Convergence Analysis of a Class of Hyvärinen–Oja's ICA Learning Algorithms With Constant Learning Rates

Jian Cheng Lv, Kok Kiong Tan, Zhang Yi, and Sunan Huang

***Abstract*—The convergence of a class of Hyvärinen–Oja's independent component analysis (ICA) learning algorithms with constant learning rates is investigated by analyzing the original stochastic discrete time (SDT) algorithms and the corresponding deterministic discrete time (DDT) algorithms. Most existing learning rates for ICA learning algorithms are required to approach zero as the learning step increases. However, this is not a reasonable requirement to impose in many practical applications. Constant learning rates overcome the shortcoming. On the other hand, the original algorithms, described by the SDT algorithms, are studied directly. Invariant sets of these algorithms are obtained so that the nondivergence of the algorithms is guaranteed in stochastic environment. In the invariant sets, the local convergence of the original algorithms is analyzed by indirectly studying the convergence of the corresponding DDT algorithms. It is rigorously proven that the trajectories of the DDT algorithms starting from the invariant sets will converge to an independent component direction with a positive kurtosis or a negative kurtosis. The convergence results can shed some light on the dynamical behaviors of the original SDT algorithms. Furthermore, the corresponding DDT algorithms are extended to the block versions of the original SDT algorithms. The block algorithms not only establish a relationship between the SDT algorithms and the corresponding DDT algorithms, but also can get a good convergence speed and accuracy in practice. Simulation examples are carried out to illustrate the theory derived.**

***Index Terms*—Deterministic discrete time (DDT) algorithm, independent component analysis (ICA), invariant set, learning algorithm, stochastic discrete time (SDT) algorithm.**

I. Introduction

INDEPENDENT component analysis (ICA) [10] can estimate the independent components from the mixed input signal. Its main applications are in blind source separation [5], [11], [17], [20], [21], feature extraction [2], [18], [31], and blind deconvolution [12], [32], [34]. ICA neural networks enable fast adaptation in a stochastic environment since the inputs can be used in their algorithms at once [13]. These networks have been extensively studied and used in many fields, see examples [1], [4], [6], [19], [22], [28], [35], and [37]–[40].

The learning algorithms of ICA neural networks play an important role in their practical applications. The convergence of the learning algorithms determines whether these applications can be successful or not. However, it is very difficult to directly study the convergence of these learning algorithms, because they are described by stochastic discrete time (SDT) algorithms [47]. Traditionally, based on some stochastic approximation theorems [3], [23], the original SDT algorithms are transformed into the corresponding deterministic continuous time (DCT) algorithms. The convergence of the SDT algorithms can be indirectly interpreted by studying the corresponding DCT algorithms. This approach is called the DCT method. The method has been widely used; see, for example, [8], [29], [30], [33], [36], [41], [43], and [44]. In order to transform a SDT algorithm to a DCT algorithm, some restrictive conditions must be applied. One critical condition is that the learning rates must converge to zero. However, this condition is unrealistic in practical applications because of computational round-off limitations and tracking requirements [42], [47]. On the other hand, learning rates approaching zero will slow down the learning procedure tremendously. Furthermore, the convergence of these algorithms critically depends on a good choice of the learning rate sequence. A poor choice of the learning rate can destroy convergence in practice [13]. More discussions about these problems can be found in [9], [24]–[26], [42], [47], and [48].

To overcome the shortcomings mentioned above, the deterministic discrete time (DDT) method has been proposed recently to indirectly study the convergence of SDT learning algorithms; see [42], [47], and [48]. The DDT method transforms the SDT algorithms into the corresponding DDT algorithms. The DDT method does not require the learning rates to converge to zero so that constant learning rates can be used. The DDT algorithms preserve the discrete time nature of the original algorithms and allow a more realistic dynamic behavior of the learning rates [48]. Compared to the DCT method, the DDT method is a more reasonable approach for indirectly studying the convergence of SDT algorithms. However, the DDT algorithms characterizes the average evolution of the original SDT algorithms. In stochastic environment, the stochastic input has a great influence on the result of DDT method, even derailing the convergent conditions of the DDT algorithms.

Manuscript received March 14, 2008; accepted November 13, 2008. First published January 23, 2009; current version published April 15, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Danilo P. Mandic.

J. C. Lv is with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610054, China, and also with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576 (e-mail: lvjiancheng@scu.edu.cn; eleljc@nus.edu.sg).

K. K. Tan and S. Huang are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576 (e-mail: kktan@nus.edu.sg; elehsn@nus.edu.sg).

Z. Yi is with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu 610054, China (e-mail: zhangyi@scu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2013887

In this paper, we study the convergence of a class of Hyvärinen–Oja's ICA learning algorithms with constant learning rates. The original algorithms, described by the SDT algorithms, are directly analyzed first. Some invariant sets are identified under certain conditions so that the nondivergence of the original algorithms is guaranteed in stochastic environment. Then, by studying the corresponding DDT algorithms in the invariant sets, it is proven that the trajectory of the algorithms starting from the invariant sets will converge to an independent component direction with a positive kurtosis or a negative kurtosis. On the one hand, the method overcomes the shortcoming of only using the DCT method or DDT method. The convergence results of DDT algorithms are guaranteed in stochastic environment. On the other hand, due to the constant learning rates used, the requirements for use in practice can be met and convergent procedure will not be slowed down tremendously.

To improve the performance of the original SDT algorithms, the corresponding DDT algorithms are extended to the block versions of the original algorithms. The block algorithms not only establish a relationship between the original SDT algorithms and the corresponding DDT algorithms, but also get a good convergence speed and accuracy. The block algorithms represent an average evolution of the original SDT algorithms over all blocks of samples. Thus, the invariant sets of the original SDT algorithms can guarantee the nondivergence of the block algorithms. Furthermore, the block algorithms can exploit more efficiently the information contained in the observed sample window. Therefore, these algorithms converge fast and with higher accuracy [45], [46].

The rest of this paper is organized as follows. Section II provides the problem formulation and gives preliminaries. Convergence results are given in Section III. The DDT algorithms are extended to the block algorithms in Section IV. Simulation results and discussions are presented in Section V. Conclusions are drawn in Section VI.

II. Preliminaries and Hyvärinen–Oja's Algorithms

Suppose a sequence of observations $\{\mathbf{y}(k), k = 1, 2, \cdots\}$ with m scalar random variables, i.e., $\mathbf{y}(k) = [y_1(k), y_2(k), \cdots, y_m(k)]^T$. In the simplest of ICA [10], it is assumed that

$$\mathbf{y}(k) = A\mathbf{s}(k),$$

where $\mathbf{s}(k) = [s_1(k), s_2(k), \cdots, s_n(k)]^T$ is unknown and $s_1(k), s_2(k), \cdots, s_n(k)$ are mutually and statistically independent with zero mean and unit variance. A is an unknown $m \times n$ mixing matrix of full rank. The basic problem of ICA is then to estimate the original $s_i(k)$ from the mixtures $y(k)$. The problem can be simplified by prewhitening of the data $\mathbf{y}(k)$. The observed data $\mathbf{y}$ is linearly transformed to a vector $\mathbf{x} = M\mathbf{y}$ such that its elements x_i are mutually uncorrelated and all have unit variance, i.e., $E\{\mathbf{x}(k)\mathbf{x}^T(k)\} = I$. It follows that

$$\mathbf{x}(k) = M\mathbf{y}(k) = MA\mathbf{s}(k) = B\mathbf{s}(k)$$

where B is an orthogonal matrix. Clearly, $E\{\mathbf{x}(k)\mathbf{x}^T(k)\} = BE\{\mathbf{s}(k)\mathbf{s}^T(k)\}B^T = BB^T = I$, where $B = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_n)$, $s_i(k) = \mathbf{b}_i^T x(k)$. Then, the problem of ICA is transformed to one of estimating the $\mathbf{b}_i$.

The kurtosis of the signals is usually used to solve the ICA problem [13]–[16]. As for a zero mean random variable v, the kurtosis is defined as $\mathrm{Kurt}(v) = E\{v^4\} - 3(E\{v^2\})^2$. It is easy to show that $\mathrm{Kurt}(v_1 + v_2) = \mathrm{Kurt}(v_1) + \mathrm{Kurt}(v_2)$ and $\mathrm{Kurt}(\alpha v_1) = \alpha^4 \mathrm{Kurt}(v_1)$, where v_1 and v_2 are two independent, zero mean random variables and α is a scalar.

For the signals $s_1(k), s_2(k), \cdots, s_n(k)$, denote by $\mathrm{Kurt}(s_i)$ the kurtosis of the signal $s_i(k)$. Clearly, $\mathrm{Kurt}(s_i) = E\{s_i^4(k)\} - 3$. Define some sets as follows:

$$\begin{aligned} P &= \{i | \mathrm{Kurt}(s_i) < 0, i \in \{1, 2, \cdots, n\}\}, \\ Q &= \{i | \mathrm{Kurt}(s_i) > 0, i \in \{1, 2, \cdots, n\}\}, \\ O &= \{i | \mathrm{Kurt}(s_i) = 0, i \in \{1, 2, \cdots, n\}\}, \\ T &= O \cup P \cup Q. \end{aligned}$$

Let V_q be the subspace spanned by $\{\mathbf{b}_i | \mathrm{Kurt}(s_i) > 0\}$, i.e., $V_q = \mathrm{span}\{\mathbf{b}_i | \mathrm{Kurt}(s_i) > 0\}$. The subspace $V_q^\perp$ is perpendicular to V_q. Let $V_p = \mathrm{span}\{\mathbf{b}_i | \mathrm{Kurt}(s_i) < 0\}$ and $V_p^\perp$ is perpendicular to V_p.

To estimate the direction of $\mathbf{b}_i$, in [16], Hyvärinen and Oja proposed the ICA learning algorithms as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)\left[\sigma \mathbf{x}(k) f\left(\mathbf{w}^T(k)\mathbf{x}(k)\right) + a\left(1 - \|\mathbf{w}(k)\|^2\right)\mathbf{w}(k)\right]$$

for all $k \geq 0$ and $a > 0$, where f is a scalar function, and $\sigma = \pm 1$ is a sign that determines whether we are minimizing or maximizing the kurtosis. $\mu(k)$ is the learning rate sequence and $\mathbf{x}(k)$ is a prewhitened input data. In this paper, let $f(y) = y^3$ and the constant learning rates are used. the algorithms with constant learning rates can be presented as

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \eta\left[b\mathbf{x}(k)\left(\mathbf{w}^T(k)\mathbf{x}(k)\right)^3 + a\left(1 - \|\mathbf{w}(k)\|^2\right)\mathbf{w}(k)\right] \quad (1)$$

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \eta\left[-b\mathbf{x}(k)\left(\mathbf{w}^T(k)\mathbf{x}(k)\right)^3 + a\left(1 - \|\mathbf{w}(k)\|^2\right)\mathbf{w}(k)\right] \quad (2)$$

for all $k \geq 0$, where η is a constant learning rate. A constant $b \geq 0$ is added to the algorithms to adjust the learning parameters more conveniently.

The algorithm (1) converges to an independent component direction with a positive kurtosis, i.e., $\mathbf{w}(k)$ will converge to a direction of a column $\mathbf{b}_i$ of B with $\mathrm{Kurt}(s_i) > 0$. The algorithm (2) converges to an independent component direction with a negative kurtosis, i.e., $\mathbf{w}(k)$ will converge to a direction of a column $\mathbf{b}_i$ of B with $\mathrm{Kurt}(s_i) < 0$.

To prove the convergence of the algorithms, a transformation is given first. Let $\mathbf{w}(k) = B\tilde{\mathbf{z}}(k)$. Clearly, to prove that $\mathbf{w}(k)$ converges to a direction of a column $\mathbf{b}_i$ of B, it only needs to prove that all components of $\tilde{\mathbf{z}}(k)$ converge to zero except $\tilde{z}_i(k)$. Since $\mathbf{x}(k) = B\mathbf{s}(k)$, from (1) and (2), it follows that

$$\tilde{\mathbf{z}}(k+1) = \tilde{\mathbf{z}}(k) + \eta\left[b\mathbf{s}(k)\left(\tilde{\mathbf{z}}^T(k)\mathbf{s}(k)\right)^3 + a\left(1 - \|\tilde{\mathbf{z}}(k)\|^2\right)\tilde{\mathbf{z}}(k)\right] \quad (3)$$

$$\tilde{\mathbf{z}}(k+1) = \tilde{\mathbf{z}}(k) + \eta \Big[-b\mathbf{s}(k)\left(\tilde{\mathbf{z}}^T(k)\mathbf{s}(k)\right)^3 + a\left(1 - \|\tilde{\mathbf{z}}(k)\|^2\right)\tilde{\mathbf{z}}(k)\Big] \quad (4)$$

for all $k \geq 0$.

In the following section, the convergence of the Hyvärinen–Oja algorithms will be analyzed in detail.

