Chapter 6
Deterministic Discrete-Time System for the Analysis of Iterative Algorithms

6.1 Introduction

The convergence of neural network-based PCA or MCA learning algorithms is a difficult topic for direct study and analysis. Traditionally, based on the stochastic approximation theorem, the convergence of these algorithms is indirectly analyzed via corresponding DCT systems. The stochastic approximation theorem requires that some restrictive conditions must be satisfied. One important condition is that the learning rates of the algorithms must approach zero, which is not a reasonable requirement to be imposed in many practical applications. Clearly, the restrictive condition is difficult to be satisfied in many practical applications, where a constant learning rate is usually used due to computational roundoff issues and tracking requirements. Besides the DCT system, Lyapunov function method, differential equations method, etc., are also used to analyze the convergence of PCA algorithms. For example, in [1], a Lyapunov function was proposed for globally characterizing Oja's DCT model with a single neuron. Another single-neuron generalized version of Oja's DCT net was studied in [2] by explicitly solving the system of differential equations. The global behavior of a several-neuron Oja's DCT net was determined in [3] by explicitly solving the equations of the model, whereas [4] addressed a qualitative analysis of the generalized forms of this DCT network.

All these studies of DCT formulations are grounded on restrictive hypotheses so that the fundamental theorem of stochastic approximation can be applied. However, when some of these hypotheses cannot be satisfied, how to study the convergence of the original stochastic discrete formulation? In order to analyze the convergence of neural network-based PCA or MCA learning algorithms, several methods have been proposed, i.e., DCT, SDT, and DDT methods. The DCT method, first formalized by [5, 6], is based on a fundamental theorem of stochastic approximation theory. Thus, it is an approximation analysis method. The SDT method is a direct analysis method and it can analyze the temporal behavior of algorithm and derive

X. Kong et al., *Principal Component Analysis Networks and Algorithms*,
DOI 10.1007/978-981-10-2915-8_6

the relation between the dynamic stability and learning rate [7]. The DDT method, as a bridge between DCT and SDT methods, transforming the original SDT system into a corresponding DDT system, and preserving the discrete-time nature of the original SDT systems, can shed some light on the convergence characteristics of SDT systems [8]. Recently, the convergence of many PCA or MCA algorithms has been widely studied via the DDT method [8–13].

The objective of this chapter is to study the DDT method, analyze the convergence of PCA or MCA algorithms via DDT method to obtain some sufficient conditions to guarantee the convergence, and analyze the stability of these algorithms. The remainder of this chapter is organized as follows. A review of performance analysis methods for neural network-based PCA/MCA algorithms is presented in Sect. 6.2. The main content, a DDT system of a novel MCA algorithm is introduced in Sect. 6.3. Furthermore, a DDT system of a unified PCA and MCA algorithm is introduced in Sect. 6.4, followed by the summary in Sect. 6.5.

6.2 Review of Performance Analysis Methods for Neural Network-Based PCA Algorithms

6.2.1 Deterministic Continuous-Time System Method

According to the stochastic approximation theory (see [5, 6]), if certain conditions are satisfied, its corresponding DCT systems can represent the SDT system effectively (i.e., their asymptotic paths are close with a large probability) and eventually the PCA/MCA solution tends with probability 1 to the uniformly asymptotically stable solution of the ODE. From a computational point of view, the most important conditions are the following:

1. $x(t)$ is zero-mean stationary and bounded with probability 1.
2. $\alpha(t)$ is a decreasing sequence of *positive* scalars.
3. $\Sigma_t \alpha(t) = \infty$.
4. $\Sigma_t \alpha^P(t) < \infty$ for some p.
5. $\lim_{t\to\infty} \sup\left[\frac{1}{\alpha(t)} - \frac{1}{\alpha(t-1)}\right] < \infty$.

For example, the sequence $\alpha(t) = \text{const} \cdot t^{-\gamma}$ satisfies Conditions 2–5 for $0 < \gamma \leq 1$. The fourth condition is less restrictive than the Robbins–Monro condition $\Sigma_t \alpha^2(t) < \infty$, which is satisfied, for example, only by $\alpha(t) = \text{const} \cdot t^{-\gamma}$ with $1/2 < \gamma \leq 1$.

For example, MCA EXIN algorithm can be written as follows:

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) - \frac{\alpha(t)y(t)}{\|\boldsymbol{w}(t)\|_2^2}\left[\boldsymbol{x}(t) - \frac{y(t)\boldsymbol{w}(t)}{\|\boldsymbol{w}(t)\|_2^2}\right], \tag{6.1}$$

and its corresponding deterministic continuous-time (DCT) systems is

$$\frac{d\boldsymbol{w}(t)}{dt} = -\frac{1}{\|\boldsymbol{w}(t)\|_2^2}\left[\boldsymbol{R} - \frac{\boldsymbol{w}^{\mathrm{T}}(t)\boldsymbol{R}\boldsymbol{w}(t)}{\|\boldsymbol{w}(t)\|_2^2}\right]\boldsymbol{w}(t) = -\frac{1}{\|\boldsymbol{w}(t)\|_2^2}[\boldsymbol{R} - r(\boldsymbol{w},\boldsymbol{R})\boldsymbol{I}]\boldsymbol{w}(t). \tag{6.2}$$

For the convergence proof using deterministic continuous-time system method, refer to the proof of Theorem 16 in [7] for details.

6.2.2 *Stochastic Discrete-Time System Method*

Using only the ODE approximation does not reveal some of the most important features of these algorithms [7]. For instance, it can be shown that the constancy of the weight modulus for OJAn, Luo, and MCA EXIN, which is the consequence of the use of the ODE, is not valid, except, as a very first approximation, in approaching the minor component [7]. The stochastic discrete-time system method has led to the very important problem of the sudden divergence [7]. In the following, we will analyze the performance of Luo MCA algorithm using the stochastic discrete-time system method.

