A Spurious Equilibria-free Learning Algorithm for the Blind Separation of Non-zero Skewness Signals

Seungjin CHOI*[1], Ruey-wen LIU**, and Andrzej CICHOCKI***

* Department of Electrical Engineering
Chungbuk National University
48 Kaeshin-Dong Cheongju
Chungbuk 361-763, KOREA

** Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556
U.S.A.

*** Laboratory for Artificial Brain Systems
Frontier Research Program, RIKEN
2-1 Hirosawa, Wako-shi
Saitama 351-01, Japan

Abstract: We present a new learning algorithm for the blind separation of independent source signals having non-zero skewness (the 3rd-order cumulant) (the source signals have non-symmetric probability distribution.), from their linear mixtures. It is shown that for a class of source signals whose probability distribution functions is not symmetric, a simple adaptive learning algorithm using quadratic function ($f(x) = x^2$) is very efficient for blind source separation task. It is proved that all stable equilibria of the proposed learning algorithm are desirable solutions. Extensive computer simulation experiments confirmed the validity of the proposed algorithm.

Key Words: blind source separation, higher-order statistics, neural networks, unsupervised learning.

This paper has not been submitted elsewhere in identical or similar form, nor will it be during the first three months after its submission to *Neural Processing Letters*

[1]Please address correspondence to Seungjin Choi, School of Electrical and Electronics Engineering, Chungbuk National University, 48 Kaeshin-Dong, Cheongju, Chungbuk 361-763, KOREA, Tel: +82-431-61-2421, Fax: +82-431-63-2419, Email: schoi@engine.chungbuk.ac.kr

1 Introduction

The problem of *blind source separation* is to recover the independent source signals from their linear mixtures without the knowledge of the mixing matrix. This is a fundamental problem motivated by many potential applications. For instance, the narrow-band signals received by antenna sensors in communication systems are often a linear transformation of statistically independent source signals, and it is desirable to restore these source signals without the knowledge of the linear transformation.

Neural computational approach to blind source separation was first introduced by Jutten and Herault [6], and further developed by others [7, 5, 3, 1, 2, 8, 4]. Most of the algorithms employ the nonlinear odd functions (for example, $f(x) = x^3$ or $f(x) = \tanh(x)$) because they assume that the probability distribution of all source signals are symmetric. Surprisingly, it is shown here that for a class of source signals whose probability distribution function is not symmetric, a simple learning algorithm using quadratic function ($f(x) = x^2$) is very efficient for blind source separation task. Specifically, if the source signals are independent and each of them has a non-zero skewness (the 3rd-order cumulant), then the source signals are separated by a linear transformation, if and only if all the 2nd- and 3rd-order *cross*-cumulants of the output are zero. Note that it does not require the 3rd-order cumulants among three different variables to be zero. In this paper, we present a linear feedforward network having auxiliary lateral feedback connections with an associated on-line learning algorithm. We prove that all stable equilibria of the proposed learning algorithm are desirable solutions to blind source separation.

2 Neural Network Model and Learning Algorithm

Consider the case where the n dimensional observation vector $\mathbf{x}(t)$ and the n dimensional source vector $\mathbf{s}(t)$ are related by

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \tag{1}$$

where $\mathbf{A}$ is an $n \times n$ mixing matrix. The problem of blind source separation is to recover the source signals $\mathbf{s}(t)$ from the observation vector $\mathbf{x}(t)$ without the knowledge of the mixing matrix $\mathbf{A}$. In other words, it is required to find a linear transformation

$\mathbf{W}(t)$, i.e.,

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

such that the composite matrix $\mathbf{G}(t) = \mathbf{W}(t)\mathbf{A}$ converges as $t \to \infty$ to a matrix

$$\mathbf{G} = \mathbf{P}\mathbf{\Lambda}, \tag{3}$$

for some permutation matrix $\mathbf{P}$ and nonsingular diagonal matrix $\mathbf{\Lambda}$.

Throughout this paper, the following assumptions hold:

A1: $\mathbf{A} \in \mathbb{R}^{n\times n}$ is nonsingular.

