International Journal of Computational Intelligence Research.
ISSN 0973-1873 Vol.2, No.2 (2006), pp. 207–215

A New Algorithm for Sequential Minor Component Analysis

Dezhong Peng and Zhang Yi

Computational Intelligence Laboratory
School of Computer Science and Engineering
University of Electronic Science and Technology of China
Chengdu 610054, P. R. China
email: {*pengdz, zhangyi*}*@uestc.edu.cn*

Abstract: **Extracting multiple minor components from the input signal is quite useful for many practical applications. In this paper, a globally convergent MCA algorithm that can extract multiple minor components sequentially is proposed. Convergence of this MCA algorithm is analyzed via the deterministic discrete time (DDT) method. Sufficient conditions are obtained to guarantee the convergence of this MCA algorithm. Simulations are carried out to further illustrate the theoretical results achieved.**
Keywords: Neural Networks, Minor Component Analysis, Deterministic Discrete Time system, Eigenvector, Eigenvalue.

I. Introduction

The minor component is the direction in which the data has the smallest covariance. The statistical method for extracting minor components from the input data is called minor component analysis (MCA). As an important tool for signal processing and data analysis, MCA has been applied to total least squares (TLS) [1] [2], moving target indication [3], clutter cancellation [4], computer vision [5], curve and surface fitting [6], digital beamforming [7], frequency estimation [8] [9], and bearing estimation [10], etc.

Extracting multiple minor components from the input data is quite useful to many practical applications. Some neural learning algorithms have been proposed to obtain multiple minor components sequentially. In [11], Luo *et al.* proposed an interesting MCA learning algorithm. By extending this algorithm, Luo *et al.* [12] gave a sequential MCA algorithm that can obtain multiple minor components. However, Chen *et al.* [13] pointed out that the sequential algorithm proposed by Luo [12] does not extract more than one minor component and presented a novel sequential MCA algorithm. Unfortunately, Chen's MCA algorithm may diverge to infinity since Chen's MCA algorithm is obtained by extending the MCA algorithm proposed by Luo [11] and the latter suffers from the divergence problem, as discussed in [14]. In [15], Feng proposed an effective MCA algorithm, called OJAm algorithm, that can extract the subspace that is spanned by minor components. However, OJAm algorithm can not obtain individual minor components. In this paper, we will extend OJAm algorithm to obtain a globally convergent MCA algorithm that can extract multiple minor components sequentially.

Convergence is crucial for MCA learning algorithm toward practical applications. However, most of neural MCA learning algorithms are described by stochastic discrete time (SDT) system and it is difficult to study the convergence of the SDT system directly. To indirectly analyze the convergence of MCA learning algorithms, a traditional method is to transform an MCA algorithm into a corresponding deterministic continuous time (DCT) system, the convergence of the MCA algorithm then can be interpreted by studying the convergence of the DCT system. The DCT method is based on a fundamental theorem of stochastic approximation theory [16]. To use this fundamental theorem of stochastic approximation, some crucial conditions must be satisfied. One important condition is that the learning rate of MCA algorithms must approach zero. However, this restrictive condition cannot be satisfied in many practical applications due to the round-off limitation and tracking requirements. Thus, from application points of view, the DCT method is not reasonable for studying the convergence of MCA algorithms.

Recently, deterministic discrete time (DDT) method has been used to study Oja's stochastic PCA learning algorithm [17, 18]. This DDT method transforms Oja's stochastic PCA learning algorithm into a deterministic discrete time system. It does not require the learning rate to approach zero. DDT systems preserve the discrete time nature of original SDT systems and can shed some light on the convergence characteristics of SDT systems. In this paper, we will analyze the convergence of the proposed sequential MCA algorithm via

DDT method.

This paper is organized as follows. We discuss some existing sequential MCA algorithms in Section 2. In Section 3, a new sequential MCA algorithm is proposed by extending OJAm MCA algorithm. The convergence analysis is given in Section 4 and Section 5. Simulations are given in Section 6. Finally, the conclusion follows in Section 7.

II. Some Discussions on Existing Sequential MCA Algorithms

Consider a single linear neuron with the following input output relation:

$$y(k) = w^T(k)x(k), \quad (k = 0, 1, 2, \ldots),$$

where $y(k)$ is the neuron output, the input sequence $\{x(k)|x(k) \in R^n(k = 0, 1, 2, \ldots)\}$ is a zero mean stationary stochastic process and $w(k) \in R^n(k = 0, 1, 2, \ldots)$ is the weight vector of the neuron. The target of MCA is to extract the minor component from the input data by updating the weight vector $w(k)$ adaptively. Let $R = E[x(k)x^T(k)]$ be the autocorrelation matrix of input signal $x(k)$. Since the autocorrelation matrix R is a symmetric nonnegative definite matrix, R has the ordered eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_n \geq 0$ and the corresponding unit eigenvectors $v_1, v_2, \ldots, v_n$. Let

$$R = \sum_{i=1}^{n} \lambda_i v_i v_i^T \tag{1}$$

be the eigenvalue decomposition of R. In [11], Luo *et al.* proposed an interesting MCA algorithm:

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

where $\eta > 0$ is the learning rate, and proved the convergence result of (2):

$$\lim_{k\to\infty} w(k) = \pm\|w(0)\|v_n, \tag{3}$$

if $w^T(0)v_n \neq 0$. By extending the algorithm (2), Chen *et al.* [13] proposed a sequential MCA algorithm to extract multiple minor components. This algorithm is as follows [13]:

(1) Extract the first minor component v_n by

$$\begin{aligned} w_1(k+1) &= w_1(k) - \eta[R_1 w_1(k)w_1^T(k)w_1(k) \\ &\quad -w_1(k)w_1^T(k)R_1 w_1(k)], \end{aligned} \tag{4}$$

where $R_1 = R = E[x(k)x^T(k)]$ and $\eta > 0$ is the learning rate.