In [14, 15], Luo proposed a MCA algorithm, which is

$$\boldsymbol{w}(t+1) = \boldsymbol{w}(t) - \alpha(t)\|\boldsymbol{w}(t)\|_2^2\left[y(t)\boldsymbol{x}(t) - \frac{y^2(t)}{\|\boldsymbol{w}(t)\|_2^2}\boldsymbol{w}(t)\right]. \tag{6.3}$$

Since (6.3) is the gradient flow of the RQ and using the property of orthogonality of RQ, it holds that

$$\boldsymbol{w}^{\mathrm{T}}(t)\left\{y(t)\boldsymbol{x}(t) - \frac{y^2(t)}{\|\boldsymbol{w}(t)\|_2^2}\boldsymbol{w}(t)\right\} = 0, \tag{6.4}$$

i.e., *the weight increment at each iteration is orthogonal to the weight direction.* The squared modulus of the weight vector at instant $t + 1$ is then given by

$$\|\boldsymbol{w}(t+1)\|_2^2 = \|\boldsymbol{w}(t)\|_2^2 + \frac{\alpha^2(t)}{4}\|\boldsymbol{w}(t)\|_2^6\|\boldsymbol{x}(t)\|_2^4\sin^2 2\vartheta_{\boldsymbol{xw}}, \tag{6.5}$$

where $\vartheta_{\boldsymbol{xw}}$ is the angle between the direction of $\boldsymbol{x}(t)$ and $\boldsymbol{w}(t)$. From (6.5), we can see that (1) Except for particular conditions, *the weight modulus always increases*, $\|\boldsymbol{w}(t+1)\|_2^2 > \|\boldsymbol{w}(t)\|_2^2$. These particular conditions, i.e., all data in exact particular directions, are too rare to be found in a noisy environment. (2) $\sin^2 2\vartheta_{\boldsymbol{xw}}$ is a positive function with peaks within the interval $(-\pi, \pi]$. This is one of the possible interpretations of the oscillatory behavior of weight modulus.

The remaining part of this section is the convergence analysis of Dougla's MCA algorithm via the SDT Method. The purpose of this section is to analyze the temporal behavior of Dougla's MCA algorithm and the relation between the dynamic stability and learning rate, by using mainly the SDT system following the approach in [7].

Indeed, using only the ODE approximation does not reveal some of the most important features of MCA algorithms, and the ODE is only the very first approximation, in approaching the minor component. After the MC direction has been approached, how is the rule of the weight modulus?

From Dougla's MCA, it holds that

$$\begin{aligned}\|w(t+1)\|^2 &= w^{\mathrm{T}}(t+1)w(t+1) = \{w(t) - \alpha(t)[\|w(t)\|^4 y(t)x(t) - y^2(t)w(t)]\}^{\mathrm{T}} \cdot \{w(t) - \alpha(t)[\|w(t)\|^4 y(t)x(t) - y^2(t)w(t)]\} \\ &= \|w(t)\|^2 - 2\alpha(t)(\|w(t)\|^4 y^2(t) - y^2(t)\|w(t)\|^2) + \alpha^2(t)(\|w(t)\|^8 y^2(t)\|x(t)\|^2 - 2\|w(t)\|^4 y^4(t) + y^4(t)\|w(t)\|^2) \\ &= \|w(t)\|^2 + 2\alpha(t)y^2(t)\|w(t)\|^2(1 - \|w(t)\|^2) + O(\alpha^2(t)) \\ &\doteq \|w(t)\|^2 + 2\alpha(t)y^2(t)\|w(t)\|^2(1 - \|w(t)\|^2).\end{aligned} \tag{6.6}$$

Hence, if the learning factor is small enough and the input vector is bounded, we can make such analysis as follows by neglecting the second-order terms of the $\alpha(t)$.

$$\frac{\|w(t+1)\|^2}{\|w(t)\|^2} \doteq 1 + 2\alpha(t)y^2(t)(1 - \|w(t)\|^2) = \begin{cases} > 1 & \text{for} \quad \|w(0)\|^2 < 1 \\ < 1 & \text{for} \quad \|w(0)\|^2 < 1\,. \\ = 1 & \text{for} \quad \|w(0)\|^2 = 1 \end{cases} \tag{6.7}$$

This means that $\|w(t+1)\|^2$ tends to one whether $\|w(t)\|^2$ is equal to one or not, which is called the one-tending property (OTP), i.e., the weight modulus remains constant $\left(\|w(t)\|^2 \to 1\right)$.

To use the stochastic discrete laws is a direct analytical method. In fact, the study of the stochastic discrete learning laws of the Douglas's algorithm is an analysis of their dynamics.

Define

$$r' = \frac{|w^{\mathrm{T}}(t+1)x(t)|^2}{\|w(t+1)\|^2}, \quad r = \frac{|w^{\mathrm{T}}(t)x(t)|^2}{\|w(t)\|^2},$$
$$\rho(\alpha) = \frac{r'}{r} \geq 1, \quad p = \|w(t)\|^2, \quad u = y^2(t).$$

The two scalars r' and r represent, respectively, the squared perpendicular distance between the input $x(t)$ and the data-fitting hyperplane whose normal is given by the weight and passes through the origin, after and before the weight increment. Recalling the definition of MC, we should have $r' \leq r$. If this inequality is not valid, this means that the learning law increases the estimation error due to the disturbances caused by noisy data. When this disturbance is too large, it will make $\boldsymbol{w}(t)$ deviate drastically from the normal learning, which may result in divergence or fluctuations (implying an increased learning time).

Theorem 6.1

$$\text{If } \alpha > \frac{2}{p\|\boldsymbol{x}(t)\|^2(p - 2\cos^2\theta_{\boldsymbol{xw}})} \quad \wedge \quad p\|\boldsymbol{x}(t)\|^2(p - 2\cos^2\theta_{\boldsymbol{xw}}) > 0,$$

then $r' > r$, which implies divergence.

Proof From Eq. (6.2), we have

$$\begin{aligned} \boldsymbol{w}^{\mathrm{T}}(t+1)\boldsymbol{x}(t) &= y(t) - \alpha[\|\boldsymbol{w}(t)\|^4 y(t)\|\boldsymbol{x}(t)\|^2 - y^3(t)] \\ &= y(t)(1 - \alpha[\|\boldsymbol{w}(t)\|^4\|\boldsymbol{x}(t)\|^2 - y^2(t)]) \end{aligned} \tag{6.8}$$

$$\begin{aligned} \|\boldsymbol{w}(t+1)\|^2 = \boldsymbol{w}^{\mathrm{T}}(t+1)\boldsymbol{w}(t+1) &= \|\boldsymbol{w}(t)\|^2 - 2\alpha(t)(\|\boldsymbol{w}(t)\|^4 y^2(t) - y^2(t)\|\boldsymbol{w}(t)\|^2) \\ &+ \alpha^2(t)(\|\boldsymbol{w}(t)\|^8 y^2(t)\|\boldsymbol{x}(t)\|^2 - 2\|\boldsymbol{w}(t)\|^4 y^4(t) + y^4(t)\|\boldsymbol{w}(t)\|^2). \end{aligned} \tag{6.9}$$

Therefore,

$$\begin{aligned} \rho(\alpha) = \frac{r'}{r} = \frac{(\boldsymbol{w}^{\mathrm{T}}(t+1)\boldsymbol{x}(t))^2}{\|\boldsymbol{w}(t+1)\|^2}\frac{\|\boldsymbol{w}(t)\|^2}{(y(t))^2} &= \frac{(1 - \alpha(t)[\|\boldsymbol{w}(t)\|^4\|\boldsymbol{x}(t)\|^2 - y^2(t)])^2}{1 - 2\alpha(t)y^2(t)(\|\boldsymbol{w}(t)\|^2 - 1) + \alpha^2 E} \\ &= \frac{(1 - \alpha q)^2}{1 - 2\alpha u(p-1) + \alpha^2 E}, \end{aligned} \tag{6.10}$$

where $q = \left(\|\boldsymbol{x}(t)\|^2 p^2 - u\right)$ and $E = \left(up^3\|\boldsymbol{x}(t)\|^2 - 2u^2 p + u^2\right)$.