A2: At each time, the components of $\mathbf{s}(t)$ are statistically independent.

A3: Each component of $\mathbf{s}(t)$ is a zero mean ergodic stationary process with a non-zero variance.

A4: Each component of $\mathbf{s}(t)$ has a non-zero skewness, i.e.,

$$E\{s_i^3(t)\} \neq 0 \quad \text{for } i = 1, \ldots, n, \tag{4}$$

where E denotes expectation operator.

Consider a linear feedforward network with auxiliary lateral feedback connections (see Figure 1) described by

$$y_i(t) = \sum_{j=1}^{n} w_{ij}(t)x_j(t) - \sum_{j<i} u_{ij}(t)y_j(t), \quad \text{for } i = 1, \cdots, n, \tag{5}$$

where $w_{ij}(t)$ is a synaptic weight between the ith output $y_i(t)$ and the jth input $x_j(t)$. $u_{ij}(t)$ is a lateral feedback connection from the jth output $y_j(t)$ to the ith output $y_i(t)$. Or in compact matrix form,

$$\mathbf{y}(t) = \mathbf{W}(t)\mathbf{x}(t) - \mathbf{U}(t)\mathbf{y}(t), \tag{6}$$

where $\mathbf{W}(t) = [w_{ij}(t)]_{n\times n}$ is a full matrix and $\mathbf{U} = [u_{ij}(t)]_{n\times n}$ is a lower triangular matrix with $u_{ij}(t) = 0$ for $i \leq j$.

For such a neural network, we have developed the following adaptive learning algorithm:

$$\begin{aligned} \frac{d\mathbf{W}(t)}{dt} &= \eta(t)\{\mathbf{\Lambda}(t) - \mathbf{y}(t)\mathbf{y}^T(t) + \mathbf{y}(t)[\mathbf{y}(t)\circ\mathbf{y}(t)]^T \\ &\quad -[\mathbf{y}(t)\circ\mathbf{y}(t)]\mathbf{y}^T(t)\}\mathbf{W}(t), && (7) \\ \frac{du_{ij}(t)}{dt} &= \gamma(t)\{y_i(t)y_j^2(t)\}, \quad \text{for } i > j, && (8) \end{aligned}$$

where $\mathbf{y}(t) \circ \mathbf{y}(t) = [y_1^2(t), \ldots, y_n^2(t)]^T$ ($\circ$ denotes Hadamard product.), and $\eta(t) > 0$, $\gamma(t) > 0$ are learning rates. We can choose $\mathbf{\Lambda}(t)$ as invertible diagonal matrix (for example, identity matrix) or $\mathbf{\Lambda}(t) = diag\{\mathbf{y}(t)\mathbf{y}^T(t)\}$. Note that the learning algorithm (7) for feedforward connections $\mathbf{W}(t)$ is similar to EASI algorithm [4] developed by Cardoso and Laheld but we employ a quadratic function while EASI adopted a cubic function. Moreover, lateral connections are incorporated in order to introduced to force all cross 3rd-order cumulants of $\mathbf{y}(t)$ vanish to zero. It can be easily shown that all stable equilibria of (7) and (8) satisfy the following conditions:

$$\begin{aligned} E\{\mathbf{y}(t)\mathbf{y}^T(t)\} &= \mathbf{\Lambda}_1, \\ E\{\mathbf{y}(t)[\mathbf{y}(t) \circ \mathbf{y}(t)]^T\} &= \mathbf{\Lambda}_2, \end{aligned} \tag{9}$$

where $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are some nonsingular diagonal matrices. This means that all 2nd- and 3rd-order cross-cumulants of the output $\mathbf{y}(t)$ are zero, when the network achieves the convergence.