(2) For a constant $\gamma > \lambda_1$, set

$$R_2(k) = R_1 + \gamma w_1(k)w_1^T(k)R_1. \tag{5}$$

(3) Extract the second minor component v_{n-1} by

$$\begin{aligned} w_2(k+1) &= w_2(k) - \eta[R_2(k)w_2(k)w_2^T(k)w_2(k) \\ &\quad -w_2(k)w_2^T(k)R_2(k)w_2(k)]. \end{aligned} \tag{6}$$

(4) Repeat the above procedure to extract further minor components. Denote the number of the extracted minor components by $p(n \geq p \geq 1)$. This algorithm is summarized for $j = 1, 2, \ldots, p$ as follows.

$$R_j(k) = R_1 + \gamma \sum_{i=1}^{j-1} w_i(k)w_i^T(k)R_1, \tag{7}$$

and

$$\begin{aligned} w_j(k+1) &= w_j(k) - \eta[R_j(k)w_j(k)w_j^T(k)w_j(k) \\ &\quad -w_j(k)w_j^T(k)R_j(k)w_j(k)]. \end{aligned} \tag{8}$$

The idea of Chen's algorithm is to by (5), make v_n become the largest principal component of $R_2(k)$ and v_{n-1} become the smallest minor component of $R_2(k)$. Thus, $w_2(k)$ will converge to v_{n-1}, as $k \to \infty$, in (6). By repeating the above procedure, the third and further minor components can be extracted sequentially.

Unfortunately, Chen's algorithm does not extract multiple minor components under some condition. Next, we will show the reason. Suppose that $\|w_1(0)\| = 1$. From (3), it holds that in (4)

$$\lim_{k\to\infty} w_1(k) = \pm v_n. \tag{9}$$

From (1), (5) and (9), it follows that

$$\lim_{k\to\infty} R_2(k) = \sum_{i=1}^{n-1} \lambda_i v_i v_i^T + \lambda_n(1+\gamma)v_n v_n^T. \tag{10}$$

Suppose that the eigenvalues of R have the following relation:

$$\lambda_n > 0 \quad \text{and} \quad \lambda_1 < \frac{\lambda_{n-1}}{\lambda_n} - 1,$$

and γ satisfies the following condition:

$$\lambda_1 < \gamma < \frac{\lambda_{n-1}}{\lambda_n} - 1, \tag{11}$$

i.e.

$$\lambda_n(1+\gamma) < \lambda_{n-1}. \tag{12}$$

From (10) and (12), clearly, v_n is still the smallest minor component of the matrix $R_2(k)$, as $k \to \infty$. Therefore, (6) can not extract the second minor component v_{n-1} under the condition (11).

Next, we will use a simulation result to illustrate the problem of Chen's algorithm. Let the autocorrelation matrix R be

$$R = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}.$$

Let us extract the first and second minor component of the matrix R by (4) and (6) with $\gamma = 5.5 > \lambda_1$, $\eta = 0.01$ and $w_1(0) = w_2(0) = [0.1 \quad 0.1 \quad 0.1]^T$. The simulation result shows that for large k,

$$w_1(k) = [0 \quad 0 \quad 0.1732]^T$$

and

$$w_2(k) = [0 \quad 0 \quad 0.1732]^T.$$

Clearly, both $w_1(k)$ and $w_2(k)$ will converge to the direction of the same minor component and the second minor component is not extracted.

III. The Proposed Algorithm

In [15], Feng *et al.* proposed an effective MCA learning algorithm that may extract the minor subspace spanned by minor components. This MCA learning algorithm, called OJAm algorithm, is as follows:

$$w(k+1) = w(k) - \eta \left\{ y(k)x(k) - \frac{y^2(k)}{\left[w^T(k)w(k)\right]^2} \cdot w(k) \right\}, \tag{13}$$

where $\eta > 0$ is the learning rate. By extending the algorithm (13), we can obtain a novel sequential MCA algorithm. This algorithm is presented as:

(1) Extract the first minor component v_n by (13), as follows

$$\begin{aligned} w_1(k+1) &= w_1(k) - \eta \Bigg\{ x(k)x^T(k)w_1(k) \\ &- \frac{w_1^T(k)x(k)x^T(k)w_1(k)}{\left[w_1^T(k)w_1(k)\right]^2} \cdot w_1(k) \Bigg\}. \end{aligned} \tag{14}$$

(2) For a constant $\gamma > \lambda_1$, set

$$C_2(k) = \gamma \frac{w_1(k)w_1^T(k)}{w_1^T(k)w_1(k)}. \tag{15}$$

(3) Extract the second minor component v_{n-1} by

$$\begin{aligned} & w_2(k+1) \\ = \; & w_2(k) - \eta \Bigg\{ \left[x(k)x^T(k) + C_2(k)\right] w_2(k) \\ - \; & \frac{w_2^T(k)\left[x(k)x^T(k) + C_2(k)\right] w_2(k)}{\left[w_2^T(k)w_2(k)\right]^2} \cdot w_2(k) \Bigg\}. \end{aligned} \tag{16}$$

(4) Repeat the above procedure to extract further minor components. Denote the number of the extracted minor components by $p(n \geq p \geq 1)$. The sequential MCA algorithm is summarized for $j = 1, 2, \ldots, p$ as follows.

$$\begin{aligned} & w_j(k+1) \\ = \; & w_j(k) - \eta \Bigg\{ \left[x(k)x^T(k) + C_j(k)\right] w_j(k) \\ - \; & \frac{w_j^T(k)\left[x(k)x^T(k) + C_j(k)\right] w_j(k)}{\left[w_j^T(k)w_j(k)\right]^2} \cdot w_j(k) \Bigg\}, \end{aligned} \tag{17}$$

where

$$C_1(k) = 0, C_j(k) = \gamma \sum_{i=1}^{j-1} \frac{w_i(k)w_i^T(k)}{w_i^T(k)w_i(k)} \tag{18}$$

for $p \geq j > 1$.

