Available online at www.sciencedirect.com
ScienceDirect
NEUROCOMPUTING
Neurocomputing 70 (2006) 362–372
www.elsevier.com/locate/neucom

Convergence analysis of Xu's LMSER learning algorithm via deterministic discrete time system method ☆

Jian Cheng Lv[a], Zhang Yi[a,*], K.K. Tan[b]

[a] Computational Intelligence Laboratory, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, PR China

[b] Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576, Singapore

Received 9 June 2005; received in revised form 5 January 2006; accepted 5 January 2006
Communicated by L. Xu
Available online 15 May 2006

Abstract

The convergence of Xu's LMSER algorithm with a constant learning rate, which is in the one unit case, is interpreted by analyzing an associated deterministic discrete time (DDT) system. Some convergent results relating to the Xu's DDT system are obtained. An invariant set and an ultimate bound are identified so that the non-divergence of the system can be guaranteed. It is rigorously proven that all trajectories of the system from points in this invariant set will converge exponentially to a unit eigenvector associated with the largest eigenvalue of the correlation matrix. By comparing Xu's algorithm with Oja's algorithm, it can be observed, on the whole, the Xu's algorithm evolves faster at a cost of larger computational complexity. Extensive simulations will be carried out to illustrate the theory.

Keywords: Principal component analysis; Oja's learning algorithm; Xu's learning algorithm; Deterministic discrete time system

1. Introduction

Principal component analysis (PCA) neural networks are useful tools in feature extraction, data compression, pattern recognition and time series prediction, especially in online data processing applications. Stemming from Oja's algorithm [13], many PCA algorithms have been proposed to update the weights of these networks [1,3,8,11–13,16,17,20,24,25]. Among the algorithms for PCA, Oja's algorithm and Xu's LMSER algorithm are commonly used in these applications. Several other algorithms for PCA are related to these basic procedures [5]. Recently, the approach to study an associated DDT system as a means to interpret the convergence of Oja's algorithm has been attracting much attention [23,26,29]. However, no work has been done so far on the study of Xu's DDT system. This paper will interpret the convergence of Xu's LMSER algorithm, which is in the one unit case, by studying an associated DDT system.

All of these PCA learning algorithms are described by stochastic discrete time (SDT) systems. It is very difficult to study the convergence of the SDT models directly [26]. Traditionally, based on the stochastic approximation theorem, the convergence of SDT algorithms is interpreted indirectly by analyzing corresponding deterministic continuous time (DCT) systems, see for example, [2,4–6,9,10,15,14,18,19,21,22,28,27]. In order to relate a SDT model to a DCT system, the stochastic approximation theorem requires that some restrictive conditions must be satisfied. One important condition is that the learning rate must converge to zero. However, in practical applications, it is difficult to satisfy these conditions, because of computational round-off limitations and tracking requirements. Usually, a constant learning rate is required. More discussions can be found in [7,23,26,29].

To overcome the shortcoming of the DCT method in practical applications, the convergence of original algorithm can be interpreted by analyzing a DDT systems

☆This work was supported by National Science Foundation of China under Grant 60471055 and Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20040614017.

*Corresponding author. Tel./fax: +86 288 3208 191.
E-mail addresses: lvjch@uestc.edu.cn (Jian Cheng Lv), zhangyi@uestc.edu.cn (Zhang Yi), kktan@nus.edu.sg (K.K. Tan).

0925-2312/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2006.01.003

[7,23,26,29]. Transforming a SDT system into an associated DDT system does not need the restrictive conditions in a DCT approach. Furthermore, a constant learning rate can be used. The DDT system preserves the discrete time nature of the original algorithm and gather a more realistic behavior of the learning gain [29]. It seems more reasonable to study the convergence of an original SDT system by analyzing the corresponding DDT system. In [23], the Oja's DDT system is analyzed in detail and some important results are derived.

In this paper, by using the DDT method, we will discuss the convergence of Xu's LMSER algorithm, in the one unit case, with a constant learning rate. An invariant set and an evolution ultimate bound are identified. A mathematical proof will be provided to prove the convergence. Then, by comparing Oja's algorithm with Xu's algorithm, the computational complexity and convergence rate of both algorithms are discussed. It could be observed, on the whole, the Xu's algorithm converges faster or diverges more rapidly with time, albeit at a cost of higher computational complexity.

The rest of this paper is organized as follows. Section 2 describes the formulation and preliminaries. In Section 3, the convergence results are obtained. An invariant set and an ultimate bound are derived. A mathematic proof is given to prove the convergence. Simulation results and discussions will be provided in Section 4. Conclusions are drawn in Section 5.

2. Formulation and preliminaries

In [20], based on the least mean square error reconstruction, Xu proposed a learning algorithm to perform the true PCA. Suppose the input sequence $\{x(k)|x(k) \in R^n\ (k = 0, 1, 2, \ldots)\}$ is a zero mean stationary stochastic process, denote by $C = E[x(k)x^T(k)]$, the correlation matrix of the input data set, and let C_k be an online observation of C, then the Xu's learning algorithm can be described by the following stochastic difference equation

$$\begin{aligned} w_l(k+1) = w_l(k) + \eta(k)\Bigg[&2C_k w_l(k) - \sum_{i=1}^{l-1} C_k w_l(k) w_i^T(k) w_i(k) \\ &- C_k(k) w_l(k) w_l^T w_l(k) - \sum_{i=1}^{l-1} w_i(k) w_i(k)^T C_k w_l(k) \\ &- (w_l^T(k) C_k w_l(k)) w_l(k)\Bigg], \end{aligned} \tag{1}$$

for $\eta(k) > 0$, where w_l is the input weight vector for the lth neuron. Obviously, if $l = 1$, the algorithm is as follows:

$$\begin{aligned} w(k+1) = w(k) + \eta(k)[&2C_k w(k) - C_k(k) w(k) w^T(k) w(k) \\ &- (w^T(k) C_k w(k)) w(k)], \end{aligned} \tag{2}$$

for $\eta(k) > 0$. The system (2) approximates the Oja's one unit rule when the normal of weight vector approaches one [20]. However, it shows some different dynamical behaviors. In this paper, by studying an associated DDT system, we will discuss the convergence of (2), which is a special case of the one unit version of Xu's LMSER algorithm. By taking the conditional expectation $E\{w(k+1)/w(0), w(i), i < k\}$, the DDT system of (2) with a constant learning rate is given as follows:

$$\begin{aligned} w(k+1) = w(k) + \eta[&2Cw(k) - C(k) w(k) w^T(k) w(k) \\ &- (w^T(k) C w(k)) w(k)], \end{aligned} \tag{3}$$

for $\eta > 0$ and $C = E[x(k)x^T(k)]$ is the correlation matrix.

It is known that the correlation matrix C is a symmetric non-negative definite matrix. There exists an orthonormal basis of R^n composed by eigenvectors of C. Let $\lambda_1, \ldots, \lambda_n$ be all the eigenvalues of C ordered by $\lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant 0$. Suppose that λ_p is the smallest non-zero eigenvalue, i.e., $\lambda_p > 0$ but $\lambda_j = 0\ (j = p+1, \ldots, n)$. Denote by σ the largest eigenvalue of C. Suppose that the multiplicity of σ is $m\ (1 \leqslant m \leqslant p \leqslant n)$, then,

$$\sigma = \lambda_1 = \cdots = \lambda_m.$$

Suppose that $\{v_i | i = 1, \ldots, n\}$ is an orthonormal basis in R^n such that each v_i is a unit eigenvector of C associated with the eigenvalue λ_i. Denote by V_σ the eigensubspace of the largest eigenvalue σ, i.e.,

$$V_\sigma = span\{v_1, \ldots, v_m\}.$$

Denoting by $V_\sigma^\perp$, the subspace which is perpendicular to V_σ, it follows that

$$V_\sigma^\perp = span\{v_{m+1}, \ldots, v_n\}.$$

Since the vector set $\{v_1, \ldots, v_n\}$ is an orthonormal basis of R^n, for each $k \geqslant 0$, $w(k) \in R^n$ can be represented as

$$w(k) = \sum_{i=1}^{n} z_i(k) v_i, \tag{4}$$

where $z_i(k)\ (i = 1, \ldots, n)$ are some constants, and then

$$Cw(k) = \sum_{i=1}^{n} \lambda_i z_i(k) v_i = \sum_{i=1}^{p} \lambda_i z_i(k) v_i. \tag{5}$$

From (3), it holds that

$$z_i(k+1) = [1 + \eta(\lambda_i(2 - \|w(k)\|^2) - w^T(k) C w(k))] z_i(k), \tag{6}$$

for $k \geqslant 0$, where $i = 1, \ldots, n$.