3 Analysis of the Algorithm

In this section, we show that all stable equilibria of (7) and (8) are desirable solutions to blind source separation. Without loss of generality, we can assume that source signals have unit variance, i.e., $\mathbf{R}_{ss} = E\{\mathbf{s}(t)\mathbf{s}^T(t)\} = \mathbf{I}$. Let $\mathbf{K}_s$ be a diagonal matrix whose ith diagonal element is $E\{s_i^3(t)\}$. Let $\mathbf{y}(t) = [\mathbf{I} + \mathbf{U}]^{-1}\mathbf{W}\mathbf{A}\mathbf{s}(t) = \mathbf{G}\mathbf{s}(t)$. Define the 3rd-order cumulant matrix of $\mathbf{y}(t)$ as follows:

$$[\mathbf{C}_{12y}]_{ij} = cum\{y_i(t), y_j(t), y_j(t)\}, \tag{10}$$

where $[\cdot]_{ij}$ denotes the (i,j)th element of the matrix $[\cdot]$. Then it can be shown that the 3rd-order cumulant matrix has the following decomposition:

$$\begin{aligned} \mathbf{C}_{12y} &= E\{\mathbf{y}(t)[\mathbf{y}(t) \circ \mathbf{y}(t)]^T\}, \\ &= \mathbf{G}\mathbf{K}_s[\mathbf{G} \circ \mathbf{G}]^T. \end{aligned} \tag{11}$$

Lemma 1 *Let $\mathbf{Q} \in \mathbb{R}^{n\times n}$ be an orthogonal matrix. Let $\mathbf{\Sigma}_1 \in \mathbb{R}^{n\times n}$ and $\mathbf{\Sigma}_2 \in \mathbb{R}^{n\times n}$ be nonsingular diagonal matrices. If $\mathbf{\Sigma}_1(\mathbf{Q} \circ \mathbf{Q}) = \mathbf{Q}\mathbf{\Sigma}_2$, then $\mathbf{Q}$ has the decomposition, $\mathbf{Q} = \mathbf{P}\overset{\circ}{\mathbf{I}}$, where $\mathbf{P}$ is some permutation matrix and $\overset{\circ}{\mathbf{I}}$ is the diagonal matrix whose diagonal entries are either +1 or -1.*

Proof of Lemma 1:

$$\mathbf{\Sigma}_1(\mathbf{Q} \circ \mathbf{Q}) = \mathbf{Q}\mathbf{\Sigma}_2, \tag{12}$$

where $\mathbf{\Sigma}_1 = diag\{\alpha_1, \cdots, \alpha_n\}$ and $\mathbf{\Sigma}_2 = diag\{\beta_1, \cdots, \beta_n\}$ Then, from (12), the ijth element of the matrix $\mathbf{Q}$, q_{ij} should satisfy

$$\alpha_i q_{ij}^2 = \beta_j q_{ij}, \quad \text{for } i, j = 1, \ldots, n, \tag{13}$$

i.e., $q_{ij} = \frac{\beta_j}{\alpha_i}$ or $q_{ij} = 0$. We need to show that exactly one entry in each row and column of $\mathbf{Q}$ is equal to either +1 or -1. Since $\mathbf{Q}$ is orthogonal, each row and column has at least one non-zero element. Without loss of generality, let $q_{11} = \frac{\beta_1}{\alpha_1}$ be non-zero. Since $\mathbf{Q}$ is an orthogonal matrix, the inner product of the 1st row and the ith row is zero, i.e.,

$$\frac{1}{(\alpha_1 \alpha_i)} \sum_{k \in \mathcal{K}} \beta_k^2 = 0, \tag{14}$$

where $\mathcal{K}$ is the set of column k in which both $q_{1k} = \frac{\beta_k}{\alpha_1}$ and $q_{ik} = \frac{\beta_k}{\alpha_i}$ are non-zero. Equation (14) is possible only if $\mathcal{K}$ is an empty set. Hence there is not a column k in which both q_{1k} and q_{ik} are non-zero. This concludes that $q_{i1} = 0$, because $q_{11} \neq 0$. Since this is true for any $i \neq 1$, there is only one non-zero element in the 1st column of $\mathbf{Q}$. Similarly, this is true for every column of $\mathbf{Q}$. Now, using the same argument for different rows, one can show that there is only one non-zero element in each row of $\mathbf{Q}$. Finally the non-zero element in each row and each column of $\mathbf{Q}$ is equal to either +1 or -1 because $\mathbf{Q}$ is orthogonal. Therefore, $\mathbf{Q} = \mathbf{P}\,\mathring{\mathbf{I}}$. □