Then, $\rho(\alpha) > 1$ (dynamic instability) if and only if

$$(1 - \alpha q)^2 > 1 - 2\alpha u(p-1) + \alpha^2\left(up^3\|\boldsymbol{x}(t)\|^2 - 2u^2 p + u^2\right). \tag{6.11}$$

Notice that $u/p = \|\boldsymbol{x}(t)\|^2\cos^2\theta_{zw}$.

From (6.11), it holds that

$$\begin{aligned} &\alpha^2\left[p^4\|\boldsymbol{x}(t)\|^4\sin^2\theta_{\boldsymbol{xw}} - 2p^3\|\boldsymbol{x}(t)\|^4\cos^2\theta_{\boldsymbol{xw}}\sin^2\theta_{\boldsymbol{xw}}\right] \\ &\quad > 2\alpha\|\boldsymbol{x}(t)\|^2 p^2\sin^2\theta_{\boldsymbol{xw}}. \end{aligned} \tag{6.12}$$

The dynamic instability condition is then

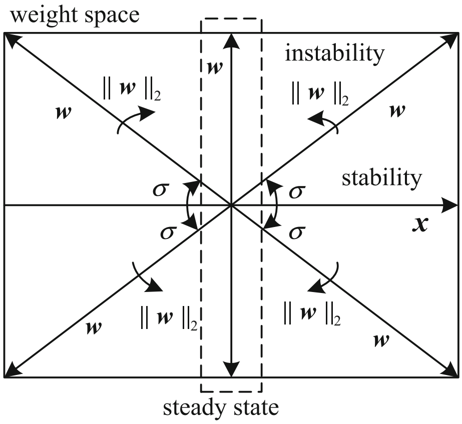

Fig. 6.1 The stability subspace of the weight space with regard to the input vector (two-dimensional space) of proposed algorithm

$$\alpha > \frac{2}{p\|\boldsymbol{x}(t)\|^2(p - 2\cos^2\theta_{\boldsymbol{xw}})} \quad \wedge \quad p\|\boldsymbol{x}(t)\|^2\left(p - 2\cos^2\theta_{\boldsymbol{xw}}\right) > 0. \tag{6.13}$$

The second condition implies the absence of the negative instability. It can be rewritten as

$$\cos^2\theta_{\boldsymbol{xw}} \le \frac{p}{2}. \tag{6.14}$$

In reality, the second condition is included in the first one. Considering the case $0 < \alpha_b \le \gamma < 1$, it holds that

$$\cos^2\theta_{\boldsymbol{xw}} \le \frac{p}{2} - \frac{1}{\gamma p\|\boldsymbol{x}(t)\|^2} = \Upsilon, \tag{6.15}$$

which is more restrictive than (6.14). Figure 6.1 shows this condition, where $\sigma = \arccos\sqrt{\Upsilon}$. From (6.15), we can see that the decrease of γ and p increases the domain of σ and then increases the stability. From Fig. 6.1, it is apparent that in the transient (in general low θ_{XW}), there are less fluctuations and this is beneficial to the stability.

This completes the proof.

6.2.3 *Lyapunov Function Approach*

Lyapunov function approach has also been applied in the convergence and stability analysis. For details, see references [7, 16, 17].

6.2.4 Deterministic Discrete-Time System Method

Traditionally, the convergence of neural network learning algorithms is analyzed via DCT systems based on a stochastic approximation theorem. However, there exist some restrictive conditions when using stochastic approximation theorem. One crucial condition is that the learning rate in the learning algorithm must converge to zero, which is not suitable in most practical applications because of the roundoff limitation and tracking requirements [8, 13]. In order to overcome the shortcomings of the DCT method, Zurifia proposed DDT method [8]. Different from the DCT method, the DDT method allows the learning rate to be a constant and can be used to indirectly analyze the dynamic behaviors of stochastic learning algorithms. Since the DDT method is more reasonable for studying the convergence of neural network algorithms than the traditional DCT method, it has been widely used to study many neural network algorithms [8, 10–13, 18–20].

6.3 DDT System of a Novel MCA Algorithm

In this section, we will analyze the convergence and stability of a class of self-stabilizing MCA algorithms via a DDT method. Some sufficient conditions are obtained to guarantee the convergence of these learning algorithms. Simulations are carried out to further illustrate the theoretical results achieved. It can be concluded that these self-stabilizing algorithms can efficiently extract the MCA, and they outperform some existing MCA methods.

In Sect. 6.3.1, a class of self-stabilizing learning algorithms is presented. In Sect. 6.3.2, the convergence and stability analysis of these algorithms via DDT method are given. In Sect. 6.3.3, computer simulation results on minor component extraction and some conclusions are presented.

6.3.1 Self-stabilizing MCA Extraction Algorithms

Consider a single linear neuron with the following input–output relation: $y(k) = \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{X}(k), k = 0, 1, 2, \cdots$, where $y(k)$ is the neuron output, the input sequence $\{\boldsymbol{X}(k)|\boldsymbol{X}(k) \in \boldsymbol{R}^n(k = 0, 1, 2, \cdots)\}$ is a zero-mean stationary stochastic process, and $\boldsymbol{W}(k) \in \boldsymbol{R}^n(k = 0, 1, 2, \cdots)$ 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 $\boldsymbol{W}(k)$ adaptively. Here, based on the OJA + algorithm [21], we add a penalty term $(1 - \|\boldsymbol{W}(t)\|^{2+\alpha})\boldsymbol{R}\boldsymbol{W}$ to OJA + and present a class of MCA algorithms as follows:

$$\dot{\boldsymbol{W}} = -\|\boldsymbol{W}\|^{2+\alpha}\boldsymbol{R}\boldsymbol{W} + (\boldsymbol{W}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{W} + 1 - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})\boldsymbol{W}, \tag{6.16}$$

where $\boldsymbol{R} = E[\boldsymbol{X}(k)\boldsymbol{X}^{\mathrm{T}}(k)]$ is the correlation matrix of the input data and the integer $0 \le \alpha \le 2$. The parameter α can be real-valued. However, for the simplicity of theoretical analysis and practical computations, it would be convenient to choose α as an integer. Considering the needs in the proofs of latter theorems, the upper limit of α is 2. It is worth noting that Algorithm (6.16) coincides with the Chen rule for minor component analysis [22] in the case $\alpha = 0$. When $\alpha > 0$, these algorithms are very similar to the Chen algorithm and can be considered as modifications of the Chen algorithm. Therefore, for simplicity, we refer to all of them as Chen algorithms.