III. Convergence Analysis

The algorithms (1) and (2) may diverge. Consider one-dimensional examples. From (1), it follows that

$$\begin{aligned} w(k+1) &= w(k) + \eta\left[bw^3(k)x^4(k) + a\left(1 - w^2(k)\right)w(k)\right] \\ &= \left[1 + \eta a - \eta\left(a - bx^4(k)\right)w(k)^2\right]w(k) \end{aligned}$$

where $\eta > 0$, $a > 0$ and $b > 0$. Suppose $a - bx^4(k) > \bar{a} > 0$ for all $k \geq 0$, where $\bar{a}$ is a constant. If $w(k) > ((1+\eta a)/\eta\bar{a})$, it holds that $\eta(a - bx^4(k))w(k)^2 - w(k) - \eta a - 1 > 0$. It follows that

$$|w(k+1)| = w^2(k)\frac{\eta\left(a - bx^4(k)\right)w^2(k) - \eta a - 1}{w(k)} > w^2(k)$$

for all $k \geq 0$. Thus, the trajectory approaches towards infinity if $w(0) > ((1+\eta a)/\eta\bar{a})$ under the condition $a - bx^4(k) > \bar{a}$ for all $k \geq 0$. From (2), it follows that

$$\begin{aligned} w(k+1) &= w(k) + \eta\left[-bw^3(k)x^4(k) + a\left(1 - w^2(k)\right)w(k)\right] \\ &= \left[1 + \eta a - \eta\left(a + bx^4(k)\right)w^2(k)\right]w(k) \end{aligned}$$

where $\eta > 0$, $a > 0$ and $b > 0$. If $w(k) > ((1+\eta a)/\eta a)$, it holds that $\eta(a + bx^4(k))w^2(k) - w(k) - \eta a - 1 > 0$. It follows that

$$|w(k+1)| = w^2(k)\frac{\eta\left(a + bx^4(k)\right)w^2(k) - \eta a - 1}{w(k)} > w^2(k)$$

for all $k \geq 0$. Thus, the trajectory approaches towards infinity if $w(0) > (1 + \eta a/\eta a)$.

The examples above show the algorithms (1) and (2) may diverge. A problem to address is therefore to find the conditions under which the algorithms are bounded. In the following section, by directly studying the SDT algorithms (1) and (2), some interesting theorems are given to guarantee the nondivergence of the algorithms in stochastic environment.

A. Invariant Sets

Definition 1: A compact set $S \subset R^n$ is called an invariant set of the algorithm $\mathbf{w}(k+1) = g(\mathbf{w}(k), \mathbf{x}(k))$, if for any $\mathbf{w}(0) \in S$, the trajectory of the algorithm starting from $\mathbf{w}(0)$ will remain in S for all $k \geq 0$.

Let $\Gamma_w = sup\{\|\mathbf{x}(k)\mathbf{x}^T(k)\|^2, k = 1, 2, \cdots\}$, assumed Γ_w is finite and $\|\mathbf{w}(0)\|^2 \neq 0$. Suppose $\|\mathbf{w}(0)\|^2 \leq (1 + \eta a/\eta a)$. Define

$$M_{w1} = \min\left\{\delta_{w1}\Big[(1+\eta a) - \eta(a - b\Gamma_w)\|\mathbf{w}(0)\|^2\Big]^2\|\mathbf{w}(0)\|^2,\ \delta_{w1}\frac{(b\Gamma_w + \eta ab\Gamma_w)^2(1+\eta a)}{\eta a^3}\right\}$$

where $a > b\Gamma_w$ and $\delta_{w1} > 0$ is a small constant. Clearly, $M_{w1} > 0$. Suppose $\|\mathbf{w}(0)\|^2 \leq ((1+\eta a)/(\eta a + \eta b\Gamma_w))$. Define

$$M_{w2} = \min\left\{\delta_{w2}\Big[(1+\eta a) - \eta(a - b\Gamma_w)\|\mathbf{w}(0)\|^2\Big]^2\|\mathbf{w}(0)\|^2,\ \delta_{w2}\frac{4(\eta b\Gamma_w)^2(1+\eta b)^3}{(\eta a + \eta b\Gamma_w)^3}\right\}$$

where $a > b\Gamma_w$ and $\delta_{w2} > 0$ is a small constant. Clearly, $M_{w2} > 0$. Define two invariant sets as

$$\begin{cases} S_{w1} = \left\{\mathbf{w}|\mathbf{w} \in R^n, M_{w1} \leq \|\mathbf{w}\|^2 \leq \frac{1+\eta a}{\eta a}\right\} \\ S_{w2} = \left\{\mathbf{w}|\mathbf{w} \in R^n, M_{w2} \leq \|\mathbf{w}\|^2 \leq \frac{1+\eta a}{\eta a + \eta b\Gamma_w}\right\}. \end{cases}$$

Theorem 1: Suppose $\eta > 0$, $a > 0$ and $b > 0$. If $b\Gamma_w/a < 0.4074$ and $\eta a < 1$, then S_{w1} is an invariant set of (1). If $b\Gamma_w/a < 0.2558$ and $\eta a < 1$, then S_{w2} is an invariant set of (2).

See the Appendix for the proof. The theorem above shows that any trajectory of algorithms (1) and (2) starting from $\mathbf{w}(0)$ in the invariant sets S_{w1} or S_{w2} will remain in S_{w1} or S_{w2}. This guarantees the nondivergence of the algorithms in stochastic environment.

The invariant sets are obtained by directly studying the original SDT algorithms (1) and (2). Γ_w is the square of the Frobenius norm and estimated from the available samples. Clearly, the size of samples does not influence the result of Theorem 1. On the other hand, Theorem 1 does not require prewhitening. Thus, the invariant sets of the original algorithms are also suitable for the data, which is not prewhitened.

Since $\|\mathbf{w}(k)\| = \|B\tilde{\mathbf{z}}(k)\| = \|\tilde{\mathbf{z}}(k)\|$, from (3) and (4), a similar theorem can be obtained as follows. Let $\Gamma_{\tilde{z}} = sup\{\|\mathbf{s}(k)\mathbf{s}^T(k)\|^2, k = 1, 2, \cdots\}$, assumed $\Gamma_{\tilde{z}}$ is finite and $\|\tilde{\mathbf{z}}(0)\|^2 \neq 0$. Suppose $\|\tilde{\mathbf{z}}(0)\|^2 \leq ((1+\eta a)/\eta a)$. Define

$$M_{\tilde{z}1} = \min\left\{\delta_{\tilde{z}1}\left[(1+\eta a) - \eta(a - b\Gamma_{\tilde{z}})\|\tilde{\mathbf{z}}(0)\|^2\right]^2\|\tilde{\mathbf{z}}(0)\|^2,\ \delta_{\tilde{z}1}\frac{(b\Gamma_{\tilde{z}} + \eta ab\Gamma_{\tilde{z}})^2(1+\eta a)}{\eta a^3}\right\} > 0$$

where $\delta_{\tilde{z}1} > 0$ is a small constant. Suppose $\|\tilde{\mathbf{z}}(0)\|^2 \leq (1 + \eta a/\eta a + \eta b\Gamma_{\tilde{z}})$. Define

$$M_{\tilde{z}2} = \min\left\{\delta_{\tilde{z}2}\left[(1+\eta a) - \eta(a - b\Gamma_{\tilde{z}})\|\tilde{\mathbf{z}}(0)\|^2\right]^2\|\tilde{\mathbf{z}}(0)\|^2,\ \delta_{\tilde{z}2}\frac{4(\eta b\Gamma_{\tilde{z}})^2(1+\eta a)^3}{(\eta a + \eta b\Gamma_{\tilde{z}})^3}\right\} > 0$$

where $\delta_{\tilde{z}2} > 0$ is a small constant. Denote $S_{\tilde{z}1}$ and $S_{\tilde{z}2}$ as two invariant set by

$$\begin{cases} S_{\tilde{z}1} = \left\{\tilde{\mathbf{z}}|\tilde{\mathbf{z}} \in R^n, M_{\tilde{z}1} \leq \|\tilde{\mathbf{z}}\|^2 \leq \frac{1+\eta a}{\eta a}\right\} \\ S_{\tilde{z}2} = \left\{\tilde{\mathbf{z}}|\tilde{\mathbf{z}} \in R^n, M_{\tilde{z}2} \leq \|\tilde{\mathbf{z}}\|^2 \leq \frac{1+\eta a}{\eta a + \eta b\Gamma_{\tilde{z}}}\right\}. \end{cases}$$

From the analysis of theorem 1, this immediately allows us to state the following theorem.

Theorem 2: Suppose $\eta > 0$, $a > 0$ and $b > 0$. If $(b\Gamma_{\tilde{z}}/a) < 0.4074$ and $\eta a < 1$, then $S_{\tilde{z}1}$ is an invariant set of (3). If $b\Gamma_{\tilde{z}}/a < 0.2558$ and $\eta a < 1$, then $S_{\tilde{z}2}$ is an invariant set of (4).

Theorems 1 and 2 give the nondivergent conditions of the algorithms in stochastic environment. It is clear the norm of the algorithms do not go to infinity under the conditions, even in stochastic environment. In these invariant sets, the local convergence of the algorithms will be further studied in the following subsection.

B. DDT Algorithms and Local Convergence

The convergence analysis of algorithms (1) and (2) can be indirectly studied by analyzing the corresponding DDT algorithms in the invariant sets. The DDT algorithms characterize the average evolution of the original algorithms. The convergence of the DDT algorithms can reflect the dynamical behaviors of the original algorithms [43], [48]. Here, we only present the convergence analysis of (1). The convergence results of (2) can be obtained by using similar method. This paper also gives the convergence results of (2), but the analysis process of its convergence is omitted.

Following Zufiria's method [48], the corresponding DDT algorithms can be obtained by taking the conditional expectation $E\{\mathbf{w}(k+1)/\mathbf{w}(0), \mathbf{w}(i), i < k\}$ on both sides of the algorithm (1). Since $E\{\mathbf{w}(k)\} = BE\{\tilde{\mathbf{z}}(k)\}$, the convergence analysis of the DDT algorithm of (1) can be transformed to studying the convergence of the DDT algorithm of (3). Taking the conditional expectation $E\{\tilde{\mathbf{z}}(k+1)/\tilde{\mathbf{z}}(0), \tilde{\mathbf{z}}(i), i < k\}$ on both sides of (3), the corresponding DDT algorithm is written as:

$$\mathbf{z}(k+1) = \mathbf{z}(k) + \eta \left[bE\left\{ \mathbf{s}(k)\left(\tilde{\mathbf{z}}^T(k)\mathbf{s}(k)\right)^3 \right\} + a\left(1 - \|\mathbf{z}(k)\|^2\right)\mathbf{z}(k) \right] \quad (5)$$

for all $k \geq 0$, where $\mathbf{z}(k) = E\{\tilde{\mathbf{z}}(k)\}$. The DDT algorithm (5) preserves the discrete time form of the original algorithm (3) and allows a more realistic behavior of the learning gain [48]. Since the DDT algorithm (5) represents an average evolution of the algorithm (3), the invariant set $S_{\tilde{z}1}$ is also an invariant set of (5).

The convergence of the algorithm (5) is analyzed in detail, which is presented in the Appendix (Lemmas 4–9). The main results are as follows.

Theorem 3: Suppose that $\mathbf{z}(0) \in S_{\tilde{z}1}$, $\mathbf{z}(0) \notin V_q^{\perp}$ and $\Gamma_{\tilde{z}} > \mathrm{Kurt}(s_i) + 3$. If $\mathrm{Kurt}(s_{j_0})z_{j_0}^2(0) > \mathrm{Kurt}(s_i)z_i^2(0), j_0 \in Q$ and $i \in T/j_0$, there must exist a constant $z_{j_0}^*$ so that $\lim_{k\to+\infty} E\{\mathbf{w}(k)\} = B\mathbf{z}(k) = \mathbf{b}_{j_0} \cdot z_{j_0}^*$.

See the Appendix for the proof. Theorem 3 shows the average evolution of the algorithm (1) converges to an independent component direction with a positive kurtosis. The analysis can be similarly used to study the convergence of (2). The results can be obtained as follows.

Theorem 4: Suppose $\mathbf{z}(0) \in S_{\tilde{z}2}$ and $\mathbf{z}(0) \notin V_p^{\perp}$. If $-\mathrm{Kurt}(s_{i_0})z_{i_0}^2(0) > -\mathrm{Kurt}(s_i)z_i^2(0), i_0 \in P$ and $i \in T/i_0$, there must exist a constant $z_{j_0}^*$ so that $\lim_{k\to+\infty} E\{\mathbf{w}(k)\} = \mathbf{b}_{j_0} \cdot z_{j_0}^*$, where $\mathrm{Kurt}(s_{i_0}) < 0$.

The theorems above show the trajectories of the DDT algorithms starting from the invariant sets will converge to an independent component direction with a positive kurtosis or a negative kurtosis. The results can shed some light on the dynamical behaviors of the original algorithms (1) and (2).