Theorem 1 (Main Theorem) *Let the sources* $\mathbf{s}(t)$ *satisfy the assumptions given in A1 through A4. Then,* $\mathbf{G}$ *has decomposition (3), i.e.,*

$$\mathbf{G} = \mathbf{P}\mathbf{\Lambda},$$

if and only if the following conditions are satisfied:

$$\mathbf{G}\mathbf{G}^T = \mathbf{\Lambda}_1, \tag{15}$$

$$\mathbf{G}\mathbf{K}_s(\mathbf{G} \circ \mathbf{G})^T = \mathbf{\Lambda}_2, \tag{16}$$

where $\mathbf{\Lambda}_1$ *and* $\mathbf{\Sigma}_2$ *are nonsingular diagonal matrices.*

Proof of Main Theorem:

i) $\Rightarrow$: It is straightforward.

ii) $\Leftarrow$: We can rewrite the equation (15) as

$$\mathbf{G}\mathbf{G}^T = (\mathbf{\Lambda}_1^{\frac{1}{2}})\mathbf{Q}\mathbf{Q}^T(\mathbf{\Lambda}_1^{\frac{1}{2}})^T. \tag{17}$$

where $\mathbf{Q}$ is any orthogonal matrix. Hence,

$$\mathbf{G} = \mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{Q}. \tag{18}$$

Substitute (18) into (16) to obtain

$$\mathbf{Q}\mathbf{K}_s(\mathbf{Q} \circ \mathbf{Q})^T = \mathbf{\Lambda}_1^{-\frac{3}{2}}\mathbf{\Lambda}_2. \tag{19}$$

Here, we have used the relations, $(\mathbf{\Lambda}\mathbf{Q}) \circ (\mathbf{\Lambda}\mathbf{Q}) = \mathbf{\Lambda}^2(\mathbf{Q} \circ \mathbf{Q})$ and $(\mathbf{Q}\mathbf{\Lambda}) \circ (\mathbf{Q}\mathbf{\Lambda}) = (\mathbf{Q} \circ \mathbf{Q})\mathbf{\Lambda}^2$. Pre-multiply (19) by $\mathbf{Q}$ to obtain

$$\mathbf{K}_s(\mathbf{Q} \circ \mathbf{Q})^T = \mathbf{Q}^T\mathbf{\Lambda}_1^{-\frac{3}{2}}\mathbf{\Lambda}_2 \tag{20}$$

Since $\mathbf{K}_s$ and $\mathbf{\Lambda}_1^{-\frac{3}{2}}\mathbf{\Lambda}_2$ are nonsingular diagonal matrices and $(\mathbf{Q} \circ \mathbf{Q})^T = \mathbf{Q}^T \circ \mathbf{Q}^T$, from Lemma 1, $\mathbf{Q} = \mathbf{P}\,\mathring{\mathbf{I}}$. Thus, (18) can be written as

$$\mathbf{G} = \mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{P}\,\mathring{\mathbf{I}}\,. \tag{21}$$

Or,

$$\begin{aligned} \mathbf{G} &= \mathbf{P}\mathbf{P}^T\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{P}\,\mathring{\mathbf{I}} \\ &= \mathbf{P}\mathbf{\Lambda}, \end{aligned} \tag{22}$$

where $\mathbf{\Lambda} = \mathbf{P}^T\mathbf{\Lambda}_1^{\frac{1}{2}}\mathbf{P}\,\mathring{\mathbf{I}}$ is a diagonal matrix. □

4 Computer Simulations

The computer simulations are conducted to evaluate the performance of the proposed learning algorithm (7) and (8). At each computer simulations, the composite system, $\mathbf{G} = (\mathbf{I}+\mathbf{U})^{-1}\mathbf{W}\mathbf{A}$, is evaluated. The $\mathbf{G}$ should be a generalized permutation matrix (a permutation matrix post-multiplied by a invertible diagonal matrix) at desirable equilibria.