The stochastic discrete-time system of (6.16) can be written as follows:

$$\boldsymbol{W}(k+1) = \boldsymbol{W}(k) - \eta\Big[\|\boldsymbol{W}(k)\|^{2+\alpha}y(k)\boldsymbol{X}(k) - (y^2(k) + 1 - \|\boldsymbol{W}(k)\|^2)\boldsymbol{W}(k)\Big], \tag{6.17}$$

where $\eta(0 < \eta < 1)$ is the learning rate. From (6.17), it follows that

$$\begin{aligned}\|\boldsymbol{W}(k+1)\|^2 - \|\boldsymbol{W}(k)\|^2 &= -2\eta\|\boldsymbol{W}(k)\|^2\Big[y^2(k)(\|\boldsymbol{W}(k)\|^{\alpha} - 1) + \Big(\|\boldsymbol{W}(k)\|^2 - 1\Big)\Big] + O(\eta^2) \\ &\doteq -2\eta\|\boldsymbol{W}(k)\|^2(\|\boldsymbol{W}(k)\| - 1)Q\big(y^2(k), \|\boldsymbol{W}(k)\|\big),\end{aligned} \tag{6.18}$$

where $Q(y^2(k), \|\boldsymbol{W}(k)\|) = y^2(k)(\|\boldsymbol{W}(k)\|^{\alpha-1} + \|\boldsymbol{W}(k)\|^{\alpha-2} + , \cdots, \|\boldsymbol{W}(k)\| + 1) + (\|\boldsymbol{W}(k)\| + 1)$ is a positive efficient. For a relatively small constant learning rate, the second-order term is very small and can be omitted. Thus, from (6.18), we can claim that Algorithm (6.17) has self-stabilizing property [23].

6.3.2 *Convergence Analysis via DDT System*

From $y(k) = \boldsymbol{X}^{\mathrm{T}}(k)\boldsymbol{W}(k) = \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{X}(k)$, by taking the conditional expectation $E\{\boldsymbol{W}(k+1)/\boldsymbol{W}(0), \boldsymbol{X}(i), i < k\}$ to (6.17) and identifying the conditional expectation as the next iterate, a DDT system can be obtained as

$$\begin{aligned}\boldsymbol{W}(k+1) &= \boldsymbol{W}(k) \\ &\quad - \eta\Big[\|\boldsymbol{W}(k)\|^{2+\alpha}\boldsymbol{R}\boldsymbol{W}(k) - \Big(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2\Big)\boldsymbol{W}(k)\Big],\end{aligned} \tag{6.19}$$

where $\boldsymbol{R} = E[\boldsymbol{X}(k)\boldsymbol{X}^{\mathrm{T}}(k)]$ is the correlation matrix of the input data. Here, we analyze the dynamics of (6.19) subject to η being some smaller constant to interpret the convergence of Algorithm (6.17) indirectly.

For the convenience of analysis, we next give some preliminaries. Since $\boldsymbol{R}$ is a symmetric positive definite matrix, there exists an orthonormal basis of $\Re^n$ composed of the eigenvectors of $\boldsymbol{R}$. Obviously, the eigenvalues of the autocorrelation matrix $\boldsymbol{R}$ are nonnegative. Assume that $\lambda_1, \lambda_2, \cdots, \lambda_n$ are all eigenvalues of $\boldsymbol{R}$ ordered by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} > \lambda_n > 0$. Suppose that $\{\boldsymbol{V}_i | i = 1, 2, \cdots, n\}$ is an orthogonal basis of $\boldsymbol{R}^n$ such that each $\boldsymbol{V}_{\mathrm{i}}$ is unit eigenvector of $\boldsymbol{R}$ associated with the eigenvalue λ_i. Thus, for each $k \geq 0$, the weight vector $\boldsymbol{W}(k)$ can be represented as

$$\boldsymbol{W}(k) = \sum_{i=1}^{n} z_i(k)\boldsymbol{V}_i, \tag{6.20}$$

where $z_i(k)(i = 1, 2, \ldots, n)$ are some constants. From (6.19) and (6.20), it holds that

$$z_i(k+1) = \left[1 - \eta\lambda_i \|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\left(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2\right)\right] z_i(k) \tag{6.21}$$

$(i = 1, 2, \ldots, n)$, for all $k \geq 0$.

According to the properties of Rayleigh Quotient [7], it clearly holds that

$$\lambda_n \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{W}(k) \leq \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) \leq \lambda_1 \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{W}(k), \tag{6.22}$$

for all $\boldsymbol{W}(k) \neq 0$, and $k \geq 0$.

Next, we perform the convergence analysis of DDT system (6.19) via the following Theorems 6.2–6.6.

Theorem 6.2 *Suppose that* $\eta\lambda_1 < 0.125$ *and* $\eta < 0.25$. *If* $\boldsymbol{W}^{\mathrm{T}}(0)\boldsymbol{V}_n \neq 0$ *and* $\|\boldsymbol{W}(0)\| \leq 1$, *then it holds that* $\|\boldsymbol{W}(k)\| < (1 + \eta\lambda_1)$, *for all* $k \geq 0$.

