensional matrix, it's row vectors are A^+'s arbitrary rows. This shows that even natural gradient algorithm (7) does not reach the balance point. Before divergence, separation system still able to separate signals. While with the iteration of W, if the learning rate $\eta(t)$ is not sufficiently small, makes the related row vector get accumulated, and cause the norm of separation matrix W tends to infinity, eventually led to the divergence of the algorithm.

IV. ADAPTIVE BLIND SEPARATION ALGORITHM WITH CHANGING SOURCE NUMBER

3.1 The derivation of the algorithm

In order to stabilize the overdetermined blind signal separation system, we adjust the natural gradient algorithm to

$$W(t+1) = W(t) + \eta(t)\left(R - \Phi[y(t)]y^T(t)\right)W(t) \qquad (7)$$

R is a $m \times m$ matrix. Obviously, when $R = I$ (unit matrix), algorithm (7) is equivalent to the natural gradient algorithm. As noted earlier, when $m > n$, the algorithm is not stably converging. Amari et al proposed the BBS nonhomogeneous orthogonal learning algorithm [4], defined R as a diagonal matrix.

$$R = diag\left\{ \varphi_1(y_1(t))y_1(t), \cdots, \varphi_1(y_m(t))y_m(t) \right\} \quad (8)$$

This orthogonal natural gradient algorithm can reduce the redundant components to zero, avoiding the shortcomings of divergence. it is especially effective when there are rapid or intermittent amplitude changes in the source signal. Its drawback is that when the separated sources' amplitude is small, it's hard to distinguish and delete the weak redundant signals. Based on these, Ye et al proposed a generalized orthogonal natural gradient algorithm [5].

$$R = E\left\{ \varphi(y(t))y(t) \right\} \quad (9)$$

Obviously, when $m > n$, R is no longer a diagonal matrix, and it is closely related with the output components $y(t)$. Reversely thinking: if the number of sources can be predicted, right choice of R can control the position of redundancy copies in the output channel according to the output components $y(t)$.

In order to make the output components intuitive, and make it easy to delete redundant components, considering the situation $n < m \le 2n$. Let the output $m - n$ copy signals appear in pairs with the source signal, and locate in the first $2(m - n)$ output channels, the rest source signals are located in $2n - m$ channels after, the output components can be expressed as:

$$y_i(t) = \begin{cases} s_i(t), i = 1, 3, \cdots, 2(m-n)+1, 2(m-n)+2, \cdots, m \\ s_j(t), j = 2, 4, \cdots, 2(m-n) \end{cases} \quad (10)$$

Substituting equation (10) into the stable condition of natural gradient algorithm:

$$E\left\{ I - \varphi(y(t))y^T(t) \right\} = 0 \quad (11)$$

And additional constraints due to the natural gradient algorithm [6]:

$$E\left\{ I - \varphi(y_i(t))y_i^T(t) \right\} = 0, i = 1, 2, \cdots, m \quad (12)$$

Specify the values of the separated state (10) in (11)

$$E\{I - \varphi(y(t))y^T(t)\} = -[I_2^1, I_1^2, I_4^3, I_3^4, \cdots, I_{2(m-n)}^{2(m-n)-1},$$
$$I_{2(m-n)-1}^{2(m-n)}, I_{2(m-n)+1}^{2(m-n)+1}, \cdots, I_m^m] = M \quad (13)$$

I_r^q represents the r-th row of the $m \times m$ unit matrix, q represents M's q-th column. Therefore get the learning rules:

$$E\left\{ \frac{dW}{dt} \right\} = -\eta_t[W_2^1, W_1^2, W_4^3, W_3^4, \cdots, W_{2(m-n)}^{2(m-n)-1},$$
$$W_{2(m-n)-1}^{2(m-n)}, W_{2(m-n)+1}^{2(m-n)+1}, \cdots, W_m^m] \quad (14)$$

W_r represents W's r-th column. From the above derivation, we can choose

$$R = I + M \quad (15)$$

Then get the new algorithm's model:

$$W(t+1) = W(t) + \eta(t)\left(R - \Phi[y(t)]y^T(t) \right)W(t) \quad (R = I + M)(16)$$

Update learning steps using the following formula[7]:

$$\eta(t) = \eta(t-1)$$
$$-\rho trace\left\{ f(t)X^T(t)W^T(t-1)\left[I - y(t-1)f^T(y(t-1)) \right] \right\} \quad (17)$$

In the above equation, $trace(\bullet)$ represents the trace of the matrix.