IV. Convergence Analysis of OJAm MCA Algorithm

Since the sequential MCA algorithm (14) - (18) is obtained by extending OJAm MCA algorithm, the convergence analysis of the original OJAm MCA algorithm is important. In [15], one crucial condition to guarantee the convergence of OJAm MCA algorithm is that the learning rate η is small enough. However, in many practical applications due to the round-off limitation and tracking requirements, the learning rate is usually a constant. Let us consider a one-dimensional example with constant input values $x(k) = 1$, the learning rate $\eta = 1$ and the initial weight vector $w(0) = 0.5$. From (13), we can get the following system:

$$w(k+1) = 1/w(k). \tag{19}$$

Clearly, $w(k)$ does not converge in (19). This example shows that OJAm MCA algorithm may diverge if the learning rate η is a constant. A problem to address is therefore to find out the conditions under which OJAm MCA algorithm can converge.

For convenience of analysis, next, some preliminaries are given. By taking conditional expectation operator $E\{w(k+1)/w(0), x(i), i < k\}$ to (13) and identifying the conditional expected value as the next iterate, a DDT system can be obtained as:

$$w(k+1) = w(k) - \eta \left[Rw(k) - \frac{w^T(k)Rw(k)}{\left[w^T(k)w(k)\right]^2} \cdot w(k) \right], \tag{20}$$

where $R = E[x(k)x^T(k)]$ is the autocorrelation matrix of $\{x(k)|x(k) \in R^n(k = 0, 1, 2, \ldots)\}$.

Since the autocorrelation matrix R is a symmetric nonnegative definite matrix, there exists an orthonormal basis of R^n composed of the unit eigenvectors $\{v_i|i = 1, 2, \ldots, n\}$ of R. Then, for each $k \geq 0$, the weight vector $w(k)$ can be represented as :

$$w(k) = \sum_{i=1}^{n} z_i(k)v_i, \tag{21}$$

where $z_i(k)(i = 1, 2, \ldots, n)$ are some constants, and then

$$Rw(k) = \sum_{i=1}^{n} \lambda_i z_i(k)v_i. \tag{22}$$

Clearly, it holds from (20) that

$$z_i(k+1) = \left[1 - \eta\lambda_i + \frac{w^T(k)Rw(k)}{w^T(k)w(k)} \cdot \frac{\eta}{w^T(k)w(k)} \right] z_i(k), \tag{23}$$

$(i = 1, 2, \ldots, n)$, for all $k \geq 0$. Since R is a symmetric matrix, according to the relevant properties of the Rayleigh Quotient, we have that

$$\lambda_1 \geq \frac{w^T R w}{w^T w} \geq \lambda_n, \tag{24}$$

for all $w \neq 0$.

Next, we will prove some interesting lemmas and theorems to obtain the conditions to guarantee the convergence of (20).

Lemma 1: It holds that

$$(1 - \eta\lambda_n) s + \frac{\eta\lambda_n}{s} \geq 1,$$

for all $s \in [1, +\infty)$, if $\eta\lambda_n < 0.5$.

Proof: Define a differentiable function

$$f(s) = (1 - \eta\lambda_n) s^2 - s + \eta\lambda_n,$$

for all $s \geq 1$. It follows that

$$\dot{f}(s) = 2(1 - \eta\lambda_n) s - 1$$

for all $s \geq 1$. Since $\eta\lambda_n < 0.5$, clearly, $2(1 - \eta\lambda_n) > 1$. Then, $2(1 - \eta\lambda_n)s - 1 \geq 0$, i.e., $\dot{f}(s) \geq 0$, for all $s \geq 1$. This means that $f(s)$ is monotone increasing on the interval $[1, +\infty)$. Thus, for all $s \geq 1$, it holds that $f(s) \geq f(1) = 0$, i.e.,

$$(1 - \eta\lambda_n) s^2 - s + \eta\lambda_n \geq 0.$$

Clearly,

$$(1 - \eta\lambda_n) s + \frac{\eta\lambda_n}{s} \geq 1,$$

for all $s \in [1, +\infty)$. This completes the proof.

Lemma 2: If $\eta\lambda_1 < 0.5$ and $w^T(0)v_n \neq 0$, then there exists a constant $\delta > 0$ such that

$$\|w(k)\| \geq \delta,$$

for all $k \geq 0$, where $\delta = \min\{\|w(0)\|, 1 - \eta\lambda_1\}$.

Proof: It follows from (24) that

$$1 - \eta\lambda_i + \frac{w^T(k)Rw(k)}{w^T(k)w(k)} \cdot \frac{\eta}{w^T(k)w(k)} \geq 1 - \eta\lambda_1, \tag{25}$$

$(i = 1, 2, \ldots, n)$, for all $k \geq 0$. By $\eta\lambda_1 < 0.5$, it holds from (25) that

$$1 - \eta\lambda_i + \frac{w^T(k)Rw(k)}{w^T(k)w(k)} \cdot \frac{\eta}{w^T(k)w(k)} > 0, (i = 1, 2, \ldots, n), \tag{26}$$

for all $k \geq 0$. Next, two cases will be considered to complete the proof.

Case 1: $\|w(k)\| \geq 1$.

By $\|w(k)\| \geq 1$, from (23), (25) and (26), it follows that

$$\begin{aligned}
& \|w(k+1)\|^2 \\
= & \sum_{i=1}^{n} z_i^2(k+1) \\
= & \sum_{i=1}^{n} z_i^2(k) \cdot \left[1 - \eta\lambda_i + \frac{w^T(k)Rw(k)}{w^T(k)w(k)} \cdot \frac{\eta}{w^T(k)w(k)}\right]^2 \\
\geq & (1 - \eta\lambda_1)^2 \cdot \sum_{i=1}^{n} z_i^2(k) \\
\geq & (1 - \eta\lambda_1)^2 .
\end{aligned}$$

Thus, it holds that if $\|w(k)\| \geq 1$, then

$$\|w(k+1)\| \geq 1 - \eta\lambda_1. \tag{27}$$

Case 2: $\|w(k)\| < 1$.