Definition 1. A point $w^* \in R^n$ is called an equilibrium of (3), if and only if

$$w^* = w^* + \eta\left[2Cw^* - Cw^*(w^*)^T w^* - ((w^*)^T C w^*) w^*\right].$$

Clearly, the set of all equilibrium points of (3) is

{all the eigenvectors with unit length} $\cup \{0\}$.

Definition 2. A compact set $S \subset R^n$ is called an invariant set of (3), if for any $w(0) \in S$, the trajectory of (3) starting from $w(0)$ will remain in S for all $k \geqslant 0$.

Definition 3. A bound M is called an ultimate bound of (3), if for any $w(0) \in S$, there exists a constant N so that $\|w(k)\|^2 < M$ for all $k > N$.

For convenience in analysis, a lemma is presented as follows.

Lemma 1. *It holds that*

$$[1 + \eta(\sigma(2 - s))]^2 s \leqslant \frac{4}{27\eta\sigma} \cdot (1 + 2\eta\sigma)^3$$

for all $s \in [0, 2 + 1/\eta\sigma]$.

See the Appendix for the proof.

3. Invariant set and ultimate bound

The system (3) is not globally convergent. Consider the same example as in [23], which is the one dimensional case,

$$w(k+1) = w(k) + \eta[2 - 2w^2(k)]w(k) \quad \text{for } k \geqslant 0.$$

If $w(k) \geqslant 1 + 1/(2\eta)$, it holds that

$$2\eta w^2(k) - w(k) - 2\eta - 1 \geqslant 0.$$

Then

$$|w(k+1)| = w^2(k) \cdot \frac{2\eta w^2(k) - 2\eta - 1}{w(k)} \geqslant w^2(k).$$

$w(k)$ will clearly tend to infinity if $w(k) \geqslant 1 + 1/(2\eta)$. So, an important problem to address is whether an invariant set can be obtained to guarantee the non-divergence. This is a crucial problem to address to ensure the success of applications.

Lemma 2. *Suppose that* $\eta\sigma \leqslant 0.25$ *and* $0 < \|w(k)\|^2 \leqslant 1 + 1/2\sigma\eta$ *, then*

$$1 + \eta(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)) > 0 \quad (i = 1, \ldots, n)$$

for all $k \geqslant 0$.

See the Appendix for the proof.

Theorem 1. *Denote*

$$S = \left\{ w(k) | w(k) \in R^n, \|w(k)\|^2 \leqslant 1 + \frac{1}{2\sigma\eta} \right\}.$$

If

$\eta\sigma \leqslant 0.25$,

then S is an invariant set of (3).

Proof. From Lemma 2, if $2 \leqslant \|w(k)\|^2 \leqslant 1 + 1/2\sigma\eta$, then

$$0 < 1 + \eta(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)) < 1 \quad (i = 1, \ldots, n)$$

for $k \geqslant 0$. From (4), (6), it follows that

$$\begin{aligned}
&\|w(k+1)\|^2 \\
&\quad = \sum_{i=1}^{n} \left[1 + \eta\left(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 z_i^2(k) \\
&\quad \leqslant \left[1 + \eta\left(\lambda_p(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 \|w(k)\|^2 \\
&\quad \leqslant 1 + \frac{1}{2\sigma\eta}.
\end{aligned}$$

If $\|w(k)\|^2 < 2$, then

$$\begin{aligned}
&\|w(k+1)\|^2 \\
&\quad = \sum_{i=1}^{n} [1 + \eta(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k))]^2 z_i^2(k) \\
&\quad \leqslant [1 + \eta(\sigma(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k))]^2 \|w(k)\|^2 \\
&\quad \leqslant [1 + \eta(\sigma(2 - \|w(k)\|^2))]^2 \|w(k)\|^2 \\
&\quad \leqslant \max_{0 < s < 2} \{[1 + \eta(\sigma(2 - s))]^2 s\}.
\end{aligned}$$

By Lemma 1,

$$\|w(k+1)\|^2 < \frac{4}{27\eta\sigma} \cdot (1 + 2\eta\sigma)^3.$$

If $\eta\sigma < 0.25$, then

$$\frac{4(1 + 2\eta\sigma)^3}{27\sigma\eta} < \frac{1}{2\eta\sigma} < 1 + \frac{1}{2\eta\sigma},$$

The proof is completed. □

Theorem 2. *Suppose that* $\eta\sigma \leqslant 0.25$. *the system* (3) *has an ultimate bound* $2\sigma/(\sigma + \lambda_p)$, *i.e., if* $w(0) \in S$ *and* $w(0) \notin V_\sigma^\perp$, *then there exists a constant N, so that*

$$\|w(k+1)\|^2 \leqslant \frac{2\sigma}{\sigma + \lambda_p} < 2,$$

for all $k > N$.

Proof. From Lemma 2 and Theorem 1, if $w(0) \in S$, we have

$$1 + \eta(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)) > 0 \quad (k \geqslant 0).$$

If $2 \leqslant \|w(k)\|^2 < 1 + 1/2\eta\sigma$, then

$$\begin{aligned}
&1 + \eta(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)) \\
&\quad < 1 + \eta(\lambda_p(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)) \\
&\quad < 1,
\end{aligned}$$

for all $k \geqslant 0$. Thus, it holds that

$$\begin{aligned}
&\|w(k+1)\|^2 \\
&\quad = \sum_{i=1}^{n} \left[1 + \eta\left(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 z_i^2(k) \\
&\quad \leqslant \left[1 + \eta\left(\lambda_p(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 \|w(k)\|^2 \\
&\quad < \|w(k)\|^2,
\end{aligned}$$

for all $k \geqslant 0$. If $2\sigma/(\sigma+\lambda_p) < \|w(k)\|^2 < 2$, then

$$\begin{aligned}
&1+\eta(\lambda_i(2-\|w(k)\|^2)-w^{\mathrm{T}}(k)Cw(k)) \\
&\quad \leqslant 1+\eta(2\sigma-(\sigma+\lambda_p)\|w(k)\|^2) \\
&\quad < 1,
\end{aligned}$$

for all $k \geqslant 0$. So, it follows

$$\begin{aligned}
&\|w(k+1)\|^2 \\
&\quad = \sum_{i=1}^{n} \left[1+\eta\left(\lambda_i(2-\|w(k)\|^2)-w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 z_i^2(k) \\
&\quad \leqslant [1+\eta(2\sigma-(\sigma+\lambda_p)\|w(k)\|^2)]^2\|w(k)\|^2 \\
&\quad < \|w(k)\|^2,
\end{aligned}$$

for all $k \geqslant 0$. If $0 < \|w(k)\|^2 \leqslant 2\sigma/(\sigma+\lambda_p)$, then

$$\begin{aligned}
&\|w(k+1)\|^2 \\
&\quad = \sum_{i=1}^{n} \left[1+\eta\left(\lambda_i(2-\|w(k)\|^2)-w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 z_i^2(k) \\
&\quad \leqslant \left[1+\eta\left(2\sigma-(\sigma+\lambda_p)\|w(k)\|^2\right)\right]^2\|w(k)\|^2 \\
&\quad \leqslant \left[1+\frac{0.25}{\sigma}(2\sigma-(\sigma+\lambda_p)\|w(k)\|^2)\right]^2\|w(k)\|^2 \\
&\quad \leqslant \left[1.5-0.25\frac{\sigma+\lambda_p}{\sigma}\|w(k)\|^2\right]^2\|w(k)\|^2 \\
&\quad \leqslant \frac{2\sigma}{\sigma+\lambda_p}.
\end{aligned}$$

Therefore, if $w(0) \in S$ and $w(0) \notin V_\sigma^{\perp}$, then there must exist a constant N so that

$$\|w(k)\|^2 \leqslant \frac{2\sigma}{\sigma+\lambda_p} < 2,$$

for all $k > N$. The proof is completed. □

The theorems above show the non-divergence of (3) is guaranteed if the learning rate satisfies a simple condition and the evolution of (3) is ultimately bounded.

4. Convergence analysis

In this section, we will prove the trajectories, arising from points in the invariant set S, will converge to a unit eigenvector associated with the largest eigenvalue of the correlation matrix. To complete the proof, the following lemmas are first given.