IV. Extension of the DDT Algorithms

The DDT algorithms of the algorithms (1) and (2) can be clearly expressed as

$$E\{\mathbf{w}(k+1)\} = E\left\{\mathbf{w}(k) + \eta\left[b\mathbf{x}(k)\left(\mathbf{w}^T(k)\mathbf{x}(k)\right)^3 + a\left(1 - \|\mathbf{w}(k)\|^2\right)\mathbf{w}(k)\right]\right\}$$

$$E\{\mathbf{w}(k+1)\} = E\left\{\mathbf{w}(k) + \eta\left[-b\mathbf{x}(k)\left(\mathbf{w}^T(k)\mathbf{x}(k)\right)^3 + a\left(1 - \|\mathbf{w}(k)\|^2\right)\mathbf{w}(k)\right]\right\}.$$

Let $\bar{\mathbf{w}}(k) = E\{\mathbf{w}(k)\}$, the DDT algorithms are rewritten as

$$\bar{\mathbf{w}}(k+1) = \bar{\mathbf{w}}(k) + \eta\left[bE\left\{\mathbf{x}(k)\left(\bar{\mathbf{w}}^T(k)\mathbf{x}(k)\right)^3\right\} + a\left(1 - \|\bar{\mathbf{w}}(k)\|^2\right)\bar{\mathbf{w}}(k)\right] \quad (6)$$

$$\bar{\mathbf{w}}(k+1) = \bar{\mathbf{w}}(k) + \eta\left[-bE\left\{\mathbf{x}(k)\left(\bar{\mathbf{w}}^T(k)\mathbf{x}(k)\right)^3\right\} + a\left(1 - \|\bar{\mathbf{w}}(k)\|^2\right)\bar{\mathbf{w}}(k)\right]. \quad (7)$$

By studying the convergence of the DDT algorithms (6) and (7), we indirectly analyze the convergence of the original SDT algorithms (1) and (2). Clearly, the algorithms will converge under certain conditions. However, the original SDT algorithms are with slow convergence speed, and convergence accuracy of these algorithms is not good [27], [45], [46]. To improve the performance of the original algorithms, the DDT algorithms can be extended to the block versions of the original algorithms. If the expectation operation is computed by

$$E\left\{\mathbf{x}(k)\left(\bar{\mathbf{w}}^T(k)\mathbf{x}(k)\right)^3\right\} = \frac{1}{L}\sum_{j=1}^{L}\mathbf{x}(j)\left(\bar{\mathbf{w}}^T(j)\mathbf{x}(j)\right)^3$$

where L is the block size [27], the algorithms (6) and (7) are also the block algorithms. The block algorithms establish a relationship between the original SDT algorithms and the corresponding DDT algorithms. Clearly, if $L = 1$, they are the original SDT algorithms. If L is the size of all examples, they are the corresponding DDT algorithms. The block algorithms represent an average evolution of the original algorithms over all blocks of samples. Thus, the invariant sets in Theorem 1 (where is $L = 1$) also are the invariant sets of the block algorithms. Furthermore, the block algorithms can get a good convergence speed and accuracy in practice. The following simulations will illustrate their performance.

V. Simulations and Discussions

In this section, three set of experiments will be carried out to illustrate the convergence of the class of Hyvärinen–Oja's ICA learning algorithms with constant learning rates.

A. Example 1

The DDT algorithms characterize the average evolution of the original algorithms. The convergence of the DDT algorithms

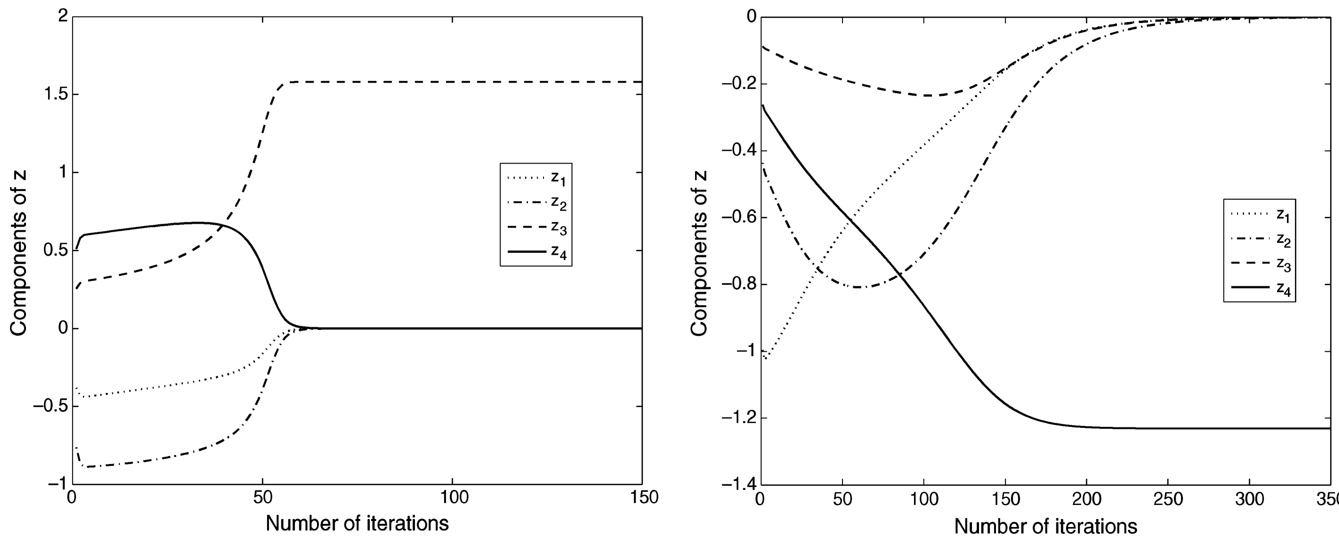


Fig. 1. The evolution of components of $\mathbf{z}$ with initial value: $\mathbf{z}(0) = [-0.3\ -0.6\ 0.2\ 0.4]^T$ (left) and $\mathbf{z}(0) = [-1\ -0.5\ -0.1\ -0.2]^T$ (right), while $\sigma = 1$ and $\mathrm{Kurt}(s_1) = -0.6, \mathrm{Kurt}(s_2) = 0, \mathrm{Kurt}(s_3) = 3, \mathrm{Kurt}(S_4) = 0.4$.

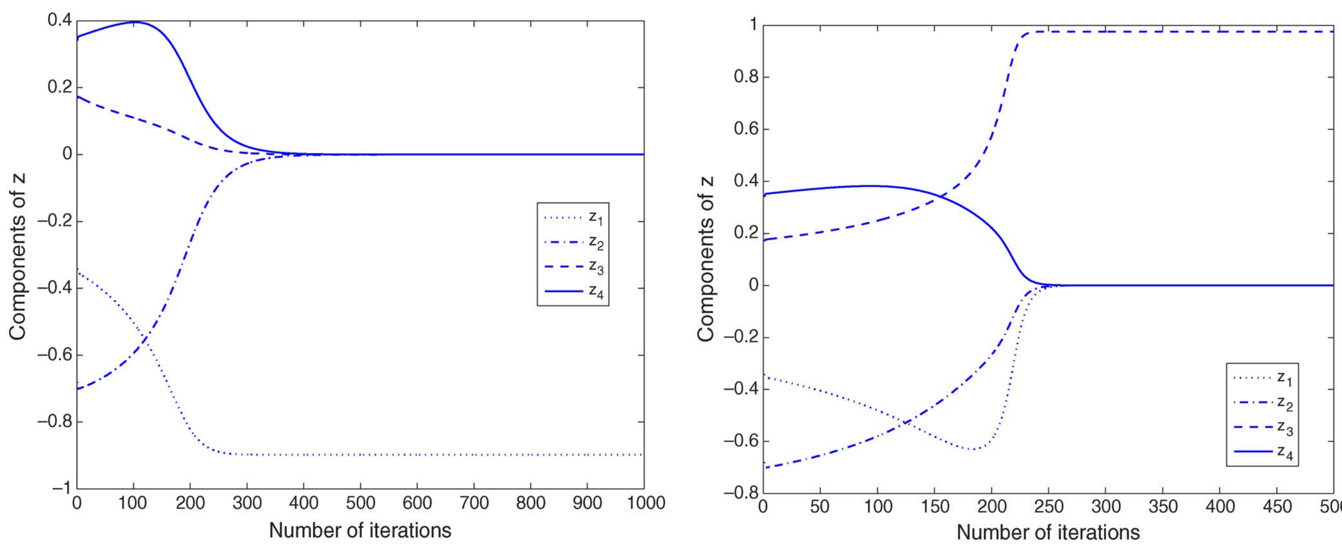


Fig. 2. The evolution of components of $\mathbf{z}$ with the different kurtosis: $\mathrm{Kurt}(s_1) = -0.6, \mathrm{Kurt}(s_2) = 0, \mathrm{Kurt}(s_3) = 3, \mathrm{Kurt}(S_4) = -0.4$ (left); $\mathrm{Kurt}(s_1) = -0.6, \mathrm{Kurt}(s_2) = 0, \mathrm{Kurt}(s_3) = -2.5, \mathrm{Kurt}(S_4) = -0.4$ (right), while $\sigma = -1$ and the initial value: $\mathbf{z}(0) = [-0.4\ -0.8\ 0.2\ 0.4]^T$.

can reflect the dynamical behaviors of the original algorithms in the invariant sets. From (10), consider four-dimensional DDT algorithms as (8) for all $k \geq 0$.

$$\begin{bmatrix} z_1(k+1) \\ z_2(k+1) \\ z_3(k+1) \\ z_4(k+1) \end{bmatrix} = \begin{bmatrix} z_1(k) \\ z_2(k) \\ z_3(k) \\ z_4(k) \end{bmatrix} + \eta \left\{ \sigma b \begin{bmatrix} \mathrm{Kurt}(s_1) z_1^3(k) \\ \mathrm{Kurt}(s_2) z_2^3(k) \\ \mathrm{Kurt}(s_3) z_3^3(k) \\ \mathrm{Kurt}(s_4) z_4^3(k) \end{bmatrix} + \left(3\sigma b \|\mathbf{z}(k)\|^2 + \eta a \left(1 - \|\mathbf{z}(k)\|^2\right)\right) \begin{bmatrix} z_1(k) \\ z_2(k) \\ z_3(k) \\ z_4(k) \end{bmatrix} \right\}. \quad (8)$$

Let $\eta = 0.1$, $a = 5$ and $b = 0.5$. While $\sigma = 1$ and $\mathrm{Kurt}(s_1) = -0.6$, $\mathrm{Kurt}(s_2) = 0$, $\mathrm{Kurt}(s_3) = 3$, $\mathrm{Kurt}(s_4) = 0.4$, in Fig. 1, the evolution of components of $\mathbf{z}$ is shown with initial vector $\mathbf{z}(0) = [-0.3\ -0.6\ 0.2\ 0.4]^T$ (left) and $\mathbf{z}(0) = [-1\ -0.5\ -0.1\ -0.2]^T$ (right). While $\sigma = -1$ and the initial value $\mathbf{z}(0) = [-0.4\ -0.8\ 0.2\ 0.4]^T$, Fig. 2 shows the evolution results of components of $\mathbf{z}$ with the kurtosis: $\mathrm{Kurt}(s_1) = -0.6$, $\mathrm{Kurt}(s_2) = 0$, $\mathrm{Kurt}(s_3) = 3$, $\mathrm{Kurt}(s_4) = -0.4$ (left) and $\mathrm{Kurt}(s_1) = -0.6$, $\mathrm{Kurt}(s_2) = 0$, $\mathrm{Kurt}(s_3) = -2.5$, $\mathrm{Kurt}(s_4) = -0.4$ (right).

Since $E\{\mathbf{w}(k)\} = B\mathbf{z}(k)$, the DDT algorithms converge to an independent component direction with a positive kurtosis or a negative kurtosis as all components of $\mathbf{z}(k)$ converge to zero except a component. The simulation illustrates the results of Theorems 3 and 4.

In addition, the initial value and kurtosis together determine which independent component will be extracted in the DDT algorithms. However, in stochastic environment, the distribution of the variable also affects which independent component will be extracted. The following experiments will further illustrate the theory derived.

B. Example 2

The example will illustrate the convergence of the block algorithm (6). There are two stochastic signals s_1 and s_2 with

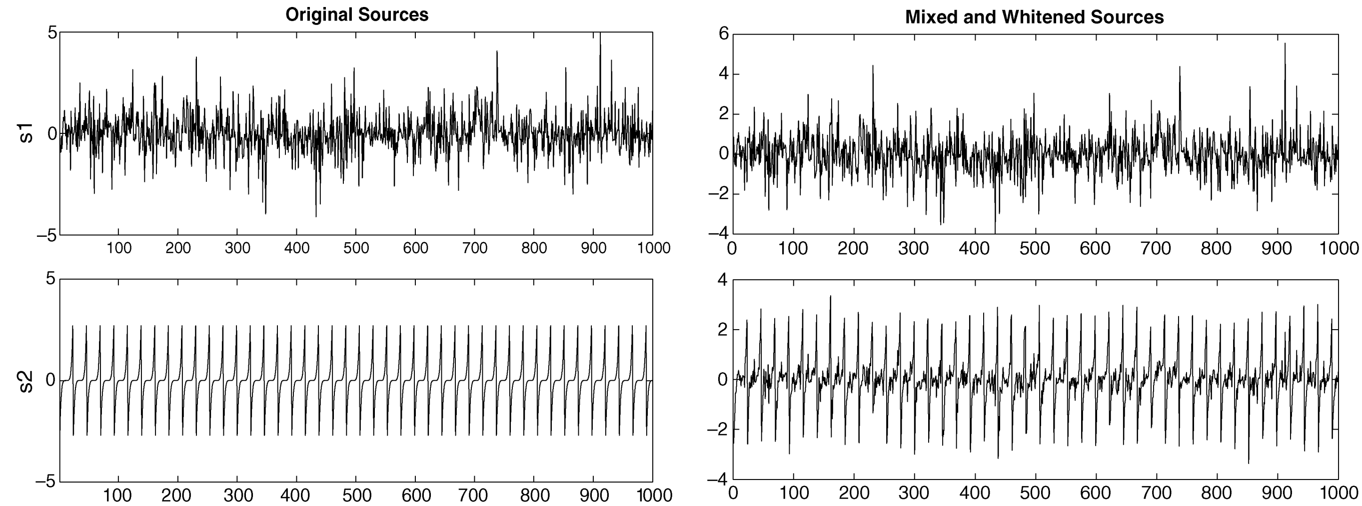


Fig. 3. Original sources s_1 and s_2 with $\mathrm{Kurt}(s_1) = 2.6822, \mathrm{Kurt}(s_2) = 2.4146$ (left); two mixed and writened signals (right).