Denote

$$\Theta(k) = w(k) - \eta R w(k) + \eta \frac{w^T(k)Rw(k)}{w^T(k)w(k)} \cdot w(k). \tag{28}$$

From (21) and (22), it holds that

$$\Theta(k) = \sum_{i=1}^{n} \left[1 - \eta\lambda_i + \eta\frac{w^T(k)Rw(k)}{w^T(k)w(k)}\right] \cdot z_i(k) v_i,$$

and

$$\|\Theta(k)\|^2 = \sum_{i=1}^{n} z_i^2(k) \cdot \left[1 - \eta\lambda_i + \eta\frac{w^T(k)Rw(k)}{w^T(k)w(k)}\right]^2 . \tag{29}$$

From (28), we have that

$$w^T(k) \cdot [\Theta(k) - w(k)] = 0,$$

i.e., $w(k) \perp [\Theta(k) - w(k)]$. Clearly,

$$\|\Theta(k)\| \geq \|w(k)\|. \tag{30}$$

By $\|w(k)\| < 1$, from (23), (26), (29) and (30), it follows that

$$\begin{aligned}
& \|w(k+1)\|^2 \\
= & \sum_{i=1}^{n} z_i^2(k+1) \\
= & \sum_{i=1}^{n} z_i^2(k) \cdot \left[1 - \eta\lambda_i + \frac{w^T(k)Rw(k)}{w^T(k)w(k)} \cdot \frac{\eta}{w^T(k)w(k)}\right]^2 \\
\geq & \sum_{i=1}^{n} z_i^2(k) \cdot \left[1 - \eta\lambda_i + \eta\frac{w^T(k)Rw(k)}{w^T(k)w(k)}\right]^2 \\
= & \|\Theta(k)\|^2 \\
\geq & \|w(k)\|^2.
\end{aligned}$$

Thus, it holds that if $\|w(k)\| < 1$, then

$$\|w(k+1)\| \geq \|w(k)\|. \tag{31}$$

From (27) and (31), clearly,

$$\|w(k)\| \geq \min\{\|w(0)\|, 1 - \eta\lambda_1\},$$

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

Lemma 3: If $\eta\lambda_1 < 0.5$ and $w^T(0)v_n \neq 0$, then there exists a constant $\zeta > 0$ such that

$$\|w(k)\| \leq \zeta,$$

for all $k \geq 0$.

Proof: Since $\eta\lambda_1 < 0.5$, clearly,

$$1-\eta\lambda_i+\frac{w^T(k)Rw(k)}{w^T(k)w(k)}\cdot\frac{\eta}{w^T(k)w(k)} \geq 1-\eta\lambda_1 > 0, \quad (32)$$

$(i = 1, 2, \ldots, n)$. Using Lemma 2, clearly, there exists a constant $\delta > 0$, such that

$$\frac{1}{w^T(k)w(k)} \leq \frac{1}{\delta^2}, \quad (33)$$

for all $k \geq 0$. In many practical applications, due to the noisy signals, the smallest eigenvalue λ_n of the covariance matrix R of the input data is usually larger than zero. Without loss of generality, we assume that $\lambda_n > 0$. Next, two cases will be considered to complete the proof.

Case 1: $\|w(k)\|^2 < \lambda_1/\lambda_n$.

By $\|w(k)\|^2 < \lambda_1/\lambda_n$, from (23), (24), (32) and (33), it follows that

$$\begin{aligned}
&\|w(k+1)\|^2 \\
= \;& \sum_{i=1}^{n} z_i^2(k+1) \\
= \;& \sum_{i=1}^{n} z_i^2(k)\cdot\left[1-\eta\lambda_i+\frac{w^T(k)Rw(k)}{w^T(k)w(k)}\cdot\frac{\eta}{w^T(k)w(k)}\right]^2 \\
\leq \;& \sum_{i=1}^{n} z_i^2(k)\cdot\left[1-\eta\lambda_n+\lambda_1\cdot\frac{\eta}{w^T(k)w(k)}\right]^2 \\
\leq \;& \sum_{i=1}^{n} z_i^2(k)\cdot\left(1-\eta\lambda_n+\frac{\eta\lambda_1}{\delta^2}\right)^2 \\
\leq \;& \left(1-\eta\lambda_n+\frac{\eta\lambda_1}{\delta^2}\right)^2\cdot\|w(k)\|^2 \\
< \;& \left(1-\eta\lambda_n+\frac{\eta\lambda_1}{\delta^2}\right)^2\cdot\frac{\lambda_1}{\lambda_n}.
\end{aligned}$$

Thus, it holds that if $\|w(k)\|^2 < \lambda_1/\lambda_n$, then

$$\|w(k+1)\| < \left(1-\eta\lambda_n+\frac{\eta\lambda_1}{\delta^2}\right)\cdot\sqrt{\frac{\lambda_1}{\lambda_n}}, \quad (34)$$

for all $k \geq 0$.

Case 2: $\|w(k)\|^2 \geq \lambda_1/\lambda_n$.

Since $\|w(k)\|^2 \geq \lambda_1/\lambda_n$, clearly,

$$1-\eta\lambda_n+\frac{\eta\lambda_1}{w^T(k)w(k)} \leq 1. \quad (35)$$

It follows from (23), (24), (32) and (35) that

$$\begin{aligned}
&\|w(k+1)\|^2 \\
= \;& \sum_{i=1}^{n} z_i^2(k+1) \\
= \;& \sum_{i=1}^{n} z_i^2(k)\cdot\left[1-\eta\lambda_i+\frac{w^T(k)Rw(k)}{w^T(k)w(k)}\cdot\frac{\eta}{w^T(k)w(k)}\right]^2 \\
\leq \;& \sum_{i=1}^{n} z_i^2(k)\cdot\left[1-\eta\lambda_n+\lambda_1\cdot\frac{\eta}{w^T(k)w(k)}\right]^2 \\
\leq \;& \left[1-\eta\lambda_n+\frac{\eta\lambda_1}{w^T(k)w(k)}\right]^2\cdot\|w(k)\|^2 \\
\leq \;& \|w(k)\|^2.
\end{aligned}$$

Thus, it holds that if $\|w(k)\|^2 \geq \lambda_1/\lambda_n$, then

$$\|w(k+1)\| \leq \|w(k)\|. \quad (36)$$

From (34) and (36), clearly, there exists a constant

$$\zeta = \max\left\{\|w(0)\|, \left(1-\eta\lambda_n+\frac{\eta\lambda_1}{\delta^2}\right)\cdot\sqrt{\frac{\lambda_1}{\lambda_n}}\right\},$$

such that $\|w(k)\| \leq \zeta$, for all $k \geq 0$. The proof is completed.