Lemma 3. *Suppose that $\eta\sigma \leqslant 0.25$. If $w(0) \in S$ and $w(0) \neq 0$, then*

$$\eta w^{\mathrm{T}}(k+1)Cw(k+1) \geqslant \gamma > 0,$$

for all $k \geqslant 0$ where

$$\gamma = \min\left\{(2\eta\lambda_p)^2(\eta\lambda_p)\frac{\lambda_p}{\sigma}, \eta w^{\mathrm{T}}(0)Cw(0)\right\} < 1.$$

See the Appendix for the proof.

Lemma 4. *If $w(0) \in S$ and $w(0) \notin V_\sigma^{\perp}$, then there exist constants $\theta_1 > 0$, $\Pi_1 \geqslant 0$ and $d > 0$ such that*

$$\sum_{j=l+1}^{n} z_j^2(k) \leqslant \Pi_1 \cdot \mathrm{e}^{-\theta_1 k},$$

for all $k > N$, where

$$\theta_1 = \ln\left(\frac{\sigma+\lambda_p+2\eta\sigma\lambda_p}{\sigma+\lambda_p+2\eta\lambda_{m+1}\lambda_p}\right)^2 > 0.$$

Proof. Since $w(0) \notin V_\sigma^{\perp}$, there must exist some i $(1 \leqslant i \leqslant m)$ such that $z_i(0) \neq 0$. Without loss of generality, assume that $z_1(0) \neq 0$.

By Theorem 1, it follows that $w(k) \in S$ for all $k \geqslant 0$. From Lemma 2, it holds that

$$1+\eta(\lambda_i(2-\|w(k)\|^2)-w^{\mathrm{T}}(k)Cw(k)) > 0,$$

for $k \geqslant 0$. From (6), for each j $(m+1 \leqslant j \leqslant n)$, it follows that

$$\begin{aligned}
\left[\frac{z_j(k+1)}{z_1(k+1)}\right]^2 &= \left[\frac{1+\eta[\lambda_i(2-\|w(k)\|^2)-w^{\mathrm{T}}(k)Cw(k)]}{1+\eta[\sigma(2-\|w(k)\|^2)-w^{\mathrm{T}}(k)Cw(k)]}\right]^2 \\
&\quad \cdot \left[\frac{z_j(k)}{z_1(k)}\right]^2.
\end{aligned}$$

From Theorem 2, then

$$\|w(k+1)\|^2 \leqslant \frac{2\sigma}{\sigma+\lambda_p} < 2,$$

for all $k > N$. So, it follows that

$$\begin{aligned}
\left[\frac{z_j(k+1)}{z_1(k+1)}\right]^2 &\leqslant \left[\frac{1+\eta[\lambda_i(2-\|w(k)\|^2)]}{1+\eta[\sigma(2-\|w(k)\|^2)]}\right]^2 \cdot \left[\frac{z_j(k)}{z_1(k)}\right]^2 \\
&\leqslant \left[\frac{1+\eta\left[\lambda_i\left(2-\frac{2\sigma}{\sigma+\lambda_p}\right)\right]}{1+\eta\left[\sigma\left(2-\frac{2\sigma}{\sigma+\lambda_p}\right)\right]}\right]^2 \cdot \left[\frac{z_j(k)}{z_1(k)}\right]^2 \\
&= \left[\frac{\sigma+\lambda_p+2\eta\lambda_{m+1}\lambda_p}{\sigma+\lambda_p+2\eta\sigma\lambda_p}\right]^2 \cdot \left[\frac{z_j(k)}{z_1(k)}\right]^2 \\
&= \left(\frac{\sigma+\lambda_p+2\eta\lambda_{m+1}\lambda_p}{\sigma+\lambda_p+2\eta\sigma\lambda_p}\right)^{2(k+1)} \cdot \left[\frac{z_j(0)}{z_1(0)}\right]^2 \\
&= \left[\frac{z_j(0)}{z_1(0)}\right]^2 \cdot \mathrm{e}^{-\theta_1(k+1)},
\end{aligned}$$

for all $k > N$.

Since $w(k) \in S$, $z_1(k)$ must be bounded, i.e., there exists a constant $d > 0$ such that $z_1^2(k) \leqslant d$. Then,

$$\sum_{j=m+1}^{n} z_j^2(k) = \sum_{j=m+1}^{n}\left[\frac{z_j(k)}{z_1(k)}\right]^2 \cdot z_1^2(k) \leqslant \Pi_1\,\mathrm{e}^{-\theta_1 k},$$

for all $k > N$, where

$$\Pi_1 = d\sum_{j=m+1}^{n}\left[\frac{z_j(0)}{z_1(0)}\right]^2 \geqslant 0.$$

This completes the proof. □

Lemma 5. *If $w(0) \in S$ and $w(0) \notin V_\sigma^\perp$, then there exists constants $\theta_2 > 0$ and $\Pi_2 > 0$ such that*

$$|\sigma(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k+1)Cw(k+1)| \leqslant k \cdot \Pi_2 \cdot [\mathrm{e}^{-\theta_2(k+1)} + \max\{\mathrm{e}^{-\theta_2 k}, \mathrm{e}^{-\theta_1 k}\}],$$

for all $k > N$, where

$$\begin{cases} \theta_2 = -\ln\delta, \quad (0 < \delta < 1), \\ \delta = \max\{(1-\gamma)^2, 2\eta\sigma\}, \\ \gamma = \min\left\{(2\eta\lambda_p)^2(\eta\lambda_p)\left(\dfrac{\lambda_p}{\sigma}\right), \eta w^{\mathrm{T}}(0)Cw(0)\right\}, \end{cases}$$

and $0 < \delta < 1$, $0 < \gamma < 1$.

See the Appendix for the proof.

Lemma 6. *Suppose there exists constants $\theta > 0$ and $\Pi > 0$ such that*

$$\eta|(\sigma(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k))z_i(k)| \leqslant k \cdot \Pi\,\mathrm{e}^{-\theta k} \quad (i = 1, \ldots, m)$$

for $k \geqslant N$. Then,

$$\lim_{k\to+\infty} z_i(k) = z_i^* \quad (i = 1, \ldots, m),$$

where z_i^* $(i = 1, \ldots, m)$ are constants.

Proof. Given any $\varepsilon > 0$, there exists a $K \geqslant 1$ such that

$$\frac{\Pi K\,\mathrm{e}^{-\theta K}}{(1 - \mathrm{e}^{-\theta})^2} \leqslant \varepsilon.$$

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

$$\begin{aligned} |z_i(k_1) - z_i(k_2)| &= \left|\sum_{r=k_2}^{k_1-1}[z_i(r+1) - z_i(r)]\right| \\ &\leqslant \eta\sum_{r=k_2}^{k_1-1}|(\sigma(2 - \|w(k)\|^2) - w^{\mathrm{T}}(r)Cw(r))z_i(r)| \\ &\leqslant \Pi\sum_{r=k_2}^{k_1-1} r\,\mathrm{e}^{-\theta r} \\ &\leqslant \Pi\sum_{r=K}^{+\infty} r\,\mathrm{e}^{-\theta r} \\ &\leqslant \Pi K\,\mathrm{e}^{-\theta K}\cdot\sum_{r=0}^{+\infty} r(\mathrm{e}^{-\theta})^{r-1} \\ &= \frac{\Pi K\,\mathrm{e}^{-\theta K}}{(1-\mathrm{e}^{-\theta})^2} \\ &\leqslant \varepsilon \quad (i = 1, \ldots, m). \end{aligned}$$

This shows that each sequence $\{z_i(k)\}$ is a *Cauchy sequence*. By *Cauchy Convergence Principle*, there must exist constants z_i^* $(i = 1, \ldots, m)$ such that

$$\lim_{k\to+\infty} z_i(k) = z_i^* \quad (i = 1, \ldots, m).$$

This completes the proof. $\square$

Theorem 3. *Suppose that*

$\eta\sigma \leqslant 0.25.$

If $w(0) \in S$ and $w(0) \notin V_\sigma^\perp$, then the trajectory of (3) starting from $w(0)$ will converge to a unit eigenvector associated with the largest eigenvalue of the correlation matrix C.

Proof. By Lemma 4, there exists constants $\theta_1 > 0$ and $\Pi_1 \geqslant 0$ such that

$$\sum_{j=m+1}^{n} z_j^2(k) \leqslant \Pi_1\,\mathrm{e}^{-\theta_1 k},$$

for all $k \geqslant N$. By Lemma 5, there exists constants $\theta_2 > 0$ and $\Pi_2 > 0$ such that

$$|\sigma(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k+1)Cw(k+1)| \leqslant k \cdot \Pi_2 \cdot [\mathrm{e}^{-\theta_2(k+1)} + \max\{\mathrm{e}^{-\theta_2 k}, \mathrm{e}^{-\theta_1 k}\}]$$

for all $k \geqslant N$.