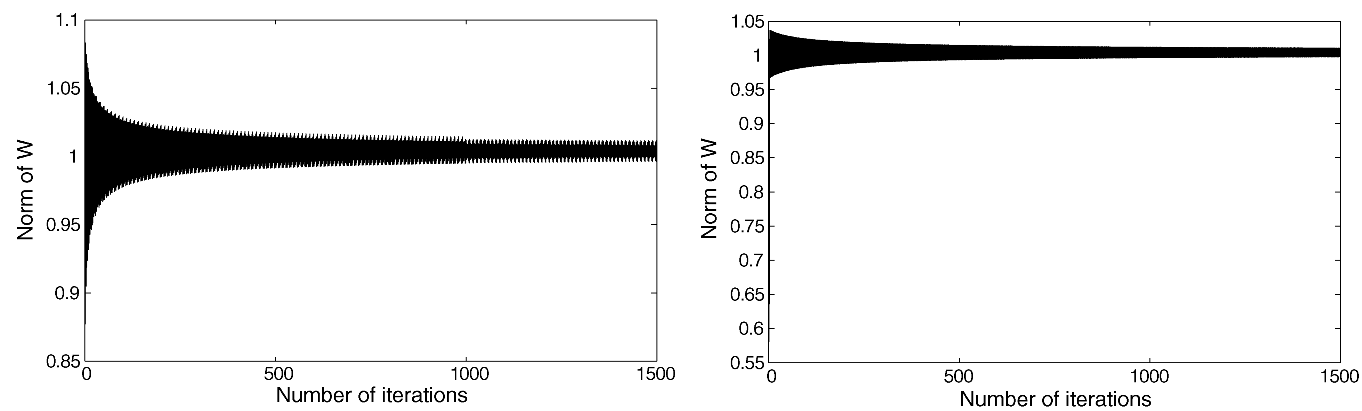


Fig. 4. Evolution of norm of $\mathbf{w}$ with $L = 200, \mathbf{w}(0) = [0.3\ 0.05]^T$ (left) and with $L = 1000, \mathbf{w}(0) = [0.03\ 0.5]^T$ (right).

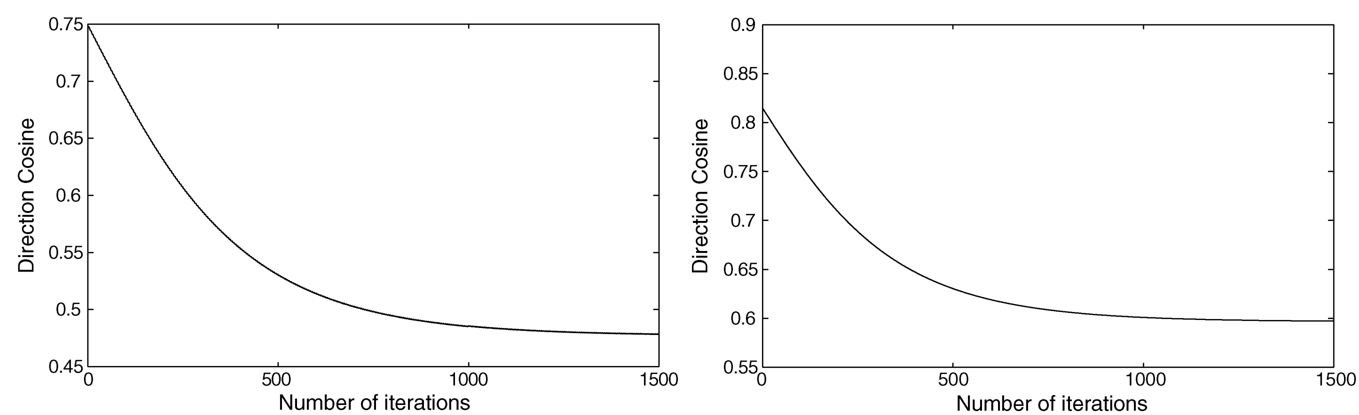


Fig. 5. Evolution of $\mathrm{DirectionCosine(k)}$ with $L = 200, \mathbf{w}(0) = [0.3\ 0.05]^T$ (left) and with $L = 1000, \mathbf{w}(0) = [0.03\ 0.5]^T$ (right).

$\mathrm{Kurt}(s_1) = 2.6822$, $\mathrm{Kurt}(s_2) = 2.4146$. The original signals are artificially generated and are samples of 1000 values with zero mean and unit variance, shown in Fig. 3 (left). Randomly mixed and written signals are shown in Fig. 3 (right). Since $\Gamma_w = 1192.7$, let $\eta = 0.05$, $a = 25$ and $b = 0.01$. The algorithm (6) is used to extracted an independent signal from the mixed signals. The evolution of norm of $\mathbf{w}$ is shown in Fig. 4 with $L = 200$, $\mathbf{w}(0) = [0.03, 0, 5]$ (left) and $L = 1000$, $\mathbf{w}(0) = [0.3, 0.05]$ (right). Clearly, the norm changes in a certain range and will not go to infinity.

To verify whether the direction of $\mathbf{w}$ converge or not, we will compute the direction cosine between $\mathbf{w}$ and a reference vector $\mathbf{v} = [1\ 1]^T$ at each k as [42]:

$$\mathrm{DirectionCosine}(k) = \frac{\left|\mathbf{w}^T(k) \cdot \mathbf{v}\right|}{\|\mathbf{w}(k)\| \cdot \|\mathbf{v}\|}$$

for all $k \geq 0$. We do not known the exact extracting direction and which signal will be extracted in advance. Thus, a reference vector is selected. Clearly, $DirectionCosine(k)$ will approach to a certain value if the direction of $\mathbf{w}$ converges. Fig. 5 shows the evolution of $\mathrm{Directioncosine}$ with $L = 200$,

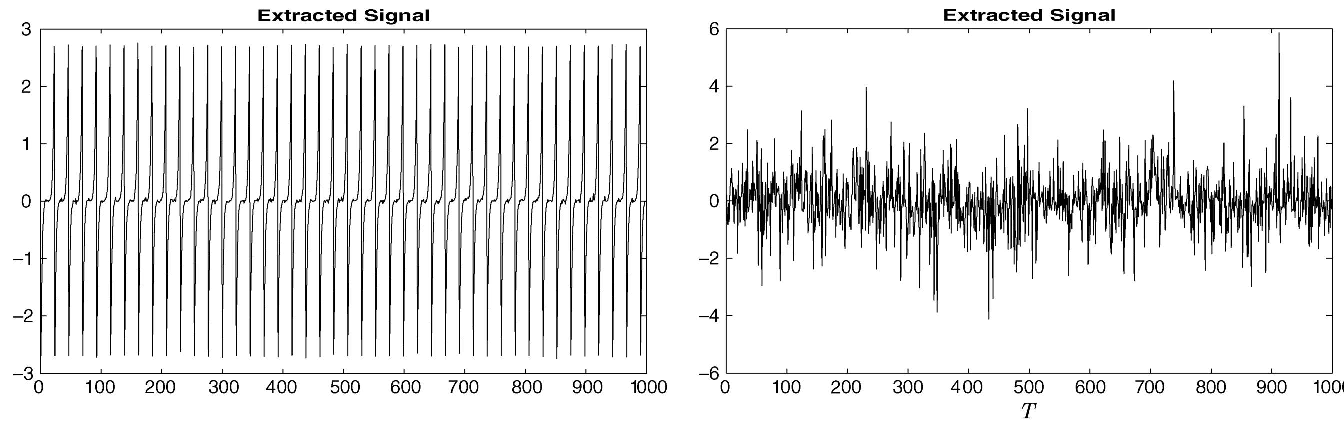


Fig. 6. Extracted signal with $L = 200, \mathbf{w}(0) = [0.3\ 0.05]^T$ (left) and with $L = 1000, \mathbf{w}(0) = [0.03\ 0.5]^T$ (right).

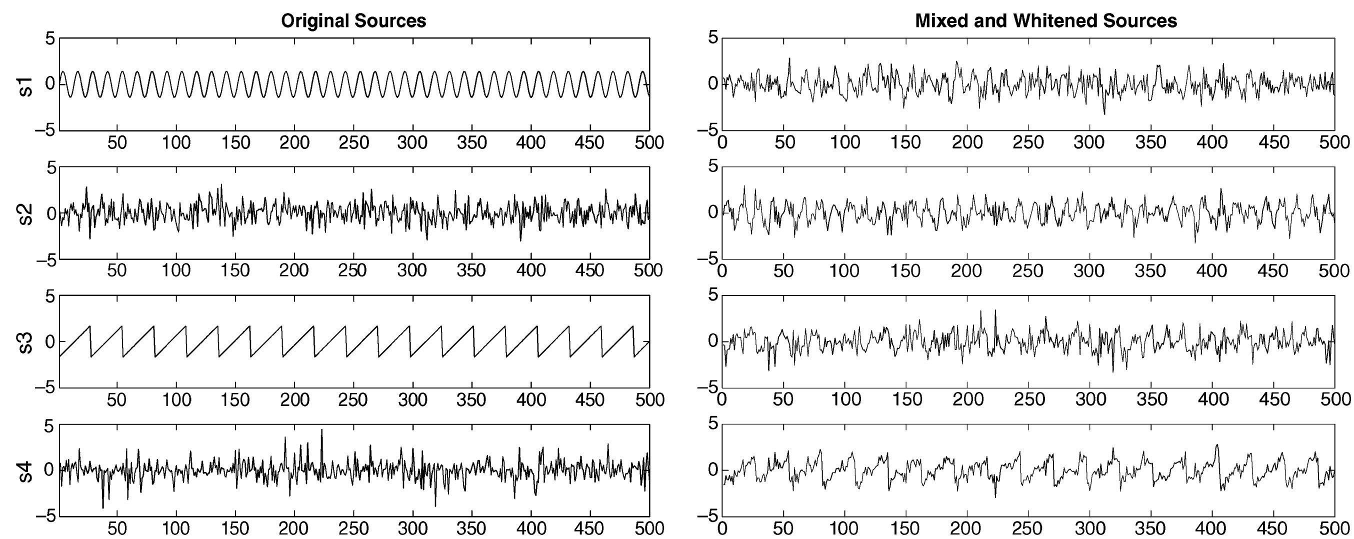


Fig. 7. Original sources s_1, s_2, s_3 and s_4 with $\mathrm{Kurt}(s_1) = -1.4984, \mathrm{Kurt}(s_2) = 0.0362, \mathrm{Kurt}(s_3) = -1.1995\ \mathrm{Kurt}(s_4) = 2.4681$ (left); four mixed and writened signals (right).

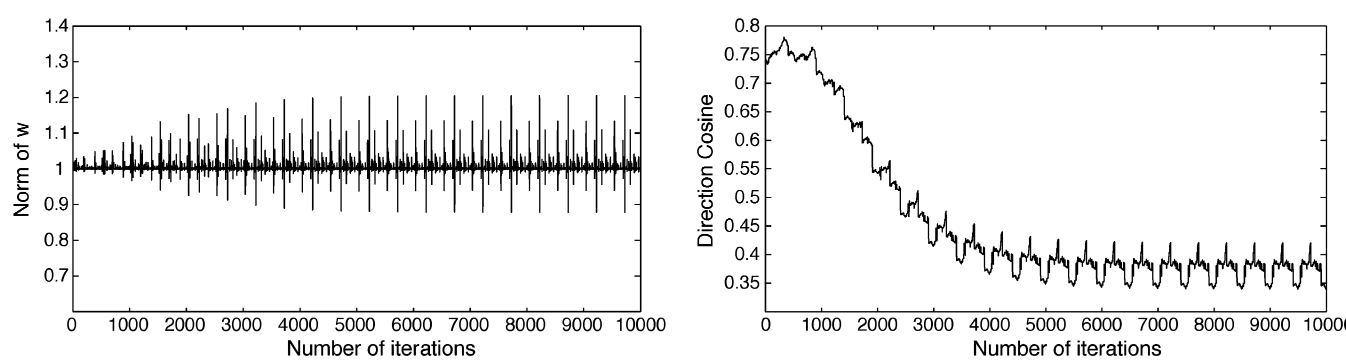


Fig. 8. Evolution of norm of $\mathbf{w}$ (left) and the evolution of $\mathrm{DirectionCosine(k)}$ (right) of the algorithm (1) with the initial value $\mathbf{w}(0) = [0.6\ 1.2\ 0.1\ 0.1]^T$.