4.1 Simulation 1

Three different sources are drawn from the same one-sided exponential distribution with unit variance and zero mean. The mixing matrix A is chosen randomly as

$$A = \begin{bmatrix} 1.0000 & 0.5413 & 0.9530 \\ 0.8685 & 0.4672 & 0.3211 \\ 0.2444 & 0.7810 & 0.3207 \end{bmatrix}. \tag{23}$$

The exemplary convergence of synaptic weights $w_{ij}(t)$ and $u_{ij}(t)$ are shown in Figure 2. The learning algorithm converged to desirable equilibria around 4000 iterations. The composite system $\mathbf{G}$ at steady state is

$$\mathbf{G} = \begin{bmatrix} 0.0524 & 0.0099 & -1.0049 \\ 0.0549 & -0.9727 & -0.0365 \\ -1.0399 & -0.0456 & -0.0614 \end{bmatrix}. \tag{24}$$

It can be observed that $\mathbf{G}$ is very close to a generalized permutation matrix.

4.2 Simulation 2: Test with Speech Signals

The biggest drawback of the criterion proposed in this paper is that it requires non-zero skewness of the source signals. Evidently the signals with symmetric distribution have zero skewness. In this case our learning algorithms can not be applied. However, there arises a natural question that what will be the minimum bound of skewness where the algorithm is working. We do not have theoretical answer for this question. However, this simulation shows very promising result. Two digitized speech signals and one random noise sequence drawn from one-sided exponential distribution, are mixed by the mixing matrix used in Simulation 1. The skewness of two different digitized speech signals are 0.04, 1.05, respectively. The skewness of the noise signal drawn from one-sided exponential distribution is 1.97. The source signals, $\mathbf{s}(t)$, the received signals at each sensor, $\mathbf{x}(t)$, the recovered signals, $\mathbf{y}(t)$ by using two different learning algorithms are shown in Figure 3, 4, and 5, respectively. In this simulation, it is observed that the signals can be recovered when the skewness of source signals are small to a certain degree (about 10^{-2}).

5 Conclusion

In this paper, a new simple learning algorithm for blind separation of source signals whose probability distribution is not symmetric, was presented. It was demonstrated that a simple learning algorithm using quadratic function was very efficient in this task. It was proved that the proposed learning algorithm does not have any spurious equilibria. This fact was confirmed by extensive computer simulation experiments

References

[1] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, Vol. 8, 1995.

[2] S. Amari, A. Cichocki, and H. H. Yang, "Recurrent neural networks for blind separation of sources," *Proceedings of International Symposium on Nonlinear Theory and Its Applications (NOLTA 95)*, Vol. 1, pp. 37-42, 1995.

[3] A. Bell and T. Sejnowski, "An information maximisation approach to blind separation and blind deconvolution," *Neural Computation* 7, pp. 1129-1159, 1995.

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

[5] A. Cichocki, R. Unbehauen, L. Moszczynski, and E. Rummert, "A new on-line adaptive learning algorithm for blind separation of source signals," *Int. Symp. on Artificial Neural Networks*, 1994.

[6] C. Jutten, J. Herault, and A. Guerin, "IN.C.A.: An independent component analyser based on an adaptive neuromimetic network," J. Demongeot, T. Hervé, V. Rialle and C. Roche, eds., *Artificial Intelligence and Cognitive Sciences* Manchester Univ. Press, Manchester, 1988.

[7] E. Moreau and O. Macchi, "Two novel architectures for the self adaptive separation of signals," *IEEE Conf. Comm.*, pp. 1154-1159, 1993.

[8] E. Oja, J. Karhunen, and A.Hyvarinen, "From Neural PCA to Neural ICA," *Neural Information Processing Systems Workshop on Blind Signal Processing and its Applications,* December 1996.