Obviously, there exists constants $\theta > 0$ and $\Pi > 0$ such that

$$\eta|(\sigma(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k))z_i(k)| \leqslant k \cdot \Pi\,\mathrm{e}^{-\theta k} \quad (i = 1, \ldots, m)$$

for $k \geqslant N$.

Using Lemmas 4 and 6, it follows that

$$\begin{cases} \lim_{t\to+\infty} z_i(k) = z_i^* \quad (i = 1, \ldots, m), \\ \lim_{t\to+\infty} z_i(k) = 0 \quad (i = m+1, \ldots, n). \end{cases}$$

Thus,

$$\lim_{t\to+\infty} w(k) = \sum_{i=1}^{m} z_i^* v_i \in V_\sigma. \tag{7}$$

From (3), after the system becomes stable, it follows that

$$\lim_{k\to\infty} 2Cw(k) = \lim_{k\to\infty}[Cw(k)w^{\mathrm{T}}(k)w(k) - (w^{\mathrm{T}}(k)Cw(k))w(k)]. \tag{8}$$

Substitute (7) into (8), we get

$$\sum_{i=1}^{m}\sigma z_i^* v_i = \sum_{i=1}^{m}\sigma z_i^* v_i\sum_{i=1}^{m}(z_i^*)^2 + \sum_{i=1}^{m}\sigma(z_i^*)^2\sum_{i=1}^{m} z_i^* v_i.$$

It is easy to see that

$$\sum_{i=1}^{m}(z_i^*)^2 = 1.$$

The proof is completed. $\square$

The above theorem requires that the initial $w(0) \notin V_\sigma^\perp$ so that there is at least one $z_i(0) \neq 0$ $(1 \leqslant i \leqslant m)$. Since any small disturbance can result in $w(0) \notin V_\sigma^\perp$, the condition $w(0) \notin V_\sigma^\perp$ is easy to meet in practical applications. Therefore, as long as the learning rate satisfies a simple condition, almost all trajectories starting from S will converge to a unit eigenvector associated with the largest eigenvalue of the correlation matrix of C.

5. Simulations and discussions

5.1. Convergence of Xu's algorithm

Simulations in this section show the convergence of Xu's algorithm with a constant learning rate. First, we randomly generate a 6×6 symmetric nonnegative definite matrix as

$$C = \begin{bmatrix} 0.1712 & 0.1538 & 0.097645 & 0.036741 & 0.07963 & 0.12897 \\ 0.1538 & 0.13855 & 0.087349 & 0.033022 & 0.072609 & 0.11643 \\ 0.097645 & 0.087349 & 0.067461 & 0.032506 & 0.043641 & 0.070849 \\ 0.036741 & 0.033022 & 0.032506 & 0.019761 & 0.016771 & 0.025661 \\ 0.07963 & 0.072609 & 0.043641 & 0.016771 & 0.041089 & 0.062108 \\ 0.12897 & 0.11643 & 0.070849 & 0.025661 & 0.062108 & 0.098575 \end{bmatrix}.$$

The maximum eigenvalue $\sigma = 0.5099$. The invariant set S can be identified as

$$S = \left\{ w(k) | w(k) \in R^2, \|w(k)\|^2 \leqslant \frac{1}{2 \times \eta \times 0.5099} \right\},$$

where $\eta \leqslant 0.25/0.5099 \approx 0.4903$.

The system (3) is not globally convergent. If the initial vector is

$$w(0) = [3.3500, 4.3000, 1.5000, 2.4000, 3.4000, 1.7000]^{\mathrm{T}}$$

and $\eta = 0.05$, $\|w(k)\|$ goes to infinity rapidly. Fig. 1 shows the result. Theorem 3 shows that trajectories starting from $S - \{0\}$ will converge to an eigenvector associated with the largest eigenvalue of the correlation matrix. The following

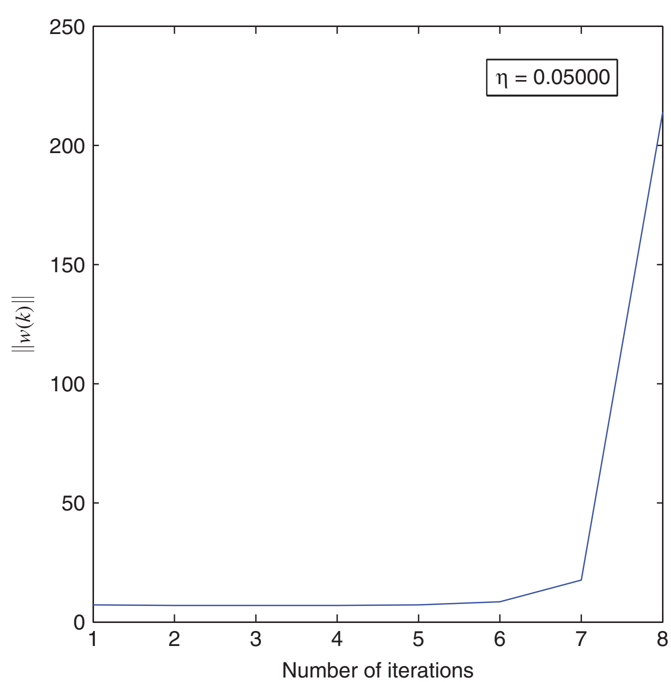


Fig. 1. Divergence of (3).

simulations will confirm it. Selecting 6 initial vectors arbitrarily in $S - \{0\}$ as

The $Norm^2$ of initial vector					
2.5214	3.1887	1.3608	0.0044	5.3374	0.1287

Fig. 2 (left) shows the six trajectories converging to the unit hyper spherical plane. The evolution result with the same initial vector and four different learning rates is presented in Fig. 2 (right).

In order to measure the convergent direction, we compute the norm of $w(k)$ and the direction cosine at kth update by [4,5,23]:

$$\text{Direction cosine}(k) = \frac{|w^{\mathrm{T}}(k) \cdot \phi|}{\|w(k)\| \cdot \|\phi\|},$$

where ϕ is the true eigenvector associated with the largest eigenvalue of C. Then, we select arbitrarily an initial vector as

$$w(0) = [0.0900, 0.0400, 0.0950, 0.0900, 0.0840, 0.0600]^{\mathrm{T}}.$$

Fig. 3 (left) shows the convergence of the direction cosine and the convergence of $\|w(k)\|$. The right one shows the components of w converge to

$$w^* = [0.58584, 0.56139, 0.39833, 0.14884, 0.24411, 0.31816]^{\mathrm{T}}.$$

This algorithm is more suitable for online learning with high performance. The online evolution behavior is illustrated by extracting a principal feature from Lenna picture in Fig. 4 (left). By numerical discretization procedure, the Xu's algorithm (3) with a constant learning rate could be rewritten as follows:

$$w(k+1) = w(k) + \eta(2C_k w(k) - w(k)^{\mathrm{T}} w(k) C_k w(k) - (w(k)^{\mathrm{T}} C_k w(k)) w(k)),$$

for $k \geqslant 0$, where C_x is a online observation sequence. The 512×512 pixel gray picture for Lenna in Fig. 4 (left) is split into 4096 vectors with 64 dimensions. The online observation sequence $\{C_k = ((k-1)C_{k-1} + x(k)x^{\mathrm{T}}(k))/k\}$ will be used to train the system with $\lim_{k\to\infty} C_k = C, (k > 0)$ [4,5]. The reconstructed image is presented in Fig. 4 (right) with SNR 32.175.