Theorem 1: If $\eta\lambda_1 < 0.5$ and $w^T(0)v_n \neq 0$, then it holds that

$$\lim_{k\to\infty} z_i(k) = 0, (i = 1, 2, \ldots, n-1).$$

Proof: Since $\eta\lambda_1 < 0.5$, clearly,

$$1-\eta\lambda_i+\frac{w^T(k)Rw(k)}{w^T(k)w(k)}\cdot\frac{\eta}{w^T(k)w(k)} \geq 1-\eta\lambda_1 > 0, \quad (37)$$

$(i = 1, 2, \ldots, n)$. Using Lemma 2, there exists a constant $\delta > 0$, such that

$$w^T(k)w(k) \geq \delta^2, \quad (38)$$

for all $k \geq 0$. From (23), (24), (37) and (38), it follows that

$$\begin{aligned}
&\left|\frac{z_i(k+1)}{z_n(k+1)}\right| \\
= \;& \frac{1-\eta\lambda_i+\eta w^T(k)Rw(k)/\left[w^T(k)w(k)\right]^2}{1-\eta\lambda_n+\eta w^T(k)Rw(k)/\left[w^T(k)w(k)\right]^2}\cdot\left|\frac{z_i(k)}{z_n(k)}\right| \\
\leq \;& \left\{1-\frac{\eta(\lambda_i-\lambda_n)}{1-\eta\lambda_n+\eta\lambda_1/\left[w^T(k)w(k)\right]}\right\}\cdot\left|\frac{z_i(k)}{z_n(k)}\right| \\
\leq \;& \left[1-\frac{\eta(\lambda_{n-1}-\lambda_n)}{1-\eta\lambda_n+\eta\lambda_1/\delta^2}\right]\cdot\left|\frac{z_i(k)}{z_n(k)}\right|, \\
\leq \;& \left[1-\frac{\eta(\lambda_{n-1}-\lambda_n)}{1-\eta\lambda_n+\eta\lambda_1/\delta^2}\right]^{k+1}\cdot\left|\frac{z_i(0)}{z_n(0)}\right|,
\end{aligned}$$

$(i = 1, 2, \ldots, n-1)$, for all $k \geq 0$. By $\eta\lambda_1 < 0.5$, clearly,

$$1 > 1-\frac{\eta(\lambda_{n-1}-\lambda_n)}{1-\eta\lambda_n+\eta\lambda_1/\delta^2} > 0.$$

Then,

$$\lim_{k\to\infty}\left|\frac{z_i(k)}{z_n(k)}\right| = 0, (i = 1, 2, \ldots, n-1).$$

Using Lemma 3, clearly, $\|w(k)\|$ is bounded and then $|z_n(k)|$ is also bounded. Thus,

$$\lim_{k\to\infty}|z_i(k)| = 0, (i = 1, 2, \ldots, n-1).$$

This completes the proof.

Theorem 2: If $\eta\lambda_1 < 0.5$ and $w^T(0)v_n \neq 0$, then it holds that

$$\lim_{k\to\infty} z_n(k) = \pm 1.$$

Proof: Using Theorem 1, clearly, $w(k)$ will converge to the direction of the minor component v_n, as $k \to \infty$. Suppose at time k_0, $w(k)$ has converged to the direction of v_n, i.e.,

$$w(k_0) = z_n(k_0)\cdot v_n. \tag{39}$$

From (20), it holds that

$$z_n(k+1) = z_n(k)\cdot\left[1 - \eta\lambda_n + \frac{\eta\lambda_n}{z_n^2(k)}\right], \tag{40}$$

for all $k \geq k_0$. Since $\eta\lambda_1 < 0.5$, clearly,

$$1 - \eta\lambda_n + \frac{\eta\lambda_n}{z_n^2(k)} > 0. \tag{41}$$

Then, it holds that from (40) and (41) that

$$|z_n(k+1)| = (1 - \eta\lambda_n)\,|z_n(k)| + \frac{\eta\lambda_n}{|z_n(k)|}. \tag{42}$$

for all $k \geq k_0$. And then,

$$\begin{aligned}\frac{|z_n(k+1)|}{|z_n(k)|} &= 1 + \eta\lambda_n\left(\frac{1}{|z_n(k)|^2} - 1\right)\\ &= \begin{cases} > 1, \text{if } |z_n(k)| < 1\\ = 1, \text{if } |z_n(k)| = 1\\ < 1, \text{if } |z_n(k)| > 1.\end{cases}\end{aligned} \tag{43}$$

From (43), clearly, $|z_n(k)| = 1$ is the stable equilibrium of (42). Next, three cases will be considered to complete the proof.

Case 1: $|z_n(k_0)| \geq 1$.

Since $\eta\lambda_1 < 0.5$, clearly, $\eta\lambda_n < \eta\lambda_1 < 0.5$. Using Lemma 1, it holds from (42) that

$$|z_n(k+1)| = (1 - \eta\lambda_n)\,|z_n(k)| + \frac{\eta\lambda_n}{|z_n(k)|} \geq 1,$$

for all $k \geq k_0$. Then, from (43), $|z_n(k)|$ is monotone decreasing for all $k \geq k_0$. Thus, $|z_n(k)|$ must converge to the equilibrium 1, as $k \to \infty$.