Proof From (6.19) and (6.20), it follows that

$$\begin{aligned}
\|\boldsymbol{W}(k+1)\|^2 &= \sum_{i=1}^{n} z_i^2(k+1) \\
&= \sum_{i=1}^{n} [1 - \eta\lambda_i \|\boldsymbol{W}(k)\|^{2+\alpha} + \eta(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2]^2 z_i^2(k) \\
&\leq \left[1 - \eta\left(\lambda_n \|\boldsymbol{W}(k)\|^{2+\alpha} - \lambda_1 \|\boldsymbol{W}(k)\|^2 + \|\boldsymbol{W}(k)\|^2 - 1\right)\right]^2 \sum_{i=1}^{n} z_i^2(k) \\
&\leq \left[1 - \eta\left(\lambda_n \|\boldsymbol{W}(k)\|^{2+\alpha} - \lambda_1 \|\boldsymbol{W}(k)\|^2 + \|\boldsymbol{W}(k)\|^2 - 1\right)\right]^2 \|\boldsymbol{W}(k)\|^2.
\end{aligned}$$

Thus, we have

$$\|\boldsymbol{W}(k+1)\|^2 \leq [1+\eta(\lambda_1\|\boldsymbol{W}(k)\|^2+1-\|\boldsymbol{W}(k)\|^2)]^2 \cdot \|\boldsymbol{W}(k)\|^2. \tag{6.23}$$

Define a differential function

$$f(s) = [1+\eta(\lambda_1 s+1-s)]^2 s, \tag{6.24}$$

over the interval [0,1]. It follows from (6.9) that

$$\dot{f}(s) = (1+\eta-\eta s(1-\lambda_1))(1+\eta-3\eta s(1-\lambda_1)),$$

for all $0<s<1$. Clearly,

$$\dot{f}(s) = 0, \ \textit{if} \quad s = (1+\eta)/(3\eta(1-\lambda_1)) \quad \textit{or} \quad s = (1+\eta)/\eta(1-\lambda_1) \ .$$

Denote

$$\theta = (1+\eta)/(3\eta(1-\lambda_1)).$$

Then,

$$\dot{f}(s)\begin{cases} >0, & \text{if} \quad 0<s<\theta \\ =0, & \text{if} \quad s=\theta \\ <0, & \text{if} \quad s>\theta. \end{cases} \tag{6.25}$$

By $\eta\lambda_1 < 0.125$ and $\eta < 0.25$, clearly,

$$\theta = (1+\eta)/(3\eta(1-\lambda_1)) = (1/\eta+1)/(3(1-\lambda_1)) > 1. \tag{6.26}$$

From (6.25) and (6.26), it holds that

$$\dot{f}(s) > 0,$$

for all $0<s<1$. This means that $f(s)$ is monotonically increasing over the interval [0,1]. Then, we have

$$f(s) \leq f(1) < (1+\eta\lambda_1)^2,$$

for all $0<s<1$.

Thus, $\|\boldsymbol{W}(k)\| < (1+\eta\lambda_1)$, for all $k \geq 0$.

This completes the proof.

Theorem 6.3 *Suppose that* $\eta\lambda_1 < 0.125$ *and* $\eta < 0.25$. *If* $\boldsymbol{W}^{\mathrm{T}}(0)\boldsymbol{V}_n \neq 0$ *and* $\|\boldsymbol{W}(0)\| \leq 1$, *then it holds that* $\|\boldsymbol{W}(k)\| > c$ *for all* $k \geq 0$, *where* $c = \min\{[1-\eta\lambda_1]\|\boldsymbol{W}(0)\|, [1-\eta\lambda_1(1+\eta\lambda_1)^4+\eta(1-(1+\eta\lambda_1)^2)]\}$.

Proof From Theorem 6.2, we have $\|\boldsymbol{W}(k)\| < (1+\eta\lambda_1)$ for all $k \geq 0$ under the conditions of Theorem 6.3. Next, two cases will be considered.

Case 1: $0 < \|\boldsymbol{W}(k)\| \leq 1$.

From (6.19) and (6.20), it follows that

$$\begin{aligned}\|\boldsymbol{W}(k+1)\|^2 &= \sum_{i=1}^{n}[1-\eta\lambda_i\|\boldsymbol{W}(k)\|^{2+\alpha}+\eta(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)+1-\|\boldsymbol{W}(k)\|^2]^2 z_i^2(k)\\ &\geq \left[1-\eta\left(\lambda_1\|\boldsymbol{W}(k)\|^{2+\alpha}-\lambda_n\|\boldsymbol{W}(k)\|^2\right)+\eta\left(1-\|\boldsymbol{W}(k)\|^2\right)\right]\sum_{i=1}^{n} z_i^2(k)\\ &\geq \left[1-\eta\left(\lambda_1\|\boldsymbol{W}(k)\|^{2+\alpha}-\lambda_n\|\boldsymbol{W}(k)\|^2\right)\right]^2\|\boldsymbol{W}(k)\|^2\\ &\geq \left[1-\eta\lambda_1\|\boldsymbol{W}(k)\|^{2+\alpha}\right]^2\|\boldsymbol{W}(k)\|^2\\ &\geq [1-\eta\lambda_1]^2\|\boldsymbol{W}(k)\|^2.\end{aligned}$$

Case 2: $1 < \|\boldsymbol{W}(k)\| < (1+\eta\lambda_1)$.

From (6.19) and (6.20), it follows that

$$\begin{aligned}\|\boldsymbol{W}(k+1)\|^2 &\geq \left[1-\eta\left(\lambda_1\|\boldsymbol{W}(k)\|^{2+\alpha}-\lambda_n\|\boldsymbol{W}(k)\|^2\right)+\eta\left(1-\|\boldsymbol{W}(k)\|^2\right)\right]^2\sum_{i=1}^{n} z_i^2(k)\\ &\geq \left[1-\eta\lambda_1\|\boldsymbol{W}(k)\|^{2+\alpha}+\eta\left(1-\|\boldsymbol{W}(k)\|^2\right)\right]^2\|\boldsymbol{W}(k)\|^2\\ &\geq \left[1-\eta\lambda_1\|\boldsymbol{W}(k)\|^{2+\alpha}+\eta\left(1-\|\boldsymbol{W}(k)\|^2\right)\right]^2\\ &\geq \left[1-\eta\lambda_1(1+\eta\lambda_1)^4+\eta\left(1-(1+\eta\lambda_1)^2\right)\right]^2.\end{aligned}$$

Using the analysis of Cases 1 and 2, clearly,

$$\|\boldsymbol{W}(k)\| > c = \min\left\{[1-\eta\lambda_1]\|\boldsymbol{W}(0)\|, \left[1-\eta\lambda_1(1+\eta\lambda_1)^4+\eta\left(1-(1+\eta\lambda_1)^2\right)\right]\right\},$$

for all $k \geq 0$. From the conditions of Theorem 6.2, clearly, $c > 0$.

This completes the proof.