$\mathbf{w}(0) = [0.03, 0, 5]$ (left) and $L = 1000$, $\mathbf{w}(0) = [0.3, 0.05]$ (right).

At the same time, to verify whether the direction that $\mathbf{w}$ goes to is right or not, a signal is extracted by projecting the mixed whitened signals to the direction. The extracted signals are shown in Fig. 6. The example shows that the block algorithm (6) exists an invariant set and converges to an independent component direction in the invariant set.

C. Example 3

In the following experiment, four signals are used, as shown in Fig. 7 (left). They are of zero mean and unit variance with $\mathrm{Kurt}(s_1) = -1.4984$, $\mathrm{Kurt}(s_2) = 0.0362$, $\mathrm{Kurt}(s_3) = -1.1995$ and $\mathrm{Kurt}(s_4) = 2.4681$. Since $\Gamma_w = 471.8463$, let $\eta = 0.05$, $a = 12$ and $b = 0.01$. The original SDT algorithms (1) and (2) are used to extract the independent signals from the mixed and written signals, as shown in Fig. 7 (right). As for the algorithm (1), the evolution of norm and $\mathrm{DirectionCosine}$ are shown in Fig. 8. An independent signal with a positive kurtosis is extracted, shown in Fig. 10 (left). As for the algorithm (2), the evolution of norm and $\mathrm{DirectionCosine}$ are shown in Fig. 9. An independent signal with a negative kurtosis is extracted, as shown in Fig. 10 (right).

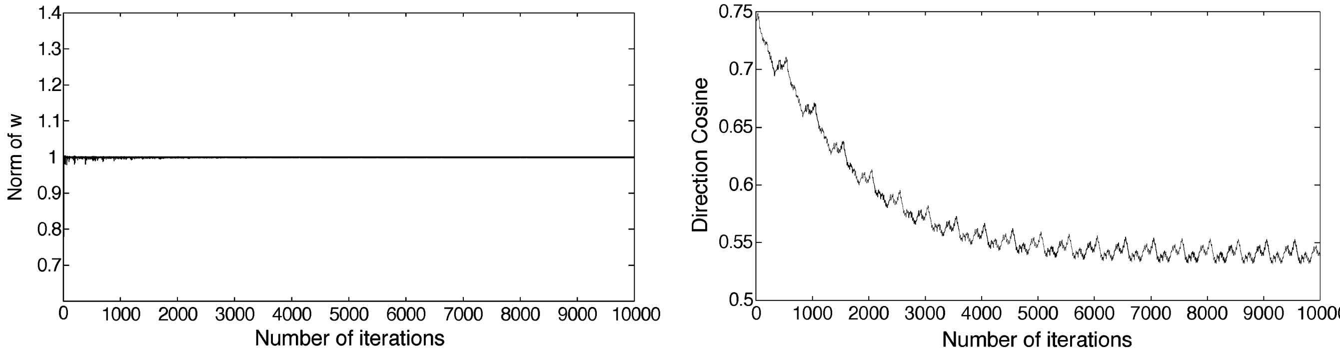


Fig. 9. Evolution of norm of $\mathbf{w}$ (left) and the evolution of $\mathrm{DirectionCosine(k)}$ (right) of the algorithm (2) with the initial value $\mathbf{w}(0) = [0.6\ 1.2\ 0.1\ 0.1]^T$.

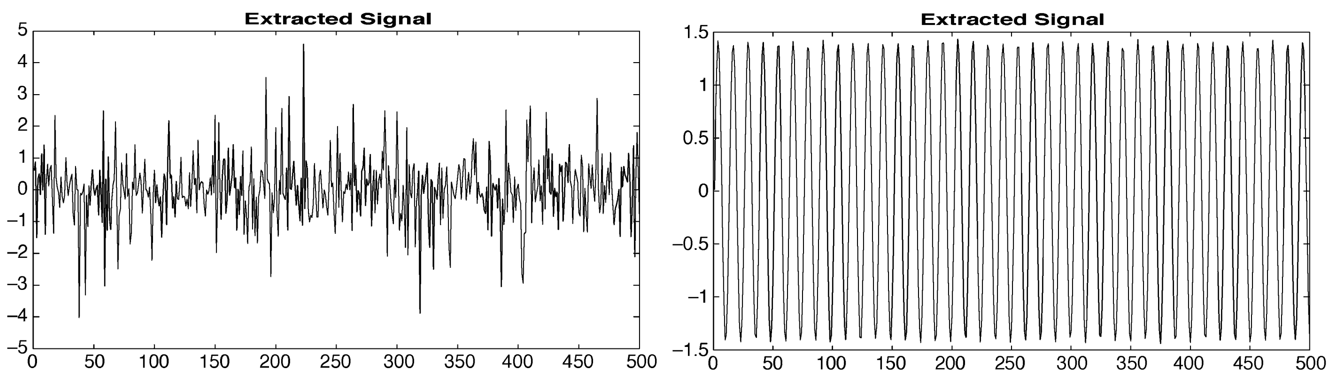


Fig. 10. Extracted signal of (1) with $\mathrm{Kurt}(s_4) = 2.4681$ (left) and the extracted signal of (2) with $\mathrm{Kurt}(s_1) = -1.4984$ (right).

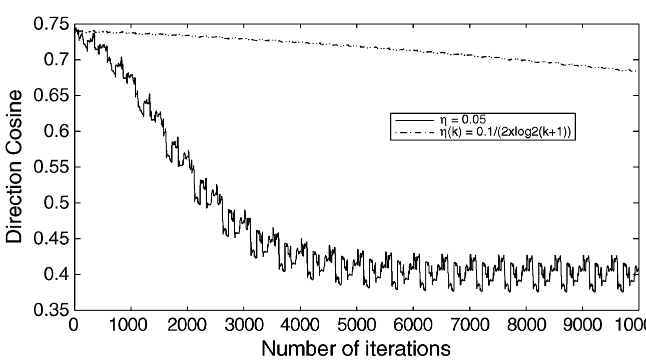


Fig. 11. Comparison of convergent rate of algorithm (1) with a constant learning rate $\eta = 0.05$ and a zero-approaching learning rate $\eta(k) = (0.1/2log_2(k+1))$. The dashed–dotted line denotes the evolution of the algorithm with the a zero-approaching learning rate.

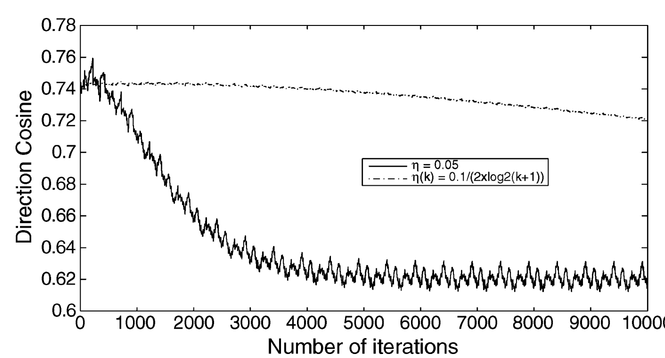


Fig. 12. Comparison of convergent rate of algorithm (2) with a constant learning rate $\eta = 0.05$ and a zero-approaching learning rate $\eta(k) = (0.1/2log_2(k+1))$. The dashed–dotted line denotes the evolution of the algorithm with the a zero-approaching learning rate.

We also compare the convergence rate of the algorithms (1) and (2) with a constant learning rate $\eta = 0.05$ and $\eta(k) = (0.1/2\log_2(k+1))$, $(k \geq 1)$. The evolution trajectories of $\mathrm{Directioncosine}$ are shown in Figs. 11 and 12. The dashed–dotted line denotes the evolution of the algorithms with the zero-approaching learning rate. Clearly, the algorithms with a constant learning rate converge faster in this experiment. Maybe it is possible to improve the performance by selecting a proper zero-approaching learning rate. However, there is no idea of selecting a proper zero-approaching learning rate so that the algorithms do not diverge and converge faster. Constant learning rates simplify the application of the original algorithms in practice and the convergent properties can be guaranteed in stochastic environment.

VI. Conclusion

The convergence of a class of Hyvärinen–Oja's ICA learning algorithms with constant learning algorithms is studied. It is shown that the algorithms converge to an independent component direction with a positive kurtosis or a negative kurtosis. Since constant learning rates are used, the requirements for the algorithms to be applied in practical applications can be met. The most important result of this paper is that some invariant sets are obtained in stochastic environment so that the nondivergence of the original SDT algorithms is guaranteed by selecting proper learning parameters. In these invariant sets, the local convergence of the original algorithms has been indirectly studied via the DDT method. To improve the performance of the original algorithms, the corresponding DDT algorithms can be extended to the block versions of the original algorithms. Simulations are carried out to further support the results obtained.

APPENDIX

Lemma 1: Suppose that $D > 0$, $E > 0$. It holds that

$$[D - Eh]^2 \cdot h \leq \frac{4D^3}{27E}$$

for all $0 < h \leq (D/E)$.

Proof: Define a differentiable function

$$f(h) = [D - Eh]^2 \cdot h$$

for $0 < h < D/E$. It follows that $\dot{f}(h) = [D - Eh] \cdot [D - 3Eh]$. It is easy to get the maximum point of $f(h)$ on the interval $(0, D/E)$ must be $h = D/3E$. Thus, it follows that

$$[D - Eh]^2 \cdot h \leq \frac{4D^3}{27E}$$

where $0 < h < D/E$. The proof is complete.

Lemma 2: Suppose $H > F > 0$. If $N > C > 0$, then

$$\frac{F + C}{H + C} < \frac{F + N}{H + N}.$$

Proof: It holds that

$$\frac{F + C}{H + C} - \frac{F + N}{H + N} = \frac{(F - H)(N - C)}{(H + C)(H + N)} < 0.$$

The proof is complete.

Lemma 3: Suppose $0 < \varrho < 1$. If $0 < \eta a < 1$ and $\varrho \leq h \leq 2$, it holds that

$$\left|(1 - h)^2 - \eta a h\right| \leq \zeta$$

where $\zeta = \max\{|(1 - \varrho)^2 - \eta a \varrho|, |1 - 2\eta a|\}$ and $0 < \zeta < 1$.

Proof: Define a differentiable function

$$f(h) = (1 - h)^2 - \eta a h$$

for all $\varrho \leq h \leq 2$, where $0 < \eta a < 1$. It follows that $\dot{f}(h) = -2(1 - h) - \eta a$. it is easy to get the maximum point of $f(h)$ on the interval $[\varrho, 2]$ must be $h = \varrho$ or $h = 2$. Thus, it follows that

$$\left|(1 - h)^2 - \eta a h\right| \leq \zeta$$

where $\zeta = \max\{|(1 - \varrho)^2 - \eta a \varrho|, |1 - 2\eta a|\} < 1$. The proof is complete.

Proof of Theorem 1: From (1), it follows that

$$\mathbf{w}(k+1) = \left[1 + \eta a - \eta a \|\mathbf{w}(k)\|^2\right] \mathbf{w}(k) + \eta b \mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k)$$

for all $k \geq 0$. Thus, there must exist $\alpha(k)$ so that

$$\|\mathbf{w}(k+1)\| = \alpha(k) \left|\left|1 + \eta a - \eta a \|\mathbf{w}(k)\|^2\right| + \eta b \Gamma_w \|\mathbf{w}(k)\|^2\right| \|\mathbf{w}(k)\|$$

for all $k \geq 0$, where $0 < \alpha(k) \leq 1$. Since

$$\alpha(k) = \frac{\|\mathbf{w}(k+1)\|}{\left|\left|1 + \eta a - \eta a \|\mathbf{w}(k)\|^2\right| + \eta b \Gamma_w \|\mathbf{w}(k)\|^2\right| \|\mathbf{w}(k)\|}$$

where

$$\|\mathbf{w}(k+1)\| = \left\|\left[1 + \eta a - \eta a \|\mathbf{w}(k)\|^2\right] \mathbf{w}(k) + \eta \mathbf{w}^T(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k)\mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}(k)\right\|$$

there must exist a small constant δ_{w1} so that $0 < \delta_{w1} < \alpha(k) \leq 1$. So, if $\|\mathbf{w}(k)\|^2 < (1 + \eta a/\eta a)$, it holds that

$$\|\mathbf{w}(k+1)\|^2 = \left[\alpha(k)(1 + \eta a) - \alpha(k) \times (\eta a - \eta b \Gamma_w) \|\mathbf{w}(k)\|^2\right]^2 \|\mathbf{w}(k)\|^2$$

for all $k \geq 0$, where $0 < \delta_{w1} < \alpha(k) \leq 1$. Since

$$\|\mathbf{w}(k)\|^2 < \frac{1 + \eta a}{\eta a} < \frac{\alpha(k)(1 + \eta a) + 1}{\alpha(k)(\eta a - \eta b \Gamma_w)}$$

where $a > b\Gamma_w$. By Lemma 1, it follows that

$$\begin{aligned} &\|\mathbf{w}(k+1)\|^2 \\ &= \left[\alpha(k)(1 + \eta a) - \alpha(k)(\eta a - \eta b \Gamma_w) \|\mathbf{w}(k)\|^2\right]^2 \|\mathbf{w}(k)\|^2 \\ &\leq \max_{0<h<\psi(k)} \left\{\left[\alpha(k)(1 + \eta a) - \alpha(k)(\eta a - \eta b \Gamma_w) h\right]^2 h\right\} \\ &\leq \frac{4\left[\alpha(k)(1 + \eta a)\right]^3}{27\alpha(k)(\eta a - \eta b \Gamma_w)} \end{aligned} \tag{9}$$

where $\psi(k) = ((1 + \eta a)/(\eta a - \eta b \Gamma_w))$. Clearly, if $(b\Gamma_w/a) < 0.4074$ and $\eta a < 1$, it follows that

$$\|w(k+1)\|^2 \leq \frac{4\left[\alpha(k)(1 + \eta a)\right]^3}{27\alpha(k)(\eta a - \eta b \Gamma_w)} < \frac{1 + \eta a}{\eta a}.$$

From (9), it follows that

$$\|\mathbf{w}(k+1)\|^2 \geq \min_{0<h<\psi(k)} \left\{[\alpha(k)(1 + \eta a) - \alpha(k) \times (\eta a - \eta b \Gamma_w) h]^2 h\right\}$$

where $\psi(k) = ((1 + \eta a)/(\eta a - \eta b \Gamma_w))$. From Lemma 1, it is easy to see the minimum point is $\|\mathbf{w}(k)\|^2 = \|\mathbf{w}(0)\|^2$ or $\|\mathbf{w}(k)\|^2 = (1 + \eta a/\eta a)$ in the interval $[0, (1 + \eta a/\eta a)]$. Thus, it can be obtained that $\|\mathbf{w}(k+1)\|^2 \geq M_{w1} > 0$, for all $k \geq 0$.