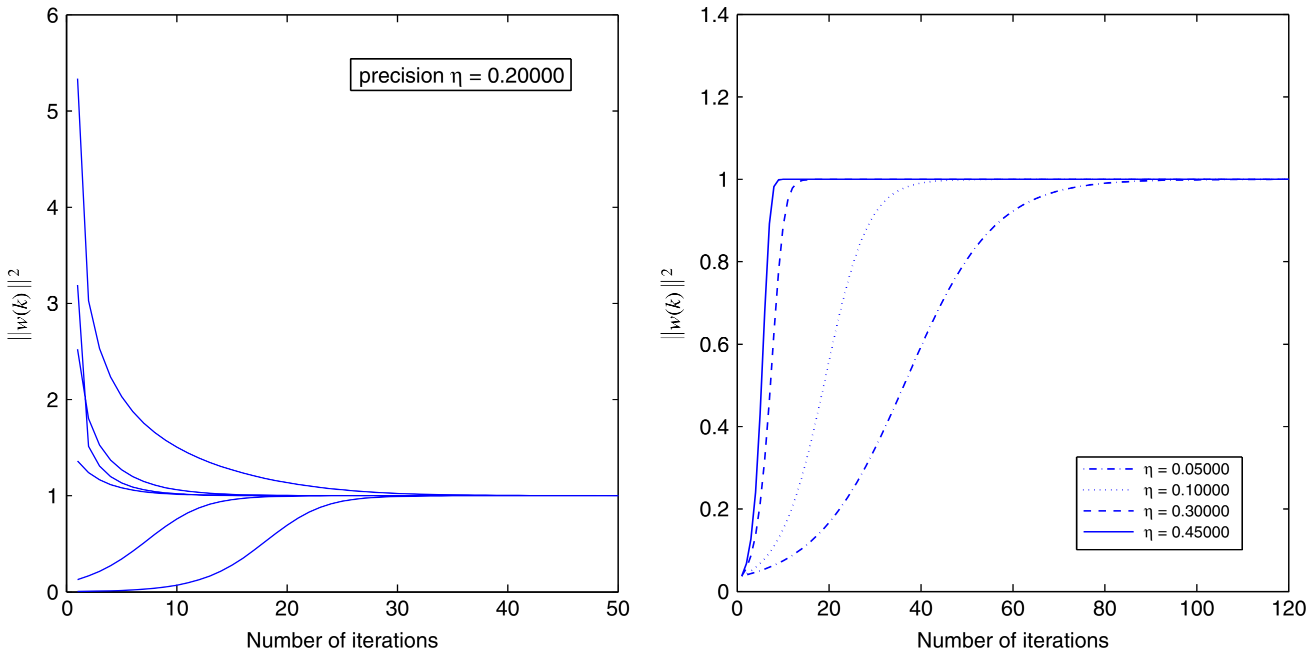


Fig. 2. Convergence of (3) with different initial vectors (left) and with different learning rates (right).

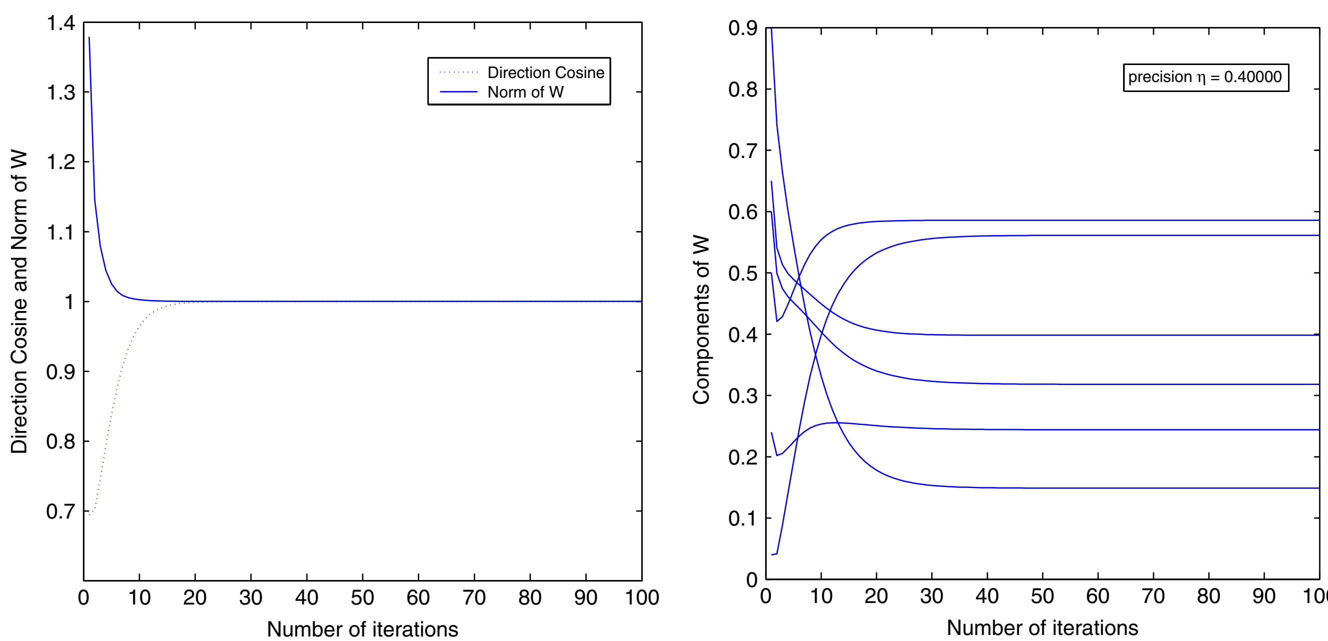


Fig. 3. Convergence of direction cosine (left) and components of w (right).

5.2. Comparison of Oja's algorithm and Xu's algorithm

In [29,23], the Oja's DDT system with a constant learning rate is analyzed in detail. Especially, in [23], Zhang presented some important results. The Oja's DDT system can be presented as follows:

$$w(k+1) = w(k) + \eta[Cw(k) - (w^{\mathrm{T}}(k)Cw(k))w(k)],$$

for $k \geqslant 0$. Clearly, Xu's algorithm (3) approximates Oja's one unit algorithm when the normal of the weight vector approaches 1. However, some different dynamical behaviors can be shown.

On the one hand, Xu's algorithm has a different attractive domain. Given a same input sequence, Oja's algorithm requires $\eta\sigma < 0.8899$ and its attractive domain [23] is

$$S_{\mathrm{oja}} = \left\{ w(k) | w(k) \in R^n, w^{\mathrm{T}}(k)Cw(k) < \frac{1}{\eta} \right\}.$$

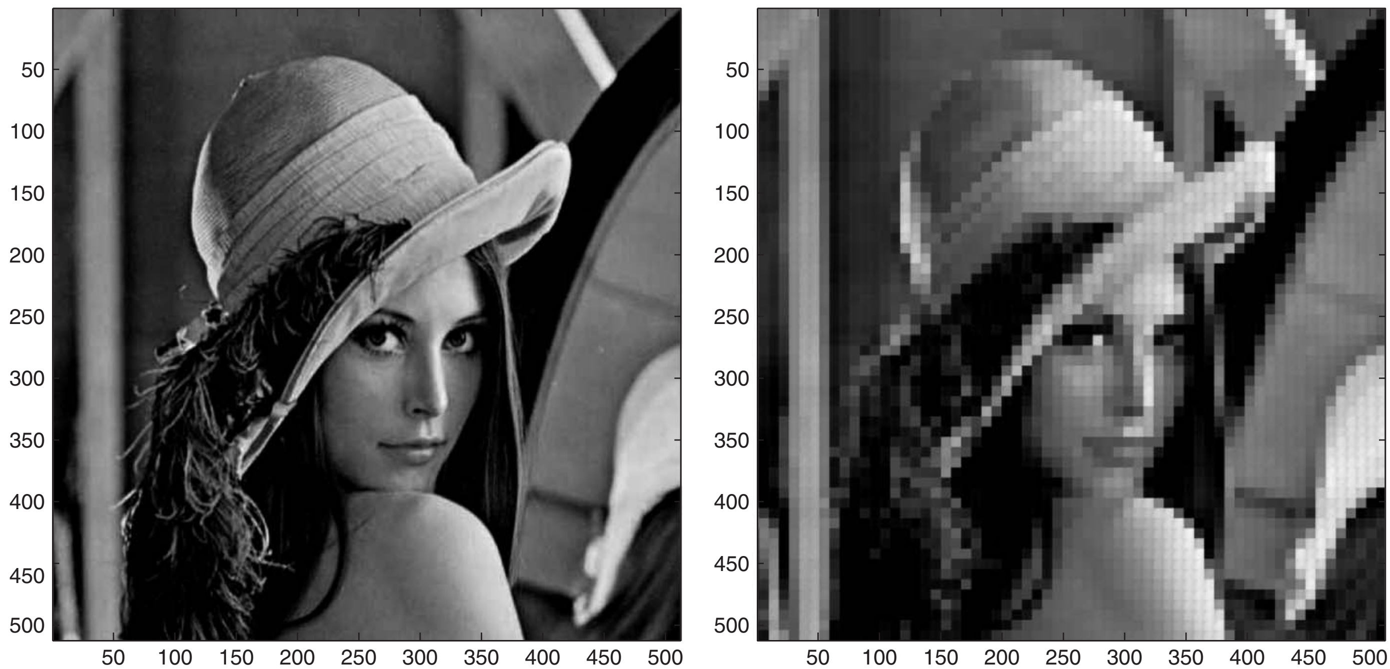


Fig. 4. The original image of Lenna (left) and the reconstructed image (right).