At this point, the boundness of DDT system (6.19) has been proven. Next, we will prove that under some mild conditions, $\lim_{k\to+\infty}\boldsymbol{W}(k) = \pm\boldsymbol{V}_n$, where $\boldsymbol{V}_n$ is the minor component. In order to analyze the convergence of DDT (6.19), we need to prove the following lemma first.

Lemma 6.1 *Suppose that* $\eta\lambda_1 < 0.125$ *and* $\eta < 0.25$. *If* $\boldsymbol{W}^{\mathrm{T}}(0)\boldsymbol{V}_n \neq 0$ *and* $\|\boldsymbol{W}(0)\| \leq 1$, *then it holds that*

$$1 - \eta\lambda_i \|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\left(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2\right) > 0.$$

Proof By Theorem 6.2, under the conditions of Lemma 6.1, it holds that $\|\boldsymbol{W}(k)\| < 1 + \eta\lambda_1$, for all $k \geq 0$. Next two cases will be considered.

Case 1: $0 < \|\boldsymbol{W}(k)\| \leq 1$.

From (6.21) and (6.22), for each $i(1 \leq i \leq n)$, we have

$$\begin{aligned}
&1 - \eta\lambda_i \|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\left(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2\right)\\
&> 1 - \eta\lambda_1 \|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\lambda_n \|\boldsymbol{W}(k)\|^2\\
&> 1 - \eta\lambda_1 \|\boldsymbol{W}(k)\|^{2+\alpha}\\
&> 1 - \eta\lambda_1\\
&> 0,
\end{aligned}$$

for $k \geq 0$.

Case 2: $1 < \|\boldsymbol{W}(k)\| < 1 + \eta\lambda_1$.

From (6.21) and (6.22), for each $i(1 \leq i \leq n)$, we have

$$\begin{aligned}
&1 - \eta\lambda_i \|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\left(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2\right)\\
&> 1 - \eta\lambda_1 \|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\lambda_n \|\boldsymbol{W}(k)\|^2 - \eta\left(2\eta\lambda_1 + \eta^2\lambda_1^2\right)\\
&> 1 - \eta\lambda_1 \|\boldsymbol{W}(k)\|^{2+\alpha} - 0.25 * \left(2\eta\lambda_1 + \eta^2\lambda_1^2\right)\\
&> 1 - \eta\lambda_1 \|\boldsymbol{W}(k)\|^{2+\alpha} - \left(0.5 * \eta\lambda_1 + 0.25 * (\eta\lambda_1)^2\right)\\
&> 1 - \eta\lambda_1\left((1 + \eta\lambda_1)^4 + 0.5 + 0.25 * \eta\lambda_1\right)\\
&> 0.
\end{aligned}$$

This completes the proof.

Lemma 6.1 means that the projection of the weight vector $\boldsymbol{W}(k)$ on eigenvector $\boldsymbol{V}_i (i = 1, 2, \ldots, n)$, which is denoted as $z_i(k) = \boldsymbol{W}^T(k)\boldsymbol{V}_i (i = 1, 2, \ldots, n)$, does not change its sign in (6.21). From (6.20), we have $z_i(t) = \boldsymbol{W}^{\mathrm{T}}(t)\boldsymbol{V}_i$. Since $\boldsymbol{W}^{\mathrm{T}}(0)\boldsymbol{V}_n \neq 0$, we have $z_n(0) \neq 0$. It follows from (6.6) and Lemma 6.1 that $z_n(k) > 0$ for all $k > 0$ if $z_n(0) > 0$; and $z_n(k) < 0$ for all $k > 0$ if $z_n(0) < 0$. Without loss of generality, we assume that $z_n(0) > 0$. Thus, $z_n(k) > 0$ for all $k > 0$.

From (6.20), for each $k \geq 0$, $\boldsymbol{W}(k)$ can be represented as

$$\boldsymbol{W}(k) = \sum_{i=1}^{n-1} z_i(k)\boldsymbol{V}_i + z_n(k)\boldsymbol{V}_n. \tag{6.27}$$

Clearly, the convergence of $\boldsymbol{W}(k)$ can be determined by the convergence of $z_i(k)$ $(i = 1,2,\ldots,n)$. Theorems 6.4 and 6.5 below provide the convergence of $z_i(k)$ $(i = 1,2,\ldots,n)$.

Theorem 6.4 *Suppose that* $\eta\lambda_1 < 0.125$ *and* $\eta < 0.25$. *If* $\boldsymbol{W}^{\mathrm{T}}(0)\boldsymbol{V}_n \neq 0$ *and* $\|\boldsymbol{W}(0)\| < 1$, *then* $\lim_{k\to\infty} z_i(k) = 0$, $(i = 1,2,\ldots,n-1)$.

Proof By Lemma 6.1, clearly,

$$1 - \eta\lambda_i\|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\left(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2\right) > 0, \quad (i = 1,2,\ldots,n) \tag{6.28}$$

for all $k \geq 0$. Using Theorems 6.2 and 6.3, it holds that $\|\boldsymbol{W}(k)\| > c$ and $\|\boldsymbol{W}(k)\| < (1+\eta\lambda_1)$ for all $k \geq 0$. Thus, it follows that for all $k \geq 0$

$$\begin{aligned}
&\left[\frac{1 - \eta\lambda_i\|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\left(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2\right)}{1 - \eta\lambda_n\|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\left(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2\right)}\right]^2 \\
&\overset{(1)}{=} \left[1 - \frac{\eta(\lambda_i - \lambda_n)\|\boldsymbol{W}(k)\|^{2+\alpha}}{1 - \eta\lambda_n\|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\left(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2\right)}\right]^2 \\
&\leq \left[1 - \frac{\eta(\lambda_i - \lambda_n)\|\boldsymbol{W}(k)\|^{2+\alpha}}{1 - \eta\lambda_n\|\boldsymbol{W}(k)\|^{2+\alpha} + \eta\left(\lambda_1\|\boldsymbol{W}(k)\|^2 + 1 - \|\boldsymbol{W}(k)\|^2\right)}\right]^2 \\
&= \left[1 - \frac{\eta(\lambda_i - \lambda_n)}{1/\|\boldsymbol{W}(k)\|^{2+\alpha} - \eta\lambda_n + \eta\left(\lambda_1\|\boldsymbol{W}(k)\|^{-\alpha} + 1/\|\boldsymbol{W}(k)\|^{2+\alpha} - \|\boldsymbol{W}(k)\|^{-\alpha}\right)}\right]^2 \\
&< \left[1 - \frac{\eta(\lambda_{n-1} - \lambda_n)}{1/c^{(2+\alpha)} - \eta\lambda_n + \eta[\lambda_1 c^{-\alpha} + 1/c^{(2+\alpha)} - (1+\eta\lambda_1)^{-\alpha}]}\right]^2, (i = 1,2,\ldots,n-1).
\end{aligned} \tag{6.29}$$

Denote

$$\theta = \left[1 - \frac{\eta(\lambda_{n-1} - \lambda_n)}{1/c^{(2+\alpha)} - \eta\lambda_n + \eta[\lambda_1 c^{-\alpha} + 1/c^{(2+\alpha)} - (1+\eta\lambda_1)^{-\alpha}]}\right]^2.$$

Clearly, θ is a constant and $0 < \theta < 1$. By $\boldsymbol{W}^{\mathrm{T}}(0)\boldsymbol{V}_n \neq 0$, clearly, $z_n(0) \neq 0$. Then, $z_n(k) \neq 0 (k > 0)$.