From (2), there must exist $\beta(k)$ and a small constant δ_{w2} so that

$$\|\mathbf{w}(k+1)\| = \beta(k) \left|\left|1 + \eta a - \eta a \|\mathbf{w}(k)\|^2\right| + \eta b \Gamma_w \|\mathbf{w}(k)\|^2\right| \|\mathbf{w}(k)\|$$

for all $k \geq 0$, where $0 < \delta_{w2} < \beta(k) \leq 1$. If $\|\mathbf{w}(k)\|^2 < ((1+\eta a)/\eta a + \eta b \Gamma_w)$, it holds that

$$\|\mathbf{w}(k+1)\|^2 = \left[\beta(k)(1+\eta a) - \beta(k) \times (\eta a - \eta b \Gamma_w) \|\mathbf{w}(k)\|^2\right]^2 \|\mathbf{w}(k)\|^2$$

for all $k \geq 0$, where $0 < \delta_{w2} < \beta(k) \leq 1$. If $b\Gamma_w/a < 0.2558$ and $\eta a < 1$, it follows that

$$M_{w2} \leq \|\mathbf{w}(k+1)\|^2 < \frac{1+\eta a}{\eta a + \eta b \Gamma_w}$$

for all $k \geq 0$. The proof is complete.

The following conclusion can be found in [15]. This paper provides its proof.

Lemma 4 (Hyvärinen–Oja [15]): For any $i \in T$, it holds that

$$E\left\{ s_i \left(\sum_j^n \tilde{z}_j(k) s_j \right)^3 \right\} = \mathrm{Kurt}(s_i) z_i^3(k) + 3\|\mathbf{z}(k)\|^2 z_i(k)$$

for all $k \geq 0$.

Proof: It holds that

$$\begin{aligned} \left(\sum_{j=1}^n \tilde{z}_j(k) s_j(k) \right)^3 &= \sum_{i=1}^n \tilde{z}_i^3(k) s_i^3(k) \\ &\quad + \sum_{i=1}^n \sum_{j \neq i} \tilde{z}_i^2(k) s_i^2(k) \tilde{z}_j(k) s_j(k) \\ &\quad + 2 \sum_{q=1}^n \tilde{z}_q(k) s_q(k) \\ &\quad \times \sum_{i \neq j} \tilde{z}_i(k) \tilde{z}_j(k) s_i(k) s_j(k) \end{aligned}$$

for all $k \geq 0$. Then, it follows that

$$\begin{aligned} &E\left\{ s_i(k) \left(\sum_{j=1}^n \tilde{z}_j(k) s_j(k) \right)^3 \right\} \\ &= E\Bigg\{ s_i(k) \sum_{i=1}^n \tilde{z}_i^3(k) s_i^3(k) \\ &\qquad + s_i(k) \sum_{i=1}^n \sum_{j \neq i} \tilde{z}_i^2(k) s_i^2(k) \tilde{z}_j(k) s_j(k) \\ &\qquad + 2 s_i(k) \sum_{q=1}^n \tilde{z}_q(k) s_q(k) \\ &\qquad \times \sum_{i \neq j} \tilde{z}_i(k) \tilde{z}_j(k) s_i(k) s_j(k) \Bigg\} \\ &= E\left\{ s_i^4(k) \right\} z_i^3(k) \\ &\quad + E\left\{ s_i(k) \sum_{i=1}^n \sum_{j \neq i} \tilde{z}_i^2(k) s_i^2(k) \tilde{z}_j(k) s_j(k) \right\} \\ &\quad + 2E\left\{ s_i(k) \sum_{q=1}^n \tilde{z}_q(k) s_q(k) \sum_{i \neq j} \tilde{z}_i(k) \tilde{z}_j(k) s_i(k) s_j(k) \right\}. \end{aligned}$$

Since s_i $(i = 1, \cdots, n)$ are mutually independent, it holds that $E\{s_i^2 s_j^2\} = 1$ and $E\{s_i^3 s_j\} = E\{S_i^2 s_j s_l\} = E\{s_i^2 s_j s_l s_m\}$ for four different indexes i, j, l, m [15]. It follows that

$$\begin{aligned} E\left\{ s_i(k) \left(\sum_{j=1}^n \tilde{z}_j(k) s_j(k) \right)^3 \right\} &= E\left\{ s_i^4(k) \right\} z_i^3(k) \\ &\quad + \sum_{j \neq i} z_j^2(k) z_i^2(k) + 2 \sum_{j \neq i} z_j^2(k) z_i(k) \end{aligned}$$

for all $k \geq 0$. Since $\mathrm{Kurt}(s_i) = E\{s_i^4(k)\} - 3$, it follows that

$$\begin{aligned} &E\left\{ s_i(k) \left(\sum_{j=1}^n \tilde{z}_j(k) s_j(k) \right)^3 \right\} \\ &= \mathrm{Kurt}(s_i) z_i^3(k) + 3 z_i^3(k) + 3 \sum_{j \neq i} z_j^2(k) z_i(k) \\ &= \mathrm{Kurt}(s_i) z_i^3(k) + 3 z_i^3(k) + 3 \sum_{j \neq i} z_j^2(k) z_i(k) \\ &= \mathrm{Kurt}(s_i) z_i^3(k) + 3 \|\mathbf{z}(k)\|^2 z_i(k) \end{aligned}$$

for all $k \geq 0$. The proof is complete.

By Lemma above, for any component $z_i(k+1)$ of $\mathbf{z}(k+1)$, it holds that

$$\begin{aligned} z_i(k+1) &= z_i(k) + \eta \Big[b\mathrm{Kurt}(s_i) z_i(k)^3 + 3b \|\mathbf{z}(k)\|^2 z_i(k) \\ &\qquad + a \left(1 - \|\mathbf{z}(k)\|^2 \right) z_i(k) \Big] \\ &= \Big[1 + \eta b \mathrm{Kurt}(s_i) z_i(k)^2 + 3\eta b \|\mathbf{z}(k)\|^2 \\ &\qquad + \eta a \left(1 - \|\mathbf{z}(k)\|^2 \right) \Big] z_i(k) \\ &= \Big[1 + \eta a - \eta a \|\mathbf{z}(k)\|^2 \\ &\qquad + \eta b \left(\mathrm{Kurt}(s_i) z_i(k)^2 + 3 \|\mathbf{z}(k)\|^2 \right) \Big] z_i(k) \end{aligned} \tag{10}$$

for all $k \geq 0$.

Lemma 5: Suppose that $\mathbf{z}(0) \in S_{\tilde{z}1}$ and $\Gamma_{\tilde{z}} > \mathrm{Kurt}(s_i) + 3$. For any $i \in T$, it holds that

$$1 + \eta a - \eta a \|\mathbf{z}(k)\|^2 + \eta b \left(\mathrm{Kurt}(s_i) z_i(k)^2 + 3 \|\mathbf{z}(k)\|^2 \right) > 0$$

for all $k \geq 0$.

Proof: It follows that

$$\begin{aligned} &1 + \eta a - \eta a \|\mathbf{z}(k)\|^2 + \eta b \left(\mathrm{Kurt}\left(s_i(k)\right) z_i(k)^2 + 3 \|\mathbf{z}(k)\|^2 \right) \\ &= 1 + \eta a - \eta \left[a - b \left(\mathrm{Kurt}\left(s_i(k)\right) \frac{z_i(k)^2}{\|\mathbf{z}(k)\|^2} + 3 \right) \right] \|\mathbf{z}(k)\|^2 \end{aligned}$$

for all $k \geq 0$. Since $\Gamma_{\tilde{z}} > \mathrm{Kurt}(s_i) + 3$ and $\mathbf{z}(0) \in S_{\tilde{z}1}$, by Theorem 2, it follows that

$$1+\eta a-\eta a\left\|\mathbf{z}(k)\right\|^2+\eta b\Big(\mathrm{Kurt}\left(s_i(k)\right)z_i(k)^2+3\left\|\mathbf{z}(k)\right\|^2\Big)>0$$

for all $k \geq 0$. The proof is completed.

Lemma 6: Suppose that $\mathbf{z}(0) \in S_{\tilde{z}1}$, $\mathbf{z}(0) \neq V_q^{\perp}$ and $\Gamma_{\tilde{z}} > \mathrm{Kurt}(s_i) + 3$. For any $z_i(k) \neq 0$, $i \in P$, if $z_i^2(k) > \bar{\rho} > 0$ for all $k \geq 0$, it holds that

$$|z_i(k+1)|^2 \leq |z_j(k+1)|^2 \cdot \left[\frac{z_i(0)}{z_j(0)}\right]^2 \cdot e^{-\theta_1 k} < \Pi_1 \cdot e^{-\theta_1 k}$$

where $j \in Q$ and

$$\Pi_1 = \frac{1+\eta a}{\eta a} \cdot \left[\frac{z_i(0)}{z_j(0)}\right]^2,$$
$$\theta_1 = ln\left(\frac{(1+\eta a)(a+3b)}{(1+\eta a)(a+3b)+\eta ab\bar{\rho}\mathrm{Kurt}(s_i)}\right)^2 > 0.$$

Proof: Since $\mathbf{z}(0) \notin V_q^{\perp}$, there must exist $j \in Q$ so that $z_j(0) \neq 0$. For any $z_i(k) \neq 0$, $i \in P$ and $z_i^2(k) > \bar{\rho} > 0$ for all $k \geq 0$, from (10), it follows that

$$\begin{aligned}\left[\frac{z_i(k+1)}{z_j(k+1)}\right]^2 &= \left[\frac{L(k)+\eta b\left(\mathrm{Kurt}\left(s_i(k)\right)z_i(k)^2+3\left\|\mathbf{z}(k)\right\|^2\right)}{L(k)+\eta b\left(\mathrm{Kurt}\left(s_j(k)\right)z_j(k)^2+3\left\|\mathbf{z}(k)\right\|^2\right)}\right]^2 \\ &\quad\cdot\left[\frac{z_i(k)}{z_j(k)}\right]^2\end{aligned}$$

where $L(k) = 1 + \eta a - \eta a\|\mathbf{z}(k)\|^2$, for all $k \geq 0$. By Lemma 5 and $i \in P$, i.e., $\mathrm{Kurt}(s_i) < 0$, it follows that

$$\begin{aligned}\left[\frac{z_i(k+1)}{z_j(k+1)}\right]^2 &\leq \left[\frac{1+\eta a-\eta a\left\|\mathbf{z}(k)\right\|^2+\eta b\left(\mathrm{Kurt}(s_i)\bar{\rho}+3\left\|\mathbf{z}(k)\right\|^2\right)}{1+\eta a-\eta a\left\|\mathbf{z}(k)\right\|^2+3\eta b\left\|\mathbf{z}(k)\right\|^2}\right]^2 \\ &\quad\cdot\left[\frac{z_i(k)}{z_j(k)}\right]^2 \\ &\leq \left[\frac{1+\eta a+\eta b\bar{\rho}\mathrm{Kurt}(s_i)+3\eta b\left\|\mathbf{z}(k)\right\|^2}{1+\eta a+3\eta b\left\|\mathbf{z}(k)\right\|^2}\right]^2\cdot\left[\frac{z_i(k)}{z_j(k)}\right]^2\end{aligned}$$

for all $k \geq 0$. Since $\|\mathbf{z}(k)\|^2 < (1 + \eta a)/\eta a$, by Lemma 2, it follows that

$$\begin{aligned}\left[\frac{z_i(k+1)}{z_j(k+1)}\right]^2 &\leq \left[\frac{1+\eta a+\eta b\bar{\rho}\mathrm{Kurt}(s_i)+3\eta b\cdot(1+\eta a)/\eta a}{1+\eta a+3\eta b\cdot(1+\eta a)/\eta a}\right]^2 \\ &\quad\cdot\left[\frac{z_i(k)}{z_j(k)}\right]^2 \\ &\leq \left[\frac{(1+\eta a)(a+3b)+\eta ab\bar{\rho}\mathrm{Kurt}(s_i)}{(1+\eta a)(a+3b)}\right]^2\cdot\left[\frac{z_i(k)}{z_j(k)}\right]^2 \\ &\leq \left[\frac{(1+\eta a)(a+3b)+\eta ab\bar{\rho}\mathrm{Kurt}(s_i)}{(1+\eta a)(a+3b)}\right]^{2(k+1)}\cdot\left[\frac{z_i(0)}{z_j(0)}\right]^2 \\ &= \left[\frac{z_i(0)}{z_j(0)}\right]^2\cdot e^{-\theta_1(k+1)}\end{aligned}$$

for all $k \geq 0$, where

$$\theta_1 = ln\left(\frac{(1+\eta a)(a+3b)}{(1+\eta a)(a+3b)+\eta ab\bar{\rho}\mathrm{Kurt}(s_i)}\right)^2 > 0.$$

Since $\mathbf{z}(0) \in S_{\tilde{z}1}$, it holds that

$$|z_i(k+1)|^2 \leq |z_j(k+1)|^2 \cdot \left[\frac{z_i(0)}{z_j(0)}\right]^2 \cdot e^{-\theta_1 k} < \Pi_1 \cdot e^{-\theta_1 k}$$

where

$$\Pi_1 = \frac{1+\eta a}{\eta a} \cdot \left[\frac{z_i(0)}{z_j(0)}\right]^2.$$

The proof is complete.