Xu's algorithm requires $\eta\sigma < 0.25$ with the attractive domain

$$S_{\text{xu}} = \left\{ w(k) | w(k) \in R^n, \|w(k)\|^2 < 1 + \frac{1}{2\eta\sigma} \right\}.$$

On the other hand, we find Xu's algorithm converges faster on the whole, though Xu's algorithm (3) has a larger computation in each iteration. In [23], the coefficients $z_i(k)$ of weight about Oja's algorithm is presented as

$$z_i(k+1)_{\text{oja}} = [1 + \eta(\lambda_i - w^{\mathrm{T}}(k)Cw(k))]z_i(k),$$

for $k \geqslant 0$, where $i = 1, \ldots, n$. Suppose $w(k)$ is same, with the same learning rate. From (6), we have

$$\begin{aligned} &\left| \frac{z_i(k+1)_{\text{xu}} - z_i(k+1)}{z_i(k+1)_{\text{oja}} - z_i(k+1)} \right| \\ &\quad = \left| \frac{\left[\eta\left(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right] z_i(k)}{[\eta(\lambda_i - w^{\mathrm{T}}(k)Cw(k))]z_i(k)} \right| \\ &\quad = \left| 1 + \frac{\lambda_i(1 - \|w(k)\|^2)}{\lambda_i - w^{\mathrm{T}}(k)Cw(k)} \right|. \end{aligned}$$

Clearly, it holds that

$$\left| \frac{z_i(k+1)_{\text{xu}} - z_i(k+1)}{z_i(k+1)_{\text{oja}} - z_i(k+1)} \right| = \left| 1 + \frac{\lambda_i(1 - \|w(k)\|^2)}{\lambda_i - w^{\mathrm{T}}(k)Cw(k)} \right| > 1,$$

for $w^{\mathrm{T}}(k)Cw(k) < \lambda_p$ or $w^{\mathrm{T}}(k)Cw(k) > \sigma$. So, in this range, Xu's algorithm converges faster. However, if $\lambda_p < w^{\mathrm{T}}(k)Cw(k) < \sigma$, it can be observed that the evolution rate of both algorithms interlace and trend to the same rate. Fig. 5 (right) shows the result. Furthermore, it is not difficult to get

$$\begin{cases} \sigma > w^{\mathrm{T}}(k)Cw(k) & \text{for } \|w(k)\|^2 < 1, \\ \lambda_p < w^{\mathrm{T}}(k)Cw(k) & \text{for } \|w(k)\|^2 > 1. \end{cases}$$

Thus, it holds that

$$\begin{cases} \left| \dfrac{z_1(k+1)_{\text{xu}} - z_1(k+1)}{z_1(k+1)_{\text{oja}} - z_1(k+1)} \right| & \text{for } \|w(k)\|^2 < 1, \\ \quad = \left| 1 + \dfrac{\sigma(1 - \|w(k)\|^2)}{\sigma - w^{\mathrm{T}}(k)Cw(k)} \right| > 1 & \\ \left| \dfrac{z_p(k+1)_{\text{xu}} - z_p(k+1)}{z_p(k+1)_{\text{oja}} - z_p(k+1)} \right| & \text{for } \|w(k)\|^2 > 1. \\ \quad = \left| 1 + \dfrac{\lambda_p(1 - \|w(k)\|^2)}{\lambda_p - w^{\mathrm{T}}(k)Cw(k)} \right| > 1 & \end{cases}$$

Therefore, in at least one direction, Xu's algorithm converges faster if $\lambda_p < w^{\mathrm{T}}(k)Cw(k) < \sigma$ and by a number of simulations, it can be observed that, on the whole, Xu's algorithm converges faster. Consider a simple example, a input matrix is randomly generated as follows. Fig. 5 shows the evolution result of both algorithms.

$$C = \begin{bmatrix} 0.9267 & 0.0633 & -0.2633 \\ 0.0633 & 0.1900 & 0.1267 \\ -0.2633 & 0.1267 & 0.9000 \end{bmatrix}.$$

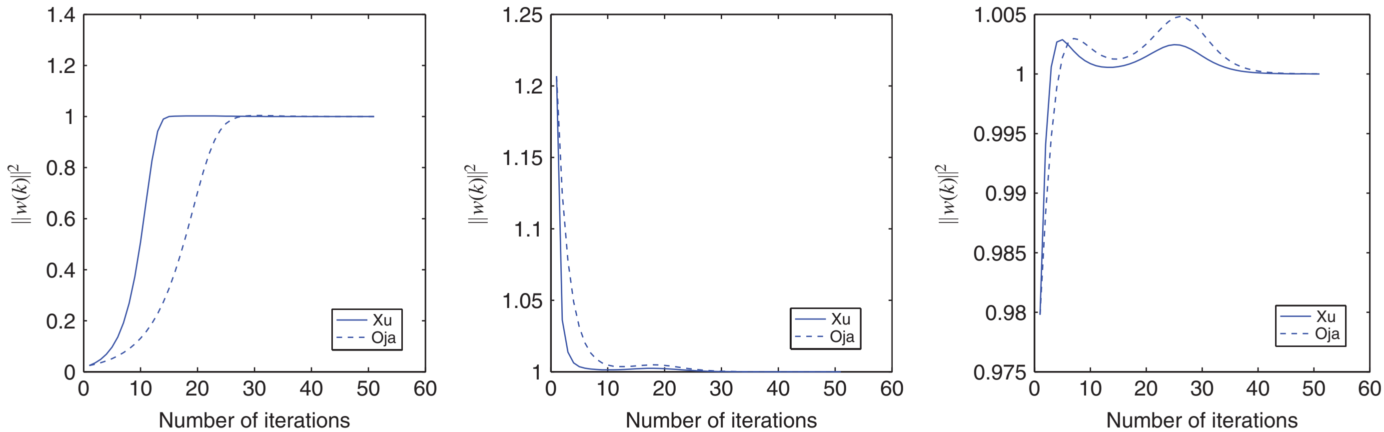


Fig. 5. Comparison of evolution rate of Xu's and Oja's algorithm with a same learning rate $\eta = 0.3$.

6. Conclusions

The convergence of Xu's LMSER algorithm, in the one unit case(2), is interpreted by using the DDT method. An invariant sets and an ultimate bound are derived to guarantee the non-divergence of Xu's DDT system. It is rigorously proven, if the learning rate satisfy some simple condition, all trajectories, starting from points in this invariant set, converge to the eigen subspace, which is spanned by the eigenvectors associated with the largest eigenvalue of the correlation matrix. At the same time, it could be concluded that Xu's algorithm converges faster on the whole. Extensive simulations have further illustrated the theory.

Appendix

Proof of Lemma 1. Define a differentiable function

$$f(s) = [1 + \eta(\sigma(2 - s))]^2 s$$

for $0 \leqslant s \leqslant 2 + 1/\eta\sigma$. It follows that

$$\dot{f}(s) = [1 + 2\eta\sigma - \eta\sigma s] \cdot [1 + 2\eta\sigma - 3\eta\sigma],$$

for $0 \leqslant s \leqslant 2 + 1/\eta\sigma$. Denote

$$\xi = \frac{1 + 2\eta\sigma}{3\eta\sigma}.$$

Then,

$$\dot{f}(s) \begin{cases} > 0 & \text{if } 0 \leqslant s \leqslant \xi, \\ = 0 & \text{if } s = \xi, \\ < 0 & \text{if } \xi \leqslant s < 2 + 1/\eta\sigma. \end{cases}$$