From (6.21), (6.28), and (6.29), it holds that

$$\left[\frac{z_i(k+1)}{z_n(k+1)}\right]^2 = \left[\frac{1-\eta\lambda_i\|\boldsymbol{W}(k)\|^{2+\alpha}+\eta\left(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)+1-\|\boldsymbol{W}(k)\|^2\right)}{1-\eta\lambda_n\|\boldsymbol{W}(k)\|^{2+\alpha}+\eta\left(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)+1-\|\boldsymbol{W}(k)\|^2\right)}\right]^2 \cdot \left[\frac{z_i(k)}{z_n(k)}\right]^2$$
$$\leq \theta \cdot \left[\frac{z_i(k)}{z_n(k)}\right]^2 \leq \theta^{k+1} \cdot \left[\frac{z_i(0)}{z_n(0)}\right]^2, (i=1,2,\ldots,n-1), \tag{6.30}$$

for all $k \geq 0$.

Thus, from $0 < \theta < 1$ $(i=1,2,\ldots,n-1)$, we have

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

By Theorems 6.2 and 6.3, $z_n(k)$ must be bounded. Then,

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

This completes the proof.

Theorem 6.5 *Suppose that* $\eta\lambda_1 < 0.125$ *and* $\eta < 0.25$. *If* $\boldsymbol{W}^{\mathrm{T}}(0)\boldsymbol{V}_n \neq 0$ *and* $\|\boldsymbol{W}(0)\| < 1$, *then it holds that* $\lim_{k\to\infty} z_n(k) = \pm 1$.

Proof Using Theorem 6.4, clearly, $\boldsymbol{W}(k)$ will converge to the direction of the minor component $\boldsymbol{V}_{\mathrm{n}}$, as $k \to \infty$. Suppose at time k_0, $\boldsymbol{W}(k)$ has converged to the direction of $\boldsymbol{V}_{\mathrm{n}}$, i.e., $\boldsymbol{W}(k_0) = z_n(k_0)\cdot\boldsymbol{V}_n$.

From (6.21), it holds that

$$\begin{aligned} z_n(k+1) &= z_n(k)\left(1-\eta\lambda_n z_n^{(2+\alpha)}(k)+\eta(\lambda_n z_n^2(k)+1-z_n^2(k))\right) \\ &= z_n(k)\left(1+\eta[\lambda_n z_n^2(k)(1-z_n^{(\alpha)}(k))+1-z_n^2(k)]\right) \\ &= z_n(k)\left(1+\eta(1-z_n(k))(\lambda_n z_n^2(k)(z_n^{(\alpha-1)}(k)+z_n^{(\alpha-2)}(k)+\cdots+1)+(1+z_n(k)))\right) \\ &= z_n(k)\left(1+\eta(1-z_n(k))(\lambda_n(z_n^{(\alpha+1)}(k)+z_n^{(\alpha)}(k)+\cdots+z_n^2(k))+(1+z_n(k)))\right) \\ &= z_n(k)(1+\eta(1-z_n(k))Q(\lambda_n, z_n(k))), \end{aligned} \tag{6.31}$$

where $Q(\lambda_n, z_n(k)) = (\lambda_n(z_n^{(\alpha+1)}(k)+z_n^{(\alpha)}(k)+\cdots+z_n^2(k))+(1+z_n(k)))$ is a positive efficient, for all $k \geq k_0$.

From (6.31), it holds that

$$\begin{aligned} z_n(k+1)-1 &= z_n(k)(1+\eta(1-z_n(k))Q(\lambda_n, z_n(k)))-1 \\ &= [1-\eta z_n(k)Q(\lambda_n, z_n(k))](z_n(k)-1), \end{aligned} \tag{6.32}$$

for $k > k_0$.

Since $z_n(k) < \|\boldsymbol{W}(\mathrm{k})\| \leq (1+\eta\lambda_1)$, we have

$$\begin{aligned}
&1-\eta z_n(k)Q(\lambda_n, z_n(k))\\
&\quad = 1-\eta z_n(k)(\lambda_n(z_n^{(\alpha+1)}(k)+z_n^{(\alpha)}(k)+\cdots+z_n^2(k))+(1+z_n(k)))\\
&\quad > 1-\eta(1+\eta\lambda_1)(\lambda_n((1+\eta\lambda_1)^{(\alpha+1)}+(1+\eta\lambda_1)^{(\alpha)}+\cdots+(1+\eta\lambda_1)^{(2)})+(1+(1+\eta\lambda_1)))\\
&\quad > 1-(1+\eta\lambda_1)(\eta\lambda_1((1+\eta\lambda_1)^{(\alpha+1)}+(1+\eta\lambda_1)^{(\alpha)}+\cdots+(1+\eta\lambda_1)^{(2)})+\eta(1+(1+\eta\lambda_1)))\\
&\quad > 1-(1+\eta\lambda_1)(\eta\lambda_1((1+\eta\lambda_1)^3+(1+\eta\lambda_1)^2)+\eta(1+(1+\eta\lambda_1)))\\
&\quad > 1-0.9980\\
&\quad > 0,
\end{aligned} \tag{6.33}$$

for all $k \geq k_0$. Thus, denote $\delta = 1-\eta z_n(k)Q(\lambda_n, z_n(k))$, Clearly, it holds that $0<\delta<1$.

It follows from (6.32) and (6.33) that

$$|z_n(k+1)-1| \leq \delta|z_n(k)-1|,$$

for all $k > k_0$. Then, for $k > k_0$

$$|z_n(k+1)-1| \leq \delta^{k+1}|z_n(0)-1| \leq (k+1)\Pi \mathrm{e}^{-\theta(k+1)},$$

where $\theta = -\ln\delta$, $\Pi = |(1+\eta\lambda_1)-1|$.