The advantage of time-varying learning step algorithm is: at the beginning of the adaptive algorithm, error $e(n)$ is relatively large; step size $\eta(t)$ is correspondingly large too, resulting in a fast convergence rate. When the error is gradually reduced, $\eta(t)$ becomes small, so the steady-state error is very small.

3.2 Determination of the source number

Now discuss how to determine the source number n, Give the following theorem:

Theorem 2: If the mixing matrix $A_{m \times n}$ is full column rank ($m \ge n$), $s = [s_1, \cdots, s_K]$ is a $n \times K$ full row rank matrix ($K \ge n$)constructed by n independent sources, and $X_{m \times K} = As$ is an observation vector excluding noise, then extend A^T and X^T to the same null space, and source number n equal to the rank of X.

From theorem 2, we know source number can be determined by the observation matrix composed of arbitrary length $K (K \ge n)$ observation data. Generally, take $K = m \sim 2m$.

From the above derivation, it can be known that as long as we follow theorem 2 to determine the source numbers, matrix R can be constructed by equation (15). The learning rule provided by the new algorithm (16) can separate m sources and $m - n$ copies well. Copies are located at the first $m - n$ even channels. Deletion can be done directly according to copy signals' location in the outputs, without based on a measure of correlation between the output components to identify and delete.

3.3 Dynamic changes of source number

In order to achieve the purpose of real-time determination of the source numbers, this paper takes first K observation data before current iteration position of the algorithm as the observation matrix, so matrix X could refresh itself constantly as a sliding window. According to the definition of rank, obviously, when suddenly increase a signal, so will be the $rank(X)$. Thus we could immediately determine current source number, reach the precise requirement of real-time: when a signal suddenly disappeared, $rank(X)$ would reduce after K iteration of the algorithm, but this effect can be ignored, and the algorithm still can meet the real-time requirement. Algorithm of specific steps is as follows:

(1) For the first K samples, learn according with the basic natural gradient algorithm.

(2) From the $K+1$-th sample, construct $m \times K$ observation matrix X, obtain $rank(X)$ to determine the source number.

(3)Construct matrix R through equation (15), learn as new algorithm (16).

(4)Delete the first $m-n$ even channels' redundant separation.

(5)If reduce the source number, separation matrix will not change, while increase the source number, $W = W + 0.1I_{m \times m}$;

(6) Repeat steps (2) to (5).

Note: the sudden reduction of sources actually is the dimension reduction transformation of the neural network, from a mathematical point of view, using the orthogonal projection method still can make neural network to restore balance state. The sudden increase of the source, make the neural network system happen an expansion-dimensional transform[3], to make the system to restore the equilibrium state is much more difficult than the dimensionality reduction transform. At present, the only effective way is to initialize system, In step (5), we add a perturbation term(the perturbation term can be adjusted) to the separation matrix, it's effect is equal to initialization.

V. SIMULATIONS

In order to verify if the adaptive natural gradient algorithm can effectively solve the problem of the radar signal sorting, we consider the separation of following random linear mixture of radar signals here:

S1:pulse $sign(\cos(2\pi \times 155t))$;

S2: high-frequency sinusoidal signal $\sin(2\pi \times 800t)$;

S3：low-frequency sinusoidal signal $\sin(2\pi \times 90t)$;

S4：phase modulation signal
$\sin((2\pi \times 300t) + 0.5\cos(2\pi \times 60t))$;

S5：amplitude modulation signal
$1 + 0.5\sin(2\pi \times 9t)\sin(2\pi \times 300t)$;

S6：In $[-1,1]$, Uniformly distributed random noise signal.