This shows that ξ must be the maximum point of the function $f(s)$ on the interval $[0, 2 + 1/\eta\sigma]$. Then, it holds that,

$$f(s) \leqslant f(\xi) = \frac{4}{27\eta\sigma} \cdot (1 + 2\eta\sigma)^3,$$

for all $0 \leqslant s \leqslant 2 + 1/\eta\sigma$. The proof is completed. □

Proof of Lemma 2. Given any i $(1 \leqslant i \geqslant n)$, if $2 \leqslant \|w(k)\|^2 < 1 + 1/2\eta\sigma$, then,

$$\begin{aligned} & 1 + \eta(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)) \\ & \quad \geqslant 1 + \eta(\sigma(2 - \|w(k)\|^2) - \sigma\|w^{\mathrm{T}}\|^2) > 0, \end{aligned}$$

for all $k \geqslant 0$. If $\|w(k)\|^2 < 2$, then

$$\begin{aligned} & 1 + \eta(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)) \\ & \quad \geqslant 1 - \eta w^{\mathrm{T}}(k)Cw(k) > 1 - 2\eta\sigma > 0, \end{aligned}$$

for all $k \geqslant 0$. The proof is completed. □

Proof of Lemma 3. Since $w(0) \in S$, by (2),we have

$$1 + \eta(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)) > 0,$$

for all $k \geqslant 0$. From Theorem 2, if $2 \leqslant \|w(k)\|^2 \leqslant 1 + 1/2\eta\sigma$, we have

$$\|w(k+1)\|^2 < \|w(k)\|^2 \quad (k \geqslant 0).$$

Case 1: $\lambda_p/\sigma < 2$.
If $\lambda_p/\sigma \leqslant \|w(k)\|^2 < 2$, from (5) and (6), it follows that

$$\begin{aligned} & \eta w^{\mathrm{T}}(k+1)Cw(k+1) \\ & \quad = \eta \sum_{i=1}^{n} \lambda_i z_i^2(k+1) \\ & \quad = \eta \sum_{i=1}^{p} \left[1 + \eta\left(\lambda_i(2 - \|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 \lambda_i z_i^2(k) \\ & \quad \geqslant \eta \sum_{i=1}^{p} \left[1 + \eta\left(2\lambda_p - \sigma\|w(k)\|^2 - \sigma\|w(k)\|^2\right)\right]^2 \lambda_i z_i^2(k) \\ & \quad \geqslant \eta\left[2\eta\lambda_p + \left(1 - 2\eta\sigma\|w(k)\|^2\right)\right]^2 w^{\mathrm{T}}(k)Cw(k) \\ & \quad \geqslant (2\eta\lambda_p)^2 \cdot \eta w^{\mathrm{T}}(k)Cw(k) \\ & \quad \geqslant (2\eta\lambda_p)^2(\eta\lambda_p)\frac{\lambda_p}{\sigma} \end{aligned} \tag{9}$$

for $k \geqslant 0$.

If $\|w(k)\|^2 \leqslant \lambda_p/\sigma$, then,

$$
\begin{aligned}
&\eta w^{\mathrm{T}}(k+1)Cw(k+1) \\
&\quad = \eta \sum_{i=1}^{p} \left[1+\eta\left(\lambda_i - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 \lambda_i z_i^2(k) \\
&\quad \geqslant \eta\left[1+\eta\left(2\lambda_p - \sigma\|w(k)\|^2 - \sigma\|w(k)\|^2\right)\right]^2 w^{\mathrm{T}}(k)Cw(k) \\
&\quad \geqslant \eta w^{\mathrm{T}}(k)Cw(k) \qquad (10)
\end{aligned}
$$

for $k \geqslant 0$.

Case 2: $\lambda_p/\sigma > 2$.

Clearly, if $2 \leqslant \|w(k)\|^2 < \lambda_p/\sigma$, it holds that

$$\|w(k+1)\|^2 < \|w(k)\|^2 \quad (k \geqslant 0).$$

If $0 < \|w(k)\|^2 < 2$, From (10), it follows that

$$\eta w^{\mathrm{T}}(k+1)Cw(k+1) \geqslant \eta w^{\mathrm{T}}(k)Cw(k) \quad (k \geqslant 0). \qquad (11)$$

Let

$$\gamma = \min\left\{4(\eta\lambda_p)^3 \frac{\lambda_p}{\sigma}, \eta w^{\mathrm{T}}(0)Cw(0)\right\},$$

where $0 < \gamma < 1$. It follows from (9),(10) and (11) that

$$\eta w^{\mathrm{T}}(k+1)Cw(k+1) \geqslant \gamma > 0,$$

for all $k \geqslant 0$. This completes the proof. □

Proof of the Lemma 5. From (5), (6), we have

$$
\begin{aligned}
&w^{\mathrm{T}}(k+1)Cw(k+1) \\
&\quad = \sum_{i=1}^{n} \lambda_i\left[1+\eta\left(\lambda_i(2-\|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 z_i^2(k) \\
&\quad = \sum_{i=1}^{n} \lambda_i\left[1+\eta\left(\sigma(2-\|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 z_i^2(k) \\
&\qquad + \sum_{i=m+1}^{n} \lambda_i\left[1+\eta\left(\lambda_i(2-\|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 z_i^2(k) \\
&\qquad - \sum_{i=m+1}^{n} \lambda_i\left[1+\eta\left(\sigma(2-\|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 z_i^2(k) \\
&\quad = \left[1+\eta\left(\sigma(2-\|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)\right)\right]^2 \\
&\qquad \cdot w^{\mathrm{T}}(k)Cw(k) - Q(k),
\end{aligned}
$$

for $k \geqslant 0$, where

$$
\begin{aligned}
Q(k) = \eta &\sum_{i=m+1}^{n} \lambda_i(\sigma-\lambda_i)(2-\|w(k)\|^2) \\
&\times\left[2+\eta\left((\sigma+\lambda_i)(2-\|w(k)\|^2)\right.\right. \\
&\left.\left.-2w^{\mathrm{T}}(k)Cw(k)\right)\right] z_i^2(k).
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\sigma(2-\|w(k)\|^2) - w^{\mathrm{T}}(k+1)Cw(k+1) \\
&\quad = \left[\sigma\left(2-\|w(k)\|^2\right) - w^{\mathrm{T}}(k)Cw(k)\right]\left[\left(1-\eta w^{\mathrm{T}}(k)Cw(k)\right)^2\right. \\
&\qquad \left. -\eta^2\sigma\left(2-\|w(k)\|^2\right)w^{\mathrm{T}}(k)Cw(k)\right] + Q(k),
\end{aligned}
$$

for $k \geqslant 0$. Denote

$$V(k) = |\sigma(2-\|w(k)\|^2) - w^{\mathrm{T}}(k)Cw(k)|,$$

for $k \geqslant 0$. By the invariance of S, $w(k) \in S$ for $k \geqslant 0$. It follows that,

$$
\begin{aligned}
&V(k+1) \\
&\quad \leqslant V(k) \cdot |(1-\eta w^{\mathrm{T}}(k)Cw(k))^2 \\
&\qquad - \eta^2\sigma(2-\|w(k)\|^2)w^{\mathrm{T}}(k)Cw(k)| + |Q(k)| \\
&\quad \leqslant \max\left\{(1-\eta w^{\mathrm{T}}(k)Cw(k))^2, \eta^2\sigma(2-\|w(k)\|^2)w^{\mathrm{T}}(k)Cw(k)\right\} \\
&\qquad \cdot V(k) + |Q(k)| \\
&\quad \leqslant \max\left\{(1-\eta w^{\mathrm{T}}(k)Cw(k))^2, \eta\sigma(2-\|w(k)\|^2)\right\} \\
&\qquad \cdot V(k) + |Q(k)|.
\end{aligned}
$$

From Theorem 2, it follows that

$$
\begin{aligned}
V(k+1) \leqslant \max&\left\{(1-\eta w^{\mathrm{T}}(k)Cw(k))^2, 2\eta\sigma\right\} \\
&\cdot V(k) + |Q(k)|,
\end{aligned}
$$

for $k > N$. Denote

$$\gamma = \min\left\{(2\eta\lambda_p)^2(\eta\lambda_p)\frac{\lambda_p}{\sigma}, \eta w^{\mathrm{T}}(0)Cw(0)\right\}.$$