Lemma 7: Suppose that $\mathrm{Kurt}(s_{j_0}) > 0$, $(j_0 \in Q)$ and $\mathrm{Kurt}(s_j) \geq 0$, $j \in (Q \cup O)/j_0$. If $\mathrm{Kurt}(s_{j_0})z_{j_0}^2(0) > \mathrm{Kurt}(s_j)z_j^2(0) \geq 0$, there must exist a constant $\xi > 0$ so that $0 < \psi(k) < \xi < 1$ for all $k \geq 0$, where

$$\psi(k) = \left[\frac{(1+\eta a)(a+3b)+\eta ab\mathrm{Kurt}(s_j)z_j^2(k)}{(1+\eta a)(a+3b)+\eta ab\mathrm{Kurt}\left(s_{j_0}\right)z_{j_0}^2(k)}\right]^2.$$

Proof: For $j_0 \in Q$ and any $j \in (Q \cup O)/j_0$, from (10), it follows that

$$\frac{z_j^2(k+1)}{z_{j_0}^2(k+1)} = \gamma(k)\cdot\frac{z_j^2(k)}{z_{j_0}^2(k)} \tag{11}$$

for all $k \geq 0$, where (see the equation at the bottom of the page). If $\mathrm{Kurt}(s_{j_0})z_{j_0}^2(k) > \mathrm{Kurt}(s_j)z_j^2(k) \geq 0$, by Lemma 5, it holds that $0 < \gamma(k) < 1$. From (11), it holds that

$$z_{j_0}^2(k+1)z_j^2(k) > z_j^2(k+1)z_{j_0}^2(k) \geq 0.$$

$$\gamma(k) = \left[\frac{1+\eta a-\eta a\left\|\mathbf{z}(k)\right\|^2+\eta b\left(\mathrm{Kurt}(s_j)z_j^2(k)+3\left\|\mathbf{z}(k)\right\|^2\right)}{1+\eta a-\eta a\left\|\mathbf{z}(k)\right\|^2+\eta b\left(\mathrm{Kurt}\left(s_{j_0}\right)z_{j_0}^2(k)+3\left\|\mathbf{z}(k)\right\|^2\right)}\right]^2.$$

So, it is easy to see that

$$\text{Kurt}\,(s_{j_0})\,z_{j_0}^2(k+1) > \text{Kurt}(s_j)z_j^2(k+1) \geq 0.$$

Thus, if $\text{Kurt}(s_{j_0})z_{j_0}^2(0) > \text{Kurt}(s_j)z_j^2(0) \geq 0$, it holds that

$$\text{Kurt}\,(s_{j_0})\,z_{j_0}^2(k+1) > \text{Kurt}(s_j)z_j^2(k+1) \tag{12}$$

for all $k \geq 0$.

Now, we will prove that if $\text{Kurt}(s_{j_0})z_{j_0}^2(0) > \text{Kurt}(s_j)z_j^2(0) \geq 0$, there must exist a small constant $\epsilon > 0$ so that $z_{j_0}^2(k) > \epsilon$ for all $k \geq 0$. Assume that this is not true. Then, it holds that $\lim_{k\to+\infty} z_{j_0}^2(k) = 0, j_0 \in Q$. From (12), it follows that $\lim_{k\to+\infty} z_j^2(k) = 0, j \in (Q\cup O)/j_0$. By Lemma 6, it holds that $\lim_{k\to+\infty} \|\mathbf{z}(k)\|^2 = 0$. This contradicts with Theorem 2.

Since $\text{Kurt}(s_{j_0})z_{j_0}^2(0) > \text{Kurt}(s_j)z_j^2(0) \geq 0$ and $0 < \gamma(k) < 1$ for all $k \geq 0$. From (11), it follows that

$$\begin{aligned}
&\text{Kurt}\,(s_{j_0})\,z_{j_0}^2(k+1) - \text{Kurt}(s_j)z_j^2(k+1)\\
&= \text{Kurt}\,(s_{j_0})z_{j_0}^2(k+1)\left(1-\gamma(k)\frac{\text{Kurt}(s_j)z_j^2(k)}{\text{Kurt}\,(s_{j_0})\,z_{j_0}^2(k)}\right)\\
&= \text{Kurt}\,(s_{j_0})z_{j_0}^2(k+1)\left(1-\prod_{l=1}^{k}\gamma(k)\cdot\frac{\text{Kurt}(s_j)z_j^2(0)}{\text{Kurt}\,(s_{j_0})\,z_{j_0}^2(0)}\right)\\
&> \text{Kurt}\,(s_{j_0})z_{j_0}^2(k+1)\left(1-\frac{\text{Kurt}(s_j)z_j^2(0)}{\text{Kurt}\,(s_{j_0})\,z_{j_0}^2(0)}\right)
\end{aligned}$$

for all $k \geq 0$. It follows that

$$\eta ab\text{Kurt}\,(s_{j_0})\,z_{j_0}^2(k+1) - \eta ab\text{Kurt}(s_j)z_j^2(k+1) > \bar{\epsilon} > 0$$

where

$$\bar{\epsilon} = \eta ab\text{Kurt}\,(s_{j_0})\,\epsilon\left(1-\frac{\text{Kurt}(s_j)z_j^2(0)}{\text{Kurt}\,(s_{j_0})\,z_{j_0}^2(0)}\right).$$

It is easy to show that

$$\psi(k) = \frac{(1+\eta a)(a+3b) + \eta ab\text{Kurt}(s_j)z_j^2(k)}{(1+\eta a)(a+3b) + \eta ab\text{Kurt}\,(s_{j_0})\,z_{j_0}^2(k)} < \xi$$

for all $k \geq 0$, where

$$\xi = 1 - \frac{a\bar{\epsilon}}{(1+\eta a)\left[a(a+3b) + \eta ab\text{Kurt}\,(s_{j_0})\right]} < 1.$$

The proof is complete.

Lemma 8: Suppose that $\mathbf{z}(0) \in S_{\tilde{z}1}$, $\mathbf{z}(0) \notin V_q^{\perp}$ and $\Gamma_{\tilde{z}} > \text{Kurt}(s_i) + 3$. If $\text{Kurt}(s_{j_0})z_{j_0}^2(0) > \text{Kurt}(s_j)z_j^2(0) \geq 0$, $j_0 \in Q$ and $j \in (Q\cup O)/j_0$, it holds that

$$|z_j(k+1)|^2 \leq |z_{j_0}(k+1)|^2 \cdot \left[\frac{z_j(0)}{z_{j_0}(0)}\right]^2 \cdot e^{-\theta_2 k} < \Pi_2 \cdot e^{-\theta_2 k}$$

for all $j \in (Q\cup O)/j_0$, where

$$\Pi_2 = \frac{1+\eta a}{\eta a}\cdot\left[\frac{z_j(0)}{z_{j_0}(0)}\right]^2 \text{ and } \theta_2 = ln\left[\frac{1}{\xi}\right]^2 > 0.$$

Proof: Since $\mathbf{z}(0) \notin V_q^{\perp}$, there must exist $j_0 (j_0 \in Q)$ so that $z_j(0) \neq 0$ and $\text{Kurt}(s_{j_0})z_{j_0}^2(0) > \text{Kurt}(s_j)z_j^2(0), j \in Q/j_0$. From (10), it follows that

$$\begin{aligned}
&\frac{z_j^2(k+1)}{z_{j_0}^2(k+1)}\\
&= \left[\frac{L(k) + \eta b\left(\text{Kurt}\,(s_j(k))\,z_j(k)^2 + 3\,\|\mathbf{z}(k)\|^2\right)}{L(k) + \eta b\left(\text{Kurt}\,(s_{j_0}(k))\,z_{j_0}(k)^2 + 3\,\|\mathbf{z}(k)\|^2\right)}\right]^2\\
&\quad\cdot\frac{z_j^2(k)}{z_{j_0}^2(k)}
\end{aligned}$$

where $L(k) = 1 + \eta a - \eta a\|\mathbf{z}(k)\|^2$. By Lemmas 2 and 7, it holds that

$$\frac{z_j^2(k+1)}{z_{j_0}^2(k+1)} \leq \xi\cdot\frac{z_j^2(k)}{z_{j_0}^2(k)}$$

for all $k \geq 0$. Thus, it follows that

$$|z_j(k+1)|^2 \leq |z_{j_0}(k+1)|^2 \cdot \left[\frac{z_j(0)}{z_{j_0}(0)}\right]^2 \cdot e^{-\bar{\theta}_1 k} < \bar{\Pi}_1 \cdot e^{-\bar{\theta}_1 k}$$

for all $j \in (Q\cup O)/j_0$, where

$$\Pi_2 = \frac{1+\eta a}{\eta a}\cdot\left[\frac{z_j(0)}{z_{j_0}(0)}\right]^2 \text{ and } \theta_2 = ln\left[\frac{1}{\xi}\right]^2 > 0.$$

The proof is complete.

Lemma 9: Suppose that $\mathbf{z}(0) \in S_{\tilde{z}1}$, $\mathbf{z}(0) \notin V_q^{\perp}$ and $\Gamma_{\tilde{z}} > \text{Kurt}(s_i) + 3$. If $\text{Kurt}(s_{j_0})z_{j_0}^2(0) > \text{Kurt}(s_i)z_i^2(0)$, $j_0 \in Q$ and $i \in T/j_0$, it holds that

$$\eta a - V(k+1) \leq \zeta^k \cdot |\eta a - V(0)|$$

where $V(k) = \eta[a - b(\text{Kurt}(s_{j_0}) + 3)]z_{j_0}^2(k)$ and $\zeta = \max\{|(1-\gamma)^2 - \eta a\gamma, |1 - 2\eta a|\}$, when k is large enough.

Proof: By Lemmas 6 and 8, it follows that

$$\begin{aligned}
&\lim_{k\to+\infty} z_i(k) = 0,\ (i \in P \text{ and } \text{Kurt}(s_i) < 0)\\
&\lim_{k\to+\infty} z_j(k) = 0,\ (j \in (Q\cup O)/j_0 \text{ and } \text{Kurt}(s_j) \geq 0)
\end{aligned}$$

where $j_0 \in Q, \text{Kurt}(s_{j_0})z_{j_0}^2(0) > \text{Kurt}(s_j)z_j^2(0)$. From (10), when k is large enough, it follows that

$$\begin{aligned}
&z_{j_0}(k+1)\\
&= z_{j_0}(k) + \left[\eta a - \eta a z_{j_0}^2(k)\right.\\
&\qquad\left. + \eta b\left(\text{Kurt}\,(s_{j_0})\,z_{j_0}^2(k) + 3z_{j_0}^2(k)\right)\right] z_{j_0}(k)\\
&= \left\{1 + \eta a - \eta\left[a - b\left(\text{Kurt}\,(s_{j_0}) + 3\right)\right] z_{j_0}^2(k)\right\} z_{j_0}(k).
\end{aligned}$$

Since $\mathbf{z}(0) \in S_{\tilde{z}1}$, there must exist a small constant ϱ so that $0 < \varrho < 1$ and $\varrho \leq V(k)$. Thus, it can be checked that

$$\begin{aligned} 0 < \varrho < V(k) &< \eta\left[a - b\left(\mathrm{Kurt}\left(s_{j_0}\right) + 3\right)\right] \frac{1+\eta a}{\eta a} \\ &= \frac{a - b\left(\mathrm{Kurt}\left(s_{j_0}\right)+3\right)}{a}(1+\eta a) \\ &\leq 2 \end{aligned}$$

where $0 < \eta a < 1$ and $a > b(\mathrm{Kurt}(s_{j_0}) + 3)$. Then, it follows that

$$\begin{aligned} &\eta a - V(k+1) \\ &\quad = \eta a - \eta\left[a - b\left(\mathrm{Kurt}\left(s_{j_0}\right)+3\right)\right] z_{j_0}^2(k+1) \\ &\quad = \eta a - \eta\left[a - b\left(\mathrm{Kurt}\left(s_{j_0}\right)+3\right)\right] \\ &\qquad \cdot \left\{1 + \eta a - \eta\left[a - b\left(\mathrm{Kurt}\left(s_{j_0}\right)+3\right)\right] z_{j_0}^2(k)\right\}^2 z_{j_0}^2(k). \end{aligned}$$

It is easy to get that

$$\begin{aligned} \eta a - V(k+1) &= [\eta a - V(k)]\left\{[1 - V(k)]^2 - \eta a V(k)\right\} \\ &\leq |\eta a - V(k)| \left|[1 - V(k)]^2 - \eta a V(k)\right|. \end{aligned}$$

By Lemma 3, It follows that

$$\begin{aligned} |\eta a - V(k+1)| &\leq |\eta a - V(k)| \cdot \left|[1 - V(k)]^2 - \eta a V(k)\right| \\ &\leq |\eta a - V(k)| \cdot \zeta \\ &= \zeta^k \left|\eta a - V(0)\right|, \end{aligned}$$

where $0 < \zeta < 1$. The proof is completed.