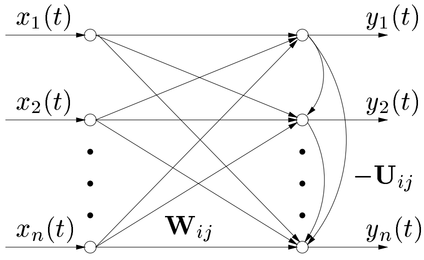
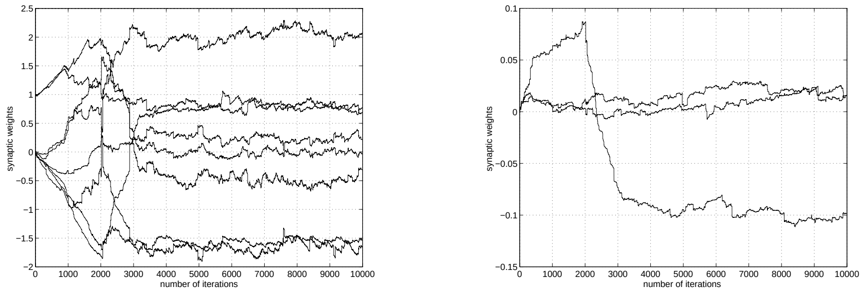
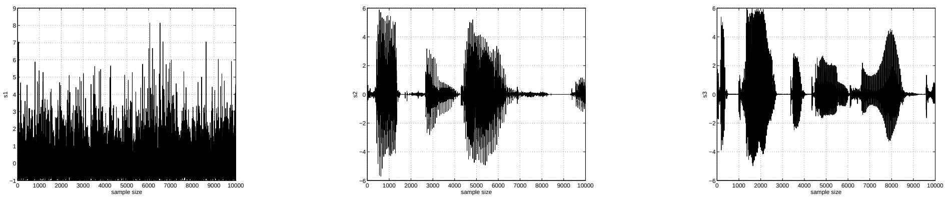


Figure 1: A linear feedforward network with lateral feedback connections



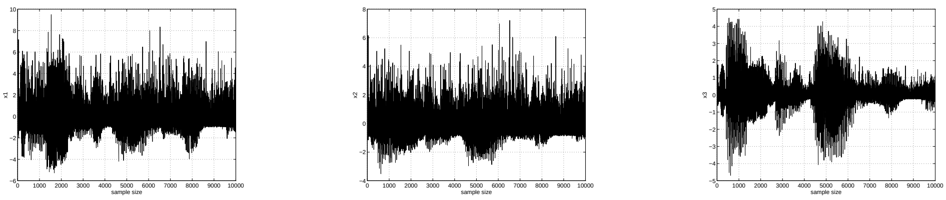
(a) (b)

Figure 2: The network is trained by the proposed learning algorithm (7), (8) with constant learning rates, $\eta(t) = .001$ and $\gamma(t) = .0001$: (a) The convergence of parameters $w_{ij}(t)$; (b) The convergence of parameters $u_{ij}(t)$



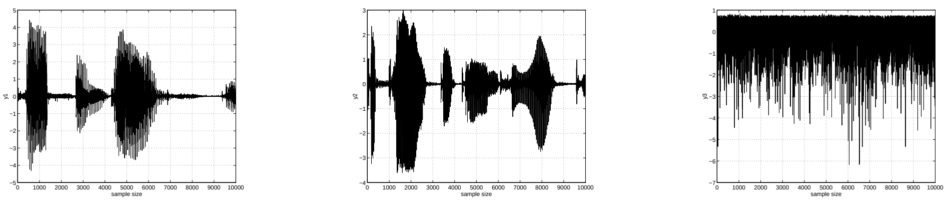
(a) (b) (c)

Figure 3: The source signals: (a) the first source signal, its skewness is 1.97; (b) the second source signal, its skewness is 0.04; (c) the third source signal, its skewness is 1.05.



(a) (b) (c)

Figure 4: The received signals at each sensor: (a) the first sensor output; (b) the second sensor output; (c) the third sensor output



(a) (b) (c)

Figure 5: The recovered signals: (a) the recovered signal 1; (b) the recovered signal 2; (c) the recovered signal 3