Case 2: $|z_n(k_0)| < 1$ and $|z_n(k)| < 1$, for all $k > k_0$.

By (43), $|z_n(k)|$ is monotone increasing for all $k \geq k_0$. Clearly, $|z_n(k)|$ will converge to the equilibrium 1, as $k \to \infty$.

Case 3: $|z_n(k_0)| < 1$ and there exists a positive integer $N(N > k_0)$, such that $|z_n(N)| \geq 1$.

Since $|z_n(N)| \geq 1$, in the same way as Case 1, it can be proven that $|z_n(k)|$ must converge to the equilibrium 1, as $k \to \infty$.

From the above three cases, we have that

$$\lim_{k\to\infty}|z_n(k)| = 1.$$

It follows from (40) and (41) that $z_n(k) > 0$ for all $k > k_0$ if $z_n(k_0) > 0$, and $z_n(k) < 0$ for all $k > k_0$ if $z_n(k_0) < 0$. Thus, $z_n(k)$ will converge if $|z_n(k)|$ converges. This completes the proof.

Using Theorem 1 and Theorem 2, we can obtain the following convergence result of (20).

Theorem 3: If $\eta\lambda_1 < 0.5$ and $w^T(0)v_n \neq 0$, then it holds that

$$\lim_{k\to\infty} w(k) = \pm v_n.$$

In the above theorem, it requires that the initial weight vector $w^T(0)v_n \neq 0$. In practical application, any small disturbance can result in $w^T(0)v_n \neq 0$, i.e., the condition is easy to meet. From this perspective, we can deduce that if $\eta\lambda_1 < 0.5$, then almost all trajectories will converge to a unit eigenvector associated with the smallest eigenvalue of the autocorrelation matrix R.

V. Convergence Analysis of The Proposed Algorithm

To establish the convergence results for the proposed sequential MCA algorithm, we will proceed the similar analysis to that of the original OJAm MCA algorithm. By taking conditional expectation operator $E\{w(k+1)/w(0), x(i), i < k\}$ to (17) and identifying the conditional expected value as the next iterate, a DDT system can be obtained for all $j = 1, 2, \ldots, p$ as:

$$\begin{aligned} w_j(k+1) = w_j(k) - \eta\Bigg\{ & [R + C_j(k)]\,w_j(k)\\ - & \frac{w_j^T(k)\,[R + C_j(k)]\,w_j(k)}{\left[w_j^T(k)w_j(k)\right]^2}\cdot w_j(k)\Bigg\},\end{aligned} \tag{44}$$

where $R = E[x(k)x^T(k)]$ is the autocorrelation matrix of $\{x(k)|x(k) \in R^n(k = 0, 1, 2, \ldots)\}$ and

$$C_1(k) = 0 \quad \text{and} \quad C_j(k) = \gamma\sum_{i=1}^{j-1}\frac{w_i(k)w_i^T(k)}{w_i^T(k)w_i(k)}, (p \geq j > 1). \tag{45}$$

Although (44) has the same structure as the original OJAm MCA algorithm (20), there exists a variant matrix $R + C_j(k)$

in (44). Using Theorem 3, clearly, the estimation of the largest eigenvalue of the matrix $R+C_j(k)$ is crucial to guarantee the convergence of (44). Next, we will prove an interesting theorem to give a practical method to estimate an upper bound of the largest eigenvalue of the matrix $R+C_j(k)$.

Theorem 4: If $\eta\,[\lambda_1+(p-1)\gamma] < 0.5$ and $w_j^T(0)v_{n-j+1}\neq 0 (j=1,2,\ldots,p)$, then it holds that

$$\lim_{k\to\infty} w_j(k)=\pm v_{n-j+1}, (j=1,2,\ldots,p),$$

where $\gamma>\lambda_1$ is a constant and $p(n\geq p\geq 1)$ is the number of the extracted minor components.

Proof: Clearly, $R+C_j(k)$ is a symmetric nonnegative definite matrix for all $k\geq 0$. Then, denote the ith eigenvalue of the matrix $R+C_j(k)$ by $\lambda_i^{(j)}$ that are ordered by

$$\lambda_1^{(j)}>\lambda_2^{(j)}>\ldots>\lambda_n^{(j)}\geq 0.$$

Since $C_1(k)=0$, clearly, $\lambda_1^{(1)}=\lambda_1$, where λ_1 is the largest eigenvalue of the autocorrelation matrix R. From (45), we have that

$$R+C_{j+1}(k)=R+C_j(k)+\gamma\frac{w_j(k)w_j^T(k)}{w_j^T(k)w_j(k)}, \quad (46)$$

$(j=1,2,\ldots,p-1)$. Using Interlacing Eigenvalue Theorem [19], it holds from (46) that

$$\lambda_1^{(j+1)}\geq\lambda_1^{(j)}\geq\lambda_2^{(j+1)}\geq\ldots\geq\lambda_n^{(j+1)}\geq\lambda_n^{(j)}, \quad (47)$$

$(j=1,2,\ldots,p-1)$, and

$$\sum_{i=1}^{n}\lambda_i^{(j+1)}-\sum_{i=1}^{n}\lambda_i^{(j)}=\gamma, (j=1,2,\ldots,p-1). \quad (48)$$

From (47) and (48), it holds that

$$\lambda_1^{(j+1)}\geq\lambda_1^{(j)} \quad\text{and}\quad \lambda_1^{(j+1)}\leq\lambda_1^{(j)}+\gamma. \quad (49)$$

Then,

$$\lambda_1^{(p)}\geq\lambda_1^{(p-1)}\geq\ldots\geq\lambda_1^{(2)}\geq\lambda_1^{(1)}=\lambda_1, \quad (50)$$

and

$$\lambda_1^{(p)}\leq\lambda_1^{(p-1)}+\gamma\leq\ldots\leq\lambda_1^{(1)}+(p-1)\gamma=\lambda_1+(p-1)\gamma. \quad (51)$$