Clearly, $0 < \gamma < 1$. By Lemma 4, it holds that

$$\gamma \leqslant \eta w^{\mathrm{T}}(k)Cw(k) \leqslant 1.$$

Denote

$$\delta = \max\{(1-\gamma)^2, 2\eta\sigma\}.$$

Clearly, $0 < \delta < 1$. Then,

$$V(k+1) \leqslant \delta \cdot V(k) + |Q(k)|, \quad k \geqslant N.$$

By Lemma 3,

$$
\begin{aligned}
|Q(k)| &\leqslant \eta \sum_{j=m+1}^{n} 2\lambda_j(\sigma-\lambda_j)\left[2+2\eta\left(\sigma+\lambda_j\right)\right] z_j^2(k) \\
&\leqslant \eta \sum_{j=m+1}^{n} 4\left[\eta\lambda_j\left(\sigma^2-\lambda_j^2\right) + \lambda_j(\sigma-\lambda_j)\right] z_j^2(k) \\
&\leqslant \eta \sum_{j=m+1}^{n} 4(\eta\lambda_j\sigma^2 + 2\lambda_j\sigma) z_j^2(k) \\
&\leqslant 4\sigma(\eta\sigma)(\eta\sigma+2) \sum_{j=m+1}^{n} z_j^2(k) \\
&\leqslant 12\sigma \cdot \sum_{j=m+1}^{n} z_j^2(k) \\
&\leqslant 12\sigma\Pi_1 \cdot \mathrm{e}^{-\theta_1 k}
\end{aligned}
$$

for all $k \geqslant N$. Then,

$$
\begin{aligned}
V(k+1) &\leqslant \delta^{k+1}V(0) + 12\sigma\Pi_1 \cdot \sum_{r=0}^{k} \left(\delta\,\mathrm{e}^{\theta_1}\right)^r \mathrm{e}^{-\theta_1 k} \\
&\leqslant \delta^{k+1}V(0) + 12k\sigma\Pi_1 \cdot \max\left\{\delta^k, \mathrm{e}^{-\theta_1 k}\right\} \\
&\leqslant k \cdot \Pi_2 \cdot \left[\mathrm{e}^{-\theta_2(k+1)} + \max\left\{\mathrm{e}^{-\theta_2 k}, \mathrm{e}^{-\theta_1 k}\right\}\right]
\end{aligned}
$$

where $\theta_2 = -\ln\delta > 0$ and

$$\Pi_2 = \max\left\{|2\sigma - w^{\mathrm{T}}(0)Cw(0)|, 12\sigma\Pi_1\right\} > 0.$$

The proof is completed. □

References

[1] H.M. Abbas, M.M. Fahmy, Neural model for Karhunen–Loeve transform with application to adaptive image compression, IEE Proc.-I 140 (2) (1993) 135–143.

[2] T. Anisse, C. Gianalvo, Against the convergence of the minor component analysis neurons, IEEE Trans. Neural Network 10 (1) (1999) 207–210.

[3] S. Bannour, R. Azimi-Sadjadi, Principal component extraction using recursive least squares learning, IEEE Trans. Neural Networks 6 (2) (1995) 457–469.

[4] C. Chatterjee, Z. Kang, V.P. Roychowdhury, Algorithms for accelerated convergence of adaptive PCA, IEEE Trans. Neural Networks 11 (2) (2000) 338–355.

[5] C. Chatterjee, V.P. Roychowdhury, E.P. Chong, On relative convergence properties of principal component analysis algorithm, IEEE Trans. Neural Networks 9 (1998) 319–329.

[6] T. Chen, Y. Hua, W. Yan, Global convergence of Oja's subspace algorithm for principal component extraction, IEEE Trans. Neural Networks 9 (1) (1998) 58–67.

[7] H. Chen, R.W. Liu, An on-line unsupervised learning machine for adaptive feature extraction, IEEE Trans. on Circuits and Systems-II: Analog and Digital Signal Processing 41 (2) (1994) 87–98.

[8] A. Cichocki, W. Kasprzak, W. Skarbek, Adaptive learning algorithm for principal component analysis with partial data, Cybernetics Systems 2 (1996) 1014–1019.

[9] G. Cirrincione, M. Cirrincione, J. Hérault, S.V. Huffel, The MCA EXIN neuron for the minor component analysis, IEEE Trans. Neural Networks 13 (1) (2002) 160–186.

[10] K. Hornik, C.M. Kuan, Convergence analysis of local feature extraction algorithm, Neural Networks 5 (1992) 229–240.

[11] S. Kung, K. Diamantaras, J. Taur, Adaptive principal component extraction (APEX) and applications, IEEE Trans. Signal Process. 42 (5) (1994) 1202–1217.

[12] R. Möller, A. Könies, Coupled principal component analysis, IEEE Trans. Neural Networks 15 (1) (2004) 214–222.

[13] E. Oja, A simplified neuron mode as a principal component analyzer, J. Math. Biol. 15 (1982) 167–273.

[14] E. Oja, Principal components, minor components, and linear neural networks, Neural Networks 5 (1992) 927–935.

[15] E. Oja, J. Karhunen, On stochastic approximation of the eigenvector and eigenvalues of the expectation of a random matrix, J. Math. Anal. Approx. 106 (1) (1985) 69–84.

[16] S. Ouyang, Z. Bao, G.S. Liao, Robust recursive least squares learning algorithm for principal component analysis, IEEE Trans. Neural Networks 11 (1) (2000) 215–221.

[17] T.D. Sanger, Optimal unsupervised learning in a single-layer linear feedforward neural network, Neural Networks 2 (1989) 459–473.

[18] J.M. Vegas, P.J. Zufiria, Generalized neural networks for spectral analysis: dynamics and Liapunov functions, Neural Networks 17 (2004) 233–245.

[19] A. Weingessel, K. Hornik, Local PCA algorithms, IEEE Trans. Neural Networks 11 (5) (2000) 1242–1250.

[20] L. Xu, Least mean square error reconstruction principle for self-organizing neural-nets, Neural Networks 6 (1993) 627–648.

[21] W.Y. Yan, U. Helmke, J.B. Moore, Global analysis of Oja's flow for neural networks, IEEE Trans. Neural Networks 5 (1994) 674–683.

[22] Z. Yi, K.K. Tan, Convergence Analysis of Recurrent Neural Networks, Kluwer Academic Publishers, Dordrecht, 2004.

[23] Z. Yi, M. Ye, J.C. Lv, K.K. Tan, Convergence analysis of a deterministic discrete time system of Oja's PCA learning algorithm, IEEE Trans. Neural Networks 16 (6) (2005) 1318–1328.

[24] X. Yu, M.O. Efeand, O. Kaynak, A general backpropagation algorithm for feedforward neural networks learning, IEEE Trans. Neural Networks 13 (5) (2002) 251–254.

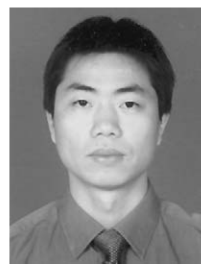
[25] W. Yu, A.S. Poznyak, X. Li, Multiplier dynamic neural networks for non-linear system on-line identification, Int. J. Control 74 (18) (2001) 1859–1864.

[26] Q. Zhang, On the discrete-time dynamics of a PCA learning algorithm, Neurocomputing 55 (2003) 761–769.

[27] Q. Zhang, Z. Bao, Dynamical system for computing the eigenvectors associated with the largest eigenvalue of a positive definite matrix, IEEE Trans. Neural Networks 6 (1995) 790–791.

[28] Q. Zhang, Y.W. Leund, Energy function for one-unit Oja algorithm, IEEE Trans. Neural Networks 6 (1995) 1291–1293.

[29] P.J. Zufiria, On the discrete-time dynamic of the basic Hebbian neural-network nods, IEEE Trans. Neural Networks 13 (6) (2002) 1252–1342.



Jian Cheng Lv received the Master degree in computer science and engineering from University of Electronic Science and Technology of China, Chengdu, China, in 2003. He is currently pursuing the Ph.D. degree in Computational Intelligence Laboratory, School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China. His research interests include neural networks and data mining.

Zhang Yi received the B.S. degree in mathematics from Sichuan Normal University, Chengdu, China, in 1983, the M.S. degree in mathematics from Hebei Normal University, Shijiazhuang, China, in 1986, and the Ph.D. degree in mathematics from the Institute of Mathematics, The Chinese Academy of Science, Beijing, China, in 1994. From 1989 to 1990, he was a Senior Visiting Scholar in the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, U.K. From February 1999 to August 2001, he was a Research Associate in the Department of Computer Science and Engineering at The Chinese University of Hong Kong. From August 2001 to December 2002, he was a Research Fellow in the Department of Electrical and Computer Engineering, National University of Singapore. He is currently a Professor at the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China. His current research interests include neural networks and data mining.

Kok Kiong Tan received the Ph.D. degree in 1995 from the National University of Singapore. Prior to joining the university, he was 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 and his current research interests are in precision motion control and instrumentation, advanced process control and autotuning, and general industrial automation.