Simulation using nine receiving antenna ($m = 9, n = 6$), elements of the mixing matrix are random generated, and obey uniform distribution in the interval $[-1,1]$. The sampling rate of the received signal is 10kHz, i.e. using the period $T_s = 0.0001$ s. Adaptive gradient algorithm's learning step $\eta(t)$ takes 65, new adaptive gradient algorithm's learning step length takes $\eta(t)$ in equation (17).

A. Simulation 1: performance comparison between new adaptive gradient algorithm with adaptive gradient algorithm.

From figure 3 it can be seen that new adaptive gradient algorithm had gotten a good convergence effect when it reached to 2000 points, while adaptive gradient algorithm got a good convergence effect when it reached to 4000 points. New adaptive gradient algorithm has significantly faster convergence than the adaptive gradient algorithm.

Additionally, new adaptive gradient algorithm's error is lower than adaptive gradient algorithm. Therefore, the new adaptive gradient algorithm has a better performance than the adaptive gradient algorithm.
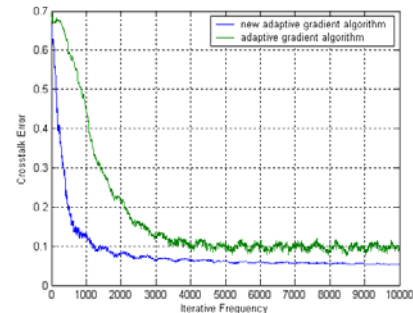

Figure3. Performance comparison between new adaptive gradient algorithm with adaptive gradient algorithm.

B. Simulation 2: separation effect validation of the new adaptive gradient algorithm when reduce source number randomly.

When $t < 0.5$ (i.e. before 5000 sampling points), there are 6 sources. When $t > 0.5$ (after 5000 sampling points), randomly reduce one source, it becomes to 5 sources. Figure 4 shows the variation waveform of the sources. figure 5 shows a separation of the random reduction of the source signal. It can be seen from figure 5, before 5000 sampling points, the algorithm separated 6 sources and 3 redundant signals well, and also output sources and it's copies in an established order: First six channels sequentially output three sources and its' copies, the last three channels output other three source signals. When $t > 0.5$, i.e. after 5000 sampling points, randomly reduce one source, the algorithm still can separate the five sources and 4 redundant signals well. To delete the redundant signals, one only needs to simply delete the specified channels' signals (i.e. 2,4,6,8 channel), without separating the sources through verifying the correlation between the output signals, which avoids the complexity of the removal process. Figure 6 is the error curve of randomly reducing a source. It can be seen from the diagram, when a source signal is reduced at the 5000-th sampling point, error is slightly increased but it quickly restored stability, and continued to maintain convergence performance of the algorithm.
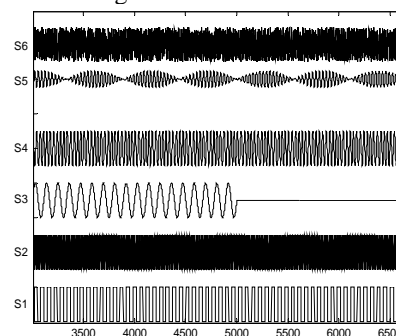
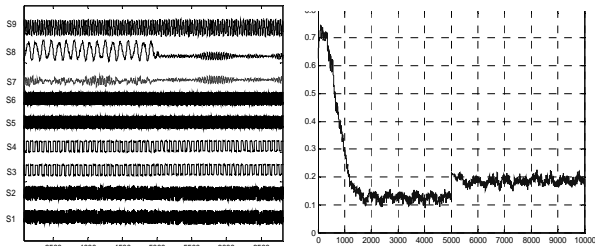

Figure4. Signals when sources reduced

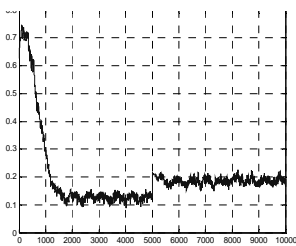Figure5. Separation process when randomly reduce one source