Since $\eta\,[\lambda_1+(p-1)\gamma]<0.5$, it follows from (50) and (51) that,

$$\eta<\frac{1}{2\,[\lambda_1+(p-1)\gamma]}\leq\frac{1}{2\lambda_1^{(p)}}\leq\ldots\leq\frac{1}{2\lambda_1^{(1)}}=\frac{1}{2\lambda_1}. \quad (52)$$

By Theorem 3, (52) guarantees that for all $j=1,2,\ldots,p$, $w_j(k)$ will converge to the unit eigenvector associated the smallest eigenvalue of the matrix $R+C_j(k)$ in (44). Clearly, we have that when $j=1$,

$$\lim_{k\to\infty} w_1(k)=\pm v_n. \quad (53)$$

From (1), (45) and (53), then,

$$\begin{aligned}&\lim_{k\to\infty} R+C_2(k)\\ =&\sum_{i=1}^{n}\lambda_i v_i v_i^T+\gamma v_n v_n^T\\ =&\sum_{i=1}^{n-1}\lambda_i v_i v_i^T+(\gamma+\lambda_n)v_n v_n^T.\end{aligned} \quad (54)$$

Clearly, as $k\to\infty$, the matrix $R+C_2(k)$ has the ordered eigenvalues:

$$\gamma+\lambda_n>\lambda_1>\ldots>\lambda_{n-2}>\lambda_{n-1},$$

and the corresponding eigenvectors are $v_n, v_1,\ldots,v_{n-2},v_{n-1}$. Thus, v_n becomes the eigenvector associated with the largest eigenvalue and v_{n-1} becomes the eigenvector associated with the smallest eigenvalue of the matrix $R+C_2(k)$, as $k\to\infty$. Using Theorem 3, it holds from (52) that

$$\lim_{k\to\infty} w_2(k)=\pm v_{n-1}.$$

In the same way, we can prove that $w_j(k)$ will converge to the corresponding minor component v_{n-j+1}, for all $j(p\geq j\geq 1)$. The proof is completed.

VI. Simulation results

The simulations presented in this section will illustrate the effectiveness of the proposed sequential MCA algorithm. We randomly generate a 4×4 correlation matrix as:

$$R=\begin{bmatrix}0.6400 & 0.1581 & 0.1936 & -0.0769\\ 0.1581 & 0.7164 & 0.0610 & 0.0322\\ 0.1936 & 0.0610 & 0.4732 & 0.3507\\ -0.0769 & 0.0322 & 0.3507 & 0.6484\end{bmatrix}.$$

The initial weight vector is taken as:

$$w(0)=[0.2643 \quad 0.3684 \quad 0.1601 \quad 0.8161]^T.$$

In the first simulation, we will extract the first minor component of the autocorrelation matrix R. Let the learning rate $\eta=0.25<0.5/\lambda_1$. Fig1 shows the convergence of the component $z_i(k)$ of $w(k)$ in (20), where $z_i(k)=w^T(k)v_i$ is the coordinate of $w(k)$ on the direction of the eigenvector $v_i (i=1,2,3,4)$. In the simulation result, $z_i(k)(i=1,2,3)$ converges to zero and $z_4(k)$ converges to -1, as $k\to\infty$, which is consistent with the convergence results obtained in Theorem 3.

In the second simulation, we will extract all minor components of the autocorrelation R, sequentially. Let $\gamma=1>\lambda_1$ and $\eta=0.12<0.5/[\lambda_1+(p-1)\gamma]$. In order to further measure the convergence and accuracy of the sequential MCA algorithm, we compute the norm of $w_i(k)$ and the

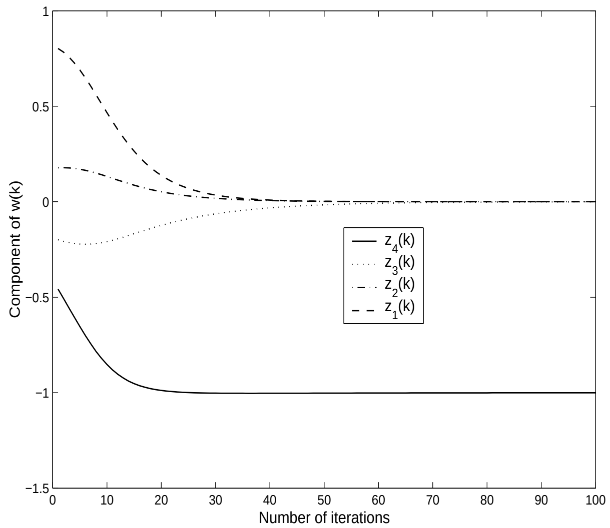

Figure. 1: Convergence of component of $w(k)$.

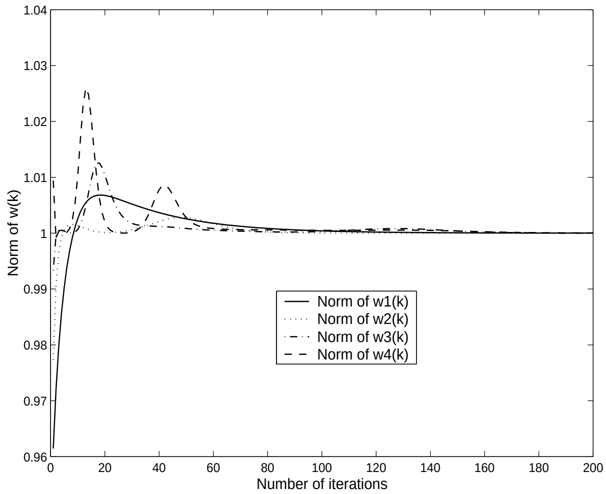

Figure. 3: Convergence of $\|w_i(k)\|$.

direction cosine of $w_i(k)$ by [20]:

$$\text{Direction Cosine of } w_i(k) = \frac{\left|w_i^T(k) \cdot v_{n-i+1}\right|}{\|w_i(k)\| \cdot \|v_{n-i+1}\|},$$

$(i = 1, 2, 3, 4)$. Fig2 shows that all Direction Cosine of $w_i(k)(i = 1, 2, 3, 4)$ converge to 1. This means that $w_i(k)(i = 1, 2, 3, 4)$ converges to the direction of the corresponding minor component $v_{n-i+1}(i = 1, 2, 3, 4)$, respectively. Fig3 shows that the norm of $w_i(k)(i = 1, 2, 3, 4)$ converge to 1. Clearly, the simulation result illustrated by Fig 2 and Fig 3 is consistent with the convergence results obtained in Theorem 4.