Proof of Theorem (3): By Lemmas 6 and 8, when k is large enough, it follows that

$$\begin{aligned} z_{j_0}(k+1) = z_{j_0}(k) &+ \left[\eta a - \eta a z_{j_0}^2(k)\right. \\ &\left. + \eta b\left(\mathrm{Kurt}\left(s_{j_0}\right) z_{j_0}^2(k) + 3 z_{j_0}^2(k)\right)\right] z_{j_0}(k). \end{aligned}$$

Give any $\tau > 0$, there exists a $K \geq 1$ such that

$$|\eta a - V(0)| \sqrt{\frac{1+\eta a}{\eta a}} \frac{\zeta^K}{1-\zeta} < \tau.$$

For any $k_1 > k_2 \geq K$, it follows that

$$\begin{aligned} \left|z_{j_0}(k_1) - z_{j_0}(k_2)\right| &= \left|\sum_{i=k_2}^{k_1-1}\left(z_{j_0}(i+1) - z_{j_0}(i)\right)\right| \\ &= \left|\sum_{i=k_2}^{k_1-1}\left[\left(\eta a - V(i)\right) z_{j_0}(i)\right]\right| \\ &\leq \sum_{i=k_2}^{k_1-1} |\eta a - V(i)| \left|z_{j_0}(i)\right| \\ &\leq \sqrt{\frac{1+\eta a}{\eta a}} \sum_{i=k_2}^{k_1-1} |\eta a - V(i)|. \end{aligned}$$

By Lemma 9, it holds that

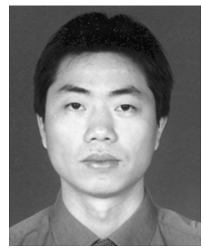
$$\begin{aligned} \left|z_{j_0}(k_1) - z_{j_0}(k_2)\right| &\leq \sqrt{\frac{1+\eta a}{\eta a}} \sum_{i=k_2}^{\infty} |\eta a - V(0)| \zeta^{i-1} \\ &\leq |\eta a - V(0)| \sqrt{\frac{1+\eta a}{\eta a}} \frac{\zeta^{k_2-1}}{1-\zeta} \\ &\leq \tau. \end{aligned}$$

This shows the sequence $\{z_{j_0}(k)\}$ is a Cauchy sequence. By Cauchy convergence principle, there must exist a $z_{j_0}^*$ so that $\lim_{k\to+\infty} z_{j_0}(k) = z_{j_0}^*$. The proof is complete.

REFERENCES

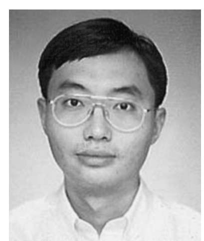
[1] T. Akuzawa, "Nested Newton's method for ICA and post factor analysis," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 839–852, Mar. 2003.
[2] A. Bell and T. J. Sejnowski, "Edges are the independent components of natural scenes," in *Advances in Neural Information Processing 9 (NIPS 96)*. Cambridge, MA: MIT, 1997, pp. 831–837.
[3] A. Benveniste *et al., Adaptive Algorithms and Stochastic Approximations*. Berlin, Germany: Springer-Verlag, 1991.
[4] D. Blanco and B. Mulgrew, "ICA in signals with multiplicative noise," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2648–2657, Aug. 2005.
[5] D. Blanco, D. P. Ruiz, E. A. Hernández, and M. C. Carrión, "A fourth-order stationarity and ergodicity conditions for harmonic processes," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1641–1649, Jun. 2004.
[6] T. Blaschke and L. Wiskott, "CuBICA: Independent component analysis by simultaneous third-and fourth-order cumulant diagonalization," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1250–1256, May 2004.
[7] C. Chatterjee, "Adaptive algorithm for first principal eigenvector computation," *IEEE Trans. Neural Netw.*, vol. 18, no. 2, pp. 145–159, 2005.
[8] A. Chen and P. J. Bickel, "Consistent independent component analysis and prewhitening," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3625–3632, Oct. 2005.
[9] H. Chen and R. W. Liu, "An on-line unsupervised learning machine for adaptive feature extraction," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 41, no. 1, pp. 87–98, 1994.
[10] P. Comon, "Independent component analysis—A new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.
[11] K. I. Diamantaras and T. Papadimitriou, "Subspace-based channel shortening for the blind separation of convolutive mixtures," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3669–3676, Oct. 2006.
[12] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*. New York: Academic, 1981, pp. 565–608.
[13] A. Hyvärinen, J. Karhunen, and E. Oja*, Independent Component Analysis*. New York: Wiley-Interscience, 2001.
[14] A. Hyvärinen and E. Oja, "Simple neuron models for independent component analysis," *Int. J. Neural Syst.*, vol. 7, no. 6, pp. 671–687, 1996.
[15] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, pp. 1483–1492, 1997.
[16] A. Hyvärinen and E. Oja, "Independent component analysis by general non-linear Hebbian-like learning rules," *Signal Process.*, vol. 64, no. 3, pp. 301–313, 1998.
[17] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, pp. 1–10, 1991.
[18] J. Karhunen, A. Hyvärinen, R. Vigario, J. Hurri, and E. Oja, "Applications of neural blind separation to signal and image processing," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Munich, Germany, 1997, pp. 131–134.
[19] J. Karhunen, E. Oja, L. Wang, R. Vigario, and J. Joutsensalo, "A class of neural networks for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 8, no. 3, pp. 486–504, 1997.
[20] C. J. Ku and T. L. Fine, "Testing for stochastic independence: Application to blind source separation," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1815–1826, May 2005.
[21] L. D. Lathauwer, J. Castaing, and J. F. Cardoso, "Fourth-order cumulant-based blind identification of underdetermined mixtures," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2965–2973, Jun. 2007.
[22] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "Independent component analysis and (simultaneous) third-order tensor diagonalization," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2262–2271, Oct. 2001.

[23] L. Ljung, "Analysis of recursive stochastic algorithm," *IEEE Trans. Autom. Control*, vol. 22, no. 4, pp. 551–575, 1977.
[24] J. C. Lv, Z. Yi, and K. K. Tan, "Global convergence of GHA learning algorithm with nonzero-approaching adaptive learning rates," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1557–1571, 2007.
[25] J. C. Lv, Z. Yi, and K. K. Tan, "Convergence analysis of Xu's LMSER learning algorithm via deterministic discrete time system method," *Neurocomput.*, vol. 7, pp. 362–372, 2006.
[26] J. C. Lv, Z. Yi, and K. K. Tan, "Global convergence of Oja's PCA learning algorithm with a non-zero-approaching adaptive learning rate," *Theoretic. Comput. Sci.*, vol. 367, pp. 286–307, 2006.
[27] W. B. Mikhael and T. Yang, "A gradient-based optimum block adaptation ICA technique for interference suppression in highly dynamic communication channels," *EURASIP J. Appl. Signal Process.*, pp. 1–10, 2006.
[28] E. Oja and Z. Yuan, "The FastICA algorithm revisited: Convergence analysis," *IEEE Trans. Neural Netw.*, vol. 17, no. 6, pp. 1370–1381, 2006.
[29] E. Oja, "Principal components, minor components, and linear neural networks," *Neural Netw.*, vol. 5, pp. 927–935, 1992.
[30] E. Oja and J. Karhunen, "On stochastic approximation of the eigenvector and eigenvalues of the expectation of a random matrix," *J. Math. Anal. App.*, vol. 106, no. 1, pp. 69–84, 1985.
[31] T. H. Sander, M. Burghoff, G. Curio, and L. Trahms, "Single evoked somatosensory MEG responses extracted by time delayed decorrelation," *IEEE Trans. Signal Process.*, vol. 53, no. 9, pp. 3384–3392, Sep. 2005.
[32] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Inf. Theory*, vol. 36, pp. 312–321, 1990.
[33] V. Solo and X. Kong, "Performance analysis of adaptive eigenanalysis algorithms," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 636–646, Mar. 1998.
[34] J. Thomas, Y. Deville, and S. Hosseini, "Differential fast fixed-point algorithms for underdetermined instantaneous and convolutive partial blind source separation," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3717–3729, Jul. 2007.
[35] P. Tichavský, Z. Koldovský, and E. Oja, "Performance analysis of the FastICA algorithm and Cramér–Rao bounds for linear independent component analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1189–1203, Apr. 2006.
[36] J. M. Vegas and P. J. Zufiria, "Generalized neural networks for spectral analysis: Dynamics and Liapunov functions," *Neural Netw.*, vol. 17, pp. 233–245, 2004.
[37] G. Wang, X. Xu, and D. Hu, "Global convergence of FastICA: Theoretical analysis and practical considerations," *Lecture Notes in Computer Science, Advances in Natural Computation*, vol. 3610, pp. 700–705, 2005.
[38] S. L. Xie, Z. S. He, and Y. L. Fu, "A note on Stone's conjecture of blind signal separation," *Neural Comput.*, vol. 17, no. 2, pp. 321–330, 2005.
[39] L. Xu, "Temporal BYY learning for state space approach, hidden Markov model, and blind source separation," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 2132–2144, Jul. 2000.
[40] M. Ye, "Global convergence analysis of a discrete time nonnegative ICA algorithm," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 253–256, 2006.
[41] Z. Yi and K. K. Tan, *Convergence Analysis of Recurrent Neural Networks*. Norwell, MA: Kluwer, 2004.
[42] Z. Yi, M. Ye, J. C. Lv, and K. K. Tan, "Convergence analysis of a deterministic discrete time system of Oja's PCA learning algorithm," *IEEE Trans. Neural Netw.*, vol. 16, no. 6, pp. 1318–1328, Nov. 2005.
[43] Z. Yi and K. K. Tan, "Global convergence of Lotka-Volterra recurrent neural networks with delays," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 11, pp. 2482–2489, Nov. 2005.
[44] V. Zarzoso, J. J. Murillo-Fuentes, R. Boloix-Tortosa, and A. K. Nandi, "Optimal pairwise fourth-order independent component analysis," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 3049–3063, Aug. 2006.
[45] V. Zarzoso and P. Comon, "How fast is FastICA?," presented at the XIV Eur. Signal Processing Conf. (EUSIPCO), Florence, Italy, Sep. 2–4, 2006.
[46] V. Zarzoso and P. Comon, "Comparative speed analysis of FastICA," presented at the 7th Int. Conf. Independent Component Analysis Signal Separation (ICA), London, U.K., Sep. 9–21, 2007.
[47] Q. Zhang, "On the discrete-time dynamics of a PCA learning algorithm," *Neurocomput.*, vol. 55, pp. 761–769, 2003.
[48] P. J. Zufiria, "On the discrete-time dynamic of the basic Hebbian neural-network nods," *IEEE Trans. Neural Networks*, vol. 13, no. 6, pp. 1342–1252, 2002.



Jian Cheng Lv received the Ph.D. degree in computer science and engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2006.

He was previously a Research Fellow in the Department of Electrical and Computer Engineering, National University of Singapore. He is currently a Professor at the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China. His research interests include machine learning, neural networks and data mining.



Kok Kiong Tan received the Ph.D. degree from the National University of Singapore in 1995.

He was previously a Research Fellow at SIMTech, where he has been involved in managing industrial projects. He is currently an Associate Professor with the National University of Singapore. His current research interests are in precision motion control and instrumentation, advanced process control and autotuning, and general industrial automation.

Zhang Yi received the Ph.D. degree in mathematics from the Institute of Mathematics, The Chinese Academy of Science, Beijing, China, in 1994.

Currently, he is a Professor at the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China. He is the coauthor of the books *Convergence Analysis of Recurrent Neural Networks* (Boston, MA; Kluwer, 2004) and *Neural Networks: Computational Models and Applications* (Berlin, Germany: Springer, 2007). His current research interests include neural networks, data mining and machine learning.

Sunan Huang received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, 1994.

Since 1997, he has been a Research Fellow in the Department of Electrical and Computer Engineering, National University of Singapore. His research interests include error compensation of high-precision machine, adaptive control, and neural network control.