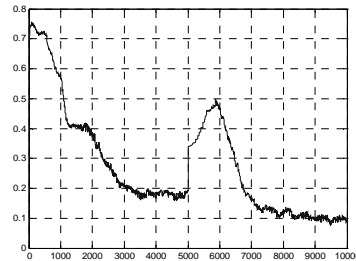Figure6. Error curve when randomly reduce one source

C. Simulation 3: separation effect validation of the new adaptive gradient algorithm when increase source number randomly.

When $t < 0.5$ (before 5000 sampling points), there are 5 sources. When $t > 0.5$ （after 5000 sampling points）, randomly increase one source, the source number become six. Figure 7 shows the variation waveform of the source, figure 8 shows a separation process when randomly increase the source number. It can be seen from figure 8,before 5000 sampling points, the algorithm well separate 5 sources and 4 redundant signals, and also output sources and it's copies in an established order: first eight channels sequentially output four sources and its' copies, the last channel output the residual. When $t > 0.5$, i.e. after 5000 sampling points, randomly increase one source, the algorithm still can separate the six sources and three redundant signals well, in which the first six channels sequentially output three source signals and its' copies, the last three channels output other three source signals. Figure 9 is the error curve of randomly increasing one source. It can be seen from the diagram, when a source is added at the 5000-th sampling point, error is slightly increased but the algorithm quickly restored stability, and continued to maintain convergence performance.
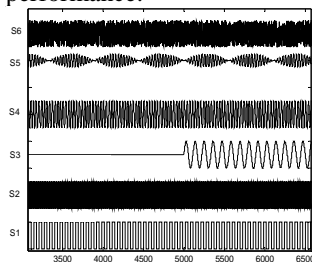

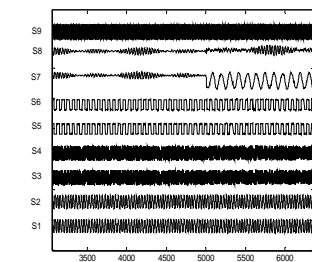Figure7. Signals when sources increased

Figure 8. Separation process when randomly increase one source

Figure9. Error curve when randomly increase one source

VI. CONCLUSIONS

Blind source separation method has been paid increasing attentions as a new research, its coverage is becoming bigger and bigger, and its application is becoming more and more practical. Blind source separation algorithm with the condition of unknown source number and dynamically changing, is more general, and has more practical value in practical applications. Radar signal sorting research based on blind source separation algorithm is the expansion and application of this field, the value of this research is obvious. This paper is a research on blind source separation algorithm at the overdetermined condition and its application in sorting radar signals, which provides a new way in radar signal processing.

REFERENCES

[1] Cichocki A,Karhunen J,Kasprzak W,et al.Neural Networks for blind separation with unkown number of sources[J],Neurocomputing. 1999,24(1).pp:55-93.

[2] Amari S. Nutural gradient works efficiently in learning[J]. Neural Computation. 1998, Vol.10, No.58.pp:251-276.

[3] Shun-tian Lou and Xian-da Zhang. Blind source Separation for Changing Source Number: A Neural Network Approach with a Variable Structure, ICA2001,December 9-12,San Diego,California, USA.

[4] Amari S. Natural gradient learning for over-and under-complete bases in ICA[J].Neural Computation.1999,11(8).pp:1875-1883.

[5] Ye J M, Zhu X L, Zhang X D. Adaptive blind separation with an unknown number of sources[J].Neural Computation.2004, 16(8) .pp: 1641-1660.

[6] Amari S.,Chen T.P, Cichocki A. Nonholonomic orthogonal learning algorithm for blind source separation[J]. Neural Computation.2000, 12(6).pp:1463-1484.

[7] Cichcki A, Amari S. Adaptive Blind Signal and Image Processing: learning Algorithm and Applications[M]. Wiley, New York,2002.