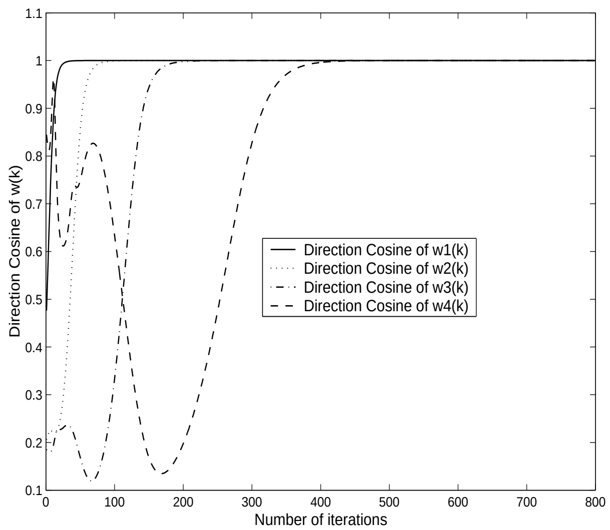

Figure. 2: Convergence of direction cosine of $w_i(k)$.

Besides the above simulation results, further simulations with high dimensions persist to show that the proposed sequential MCA algorithm has a satisfactory convergent result.

VII. Conclusions

By extending OJAm MCA algorithm, we propose a globally convergent MCA algorithm that can extract multiple minor components from the input data sequentially. The convergence of the proposed learning algorithm is studied via DDT method. Some sufficient conditions for convergence of the learning algorithm with constant learning rate are obtained.

Acknowledgments

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.

References

[1] K. Gao, M. O. Ahmad, and M. N. Swamy. Learning algorithm for total least squares adaptive signal processing, Electron. Lett., 28 (4), pp. 430-432, 1992.

[2] K. Gao, M. O. Ahmad, and M. N. Swamy. A constrained anti-Hebbian learning algorithm for total least-squares estimation with applications to adaptive FIR and IIR filtering, IEEE Trans. Circuits Syst. Part II, 41, pp. 718-729, 1994.

[3] R. Klemm. Adaptive airborne mti: An auxiliary channel approach, Proc. Inst. Elect. Eng., pt. F, 134, pp. 269-276, 1987.

[4] S. Barbarossa, E. Daddio, and G. Galati. Comparison of optimum and linear prediction technique for clutter cancellation, Proc. Inst. Elect. Eng., pt. F, 134, pp. 277-282, 1987.

[5] G. Cirrincione. A neural approach to the structure from motion problem, Ph.D. dissertation, LIS INPG Grenoble, 1998.

[6] L. Xu, E. Oja, and C. Suen. Modified Hebbian learning for curve and surface fitting, Neural Networks, 5, pp. 441-457, 1992.

[7] J. W. Griffiths. Adaptive array processing, a tutorial, Proc. Inst. Elect. Eng., pt. F, 130, pp. 3-10, 1983.

[8] G. Mathew and V. Reddy. Development and analysis of a neural network approach to Pisarenko's harmonic retrieval method, IEEE Trans. Signal Processing, 42, pp. 663-667, 1994.

[9] G. Mathew and V. Reddy. Orthogonal eigensubspace estimation using neural networks, IEEE Trans. Signal Processing, 42, pp. 1803-1811, 1994.

[10] R. Schmidt. Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propagation, 34, pp. 276-280, 1986.

[11] F. L. Luo, R. Unbehauen and A. Cichocki. A Minor Component Analysis Algorithm, Neural Networks, 10(2), pp. 291-297, 1997.

[12] F. L. Luo and R. Unbehauen. A Minor Subspace Analysis Algorithm, IEEE Trans. Neural Networks, 8(5), pp. 1149-1155, 2005.

[13] T. P. Chen, S. Amari and N. Murata. Sequential Extraction of Minor Components, Neural Processing Letters, 13, pp. 195-201, 2001.

[14] A. Taleb and G. Cirrincione. Against the convergence of the minor component analysis neurons, IEEE Trans. Neural Networks, 10, pp. 207-210, 1999.

[15] D. Z. Feng, W. X. Zheng and Y. Jia. Neural Network Learning Algorithms for Tracking Minor Subspace in Hing-Dimensional Data Stream, IEEE Trans. Neural Networks, 16(3), pp. 513-521, 2005.

[16] L. Ljung. Aanlysis of recursive stochastic algorithms, IEEE Trans. Automat. Contr., 22, pp. 551-575, 1977.

[17] P. J. Zufiria. On the discrete-time dynamics of the basic Hebbian neural-network nodes, IEEE Trans. Neural Networks, 13(6), pp. 1342-1352, 2002.

[18] 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 Networks, 16(6), pp. 1318-1327, 2005.

[19] G. H. Golub and C. F. van Loan. Matrix Computation, 2nd ed., The John Hopkins Unversity Press, 1989.

[20] C. Chatterjee, Z. Kang and V. P. Roychowdhury. Algorithm for accelerated convergence of adaptive PCA, IEEE Trans. Neural Networks, 11(2), pp. 338-355, 2000.

Author Biographies

Dezhong Peng received the B.S. degree in applied mathematics from University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1998, the M.S. degree in computer software and theory from UESTC, in 2001.

He is currently working toward the Ph.D. degree in Computational Intelligence Laboratory, School of Computer Science and Engineering, UESTC. His current 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 1999 to 2001, he was a Research Associate in the Department of Computer Science and Engineering at The Chinese University of Hong Kong. From 2001 to 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.