Given any $\varepsilon > 0$, there exists a $K \geq 1$ such that

$$\frac{\Pi_2 K \mathrm{e}^{-\theta K}}{(1-\mathrm{e}^{-\theta})^2} \leq \varepsilon.$$

For any $k_1 > k_2 > k$, it follows from (6.21) that

$$\begin{aligned}
|z_n(k_1)-z_n(k_2)| &= \left|\sum_{r=k_2}^{k_1-1}[z_n(r+1)-z_n(r)]\right| \leq \left|\sum_{r=k_2}^{k_1-1}\eta z_n(r)((1-z_n(r))Q(\lambda_n, z_n(r)))\right|\\
&\leq \sum_{r=k_2}^{k_1-1}|\eta z_n(r)((1-z_n(r))Q(\lambda_n, z_n(r)))| \leq \sum_{r=k_2}^{k_1-1}|\eta z_n(r)Q(\lambda_n, 1+\eta\lambda_1)(z_n(r)-1)|\\
&\leq \eta(1+\eta\lambda_1)Q(\lambda_n, 1+\eta\lambda_1)\sum_{r=k_2}^{k_1-1}|(z_n(r)-1)| \leq \Pi_2\sum_{r=k_2}^{k_1-1} r\mathrm{e}^{-\theta r}\\
&\leq \Pi_2\sum_{r=k}^{+\infty} r\mathrm{e}^{-\theta r} \leq \Pi_2 K\mathrm{e}^{-\theta K}\sum_{r=0}^{+\infty} r(\mathrm{e}^{-\theta})^{r-1} \leq \frac{\Pi_2 K\mathrm{e}^{-\theta K}}{(1-\mathrm{e}^{-\theta})^2}\\
&\leq \varepsilon.
\end{aligned}$$

where $\Pi_2 = \eta(1+\eta\lambda_1)Q(\lambda_n, 1+\eta\lambda_1)(z_n(0)-1)$. This means that the sequence $\{z_n(k)\}$ is a Cauchy sequence. By the Cauchy convergence principle, there must exist a constant z^* such that $\lim_{x\to\infty} z_n(k) = z^*$.

From (6.27), we have $\lim_{k\to+\infty} \boldsymbol{W}(k) = z_n^* \cdot \boldsymbol{V}_n$. Since (6.17) has self-stabilizing property, it follows that $\lim_{x\to\infty} \boldsymbol{W}(k+1)/\boldsymbol{W}(k) = 1$. From (6.21), we have $1 = 1 - \eta[\lambda_n(z_n^*)^{2+\alpha} - (\lambda_n(z_n^*)^2 + 1 - (z_n^*)^2)]$, which means $z_n^* = \pm 1$.

This completes the proof.

Using (6.27), along with Theorems 6.4 and 6.5, we can draw the following conclusion:

Theorem 6.6 *Suppose that* $\eta\lambda_1 < 0.125$ *and* $\eta < 0.25$. *If* $\boldsymbol{W}^{\mathrm{T}}(0)\boldsymbol{V}_n \neq 0$ *and* $\|\boldsymbol{W}(0)\| < 1$, *then it holds that* $\lim_{k\to\infty} \boldsymbol{W}(k) = \pm\boldsymbol{V}_n$.

At this point, we have completed the proof of the convergence of DDT system (6.19). Next we will further study the stability of (6.19).

Theorem 6.7 *Suppose that* $\eta\lambda_1 < 0.125$ *and* $\eta < 0.25$. *Then the equilibrium points* $\boldsymbol{V}_n$ *and* $-\boldsymbol{V}_n$ *are locally asymptotical stable and other equilibrium points (6.19) are unstable.*

Proof Clearly, the set of all equilibrium points of (6.21) is $\{\boldsymbol{V}_1, \cdots, \boldsymbol{V}_n\} \cup \{-\boldsymbol{V}_1, \cdots, -\boldsymbol{V}_n\} \cup \{0\}$.

Denote

$$\begin{aligned}\boldsymbol{G}(\boldsymbol{W}) &= \boldsymbol{W}(k+1) \\ &= \boldsymbol{W}(k) - \eta\left[\|\boldsymbol{W}(k)\|^{2+\alpha}\boldsymbol{R}\boldsymbol{W}(k) - (\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) + 1 - \|\boldsymbol{W}(k)\|^2)\boldsymbol{W}(k)\right].\end{aligned} \tag{6.34}$$

Then, we have

$$\begin{aligned}\frac{\partial \mathbf{G}}{\partial \mathbf{W}} = \mathbf{I} + \eta[(\mathbf{W}^{\mathrm{T}}(k)\mathbf{R}\mathbf{W}(k) + 1 - \|\mathbf{W}(k)\|^2)\mathbf{I} - \|\mathbf{W}(k)\|^{2+\alpha}\mathbf{R} + 2\mathbf{R}\mathbf{W}(k)\mathbf{W}^{\mathrm{T}}(k) \\ - 2\mathbf{W}(k)\mathbf{W}^{\mathrm{T}}(k) - (2+\alpha)\|\mathbf{W}(k)\|^{\alpha}\mathbf{R}\mathbf{W}(k)\mathbf{W}^{\mathrm{T}}(k)],\end{aligned} \tag{6.35}$$

where $\boldsymbol{I}$ is a unity matrix.

For the equilibrium point 0, it holds that

$$\left.\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{W}}\right|_0 = \boldsymbol{I} + \eta\boldsymbol{I} = \boldsymbol{J}_0.$$

The eigenvalues of $\boldsymbol{J}_0$ are $\alpha_0^{(i)} = 1 + \eta > 1 \quad (i = 1, 2, \cdots, n)$. Thus, the equilibrium point is unstable.

For the equilibrium points $\pm \boldsymbol{V}_j (j = 1, 2, \cdots, n)$, it follows from (6.35) that

$$\left.\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{W}}\right|_{\boldsymbol{V}_j} = \boldsymbol{I} + \eta \left[\lambda_j \boldsymbol{I} - \boldsymbol{R} - 2\boldsymbol{V}_j \boldsymbol{V}_j^{\mathrm{T}} - \alpha\lambda_j \boldsymbol{V}_j \boldsymbol{V}_j^{\mathrm{T}} \right] = \boldsymbol{J}_j. \tag{6.36}$$

After some simple manipulations, the eigenvalues of $\boldsymbol{J}_j$ are given by

$$\begin{cases} \alpha_j^{(i)} = 1 + \eta(\lambda_j - \lambda_i) & \text{if} \quad i \neq j. \\ \alpha_j^{(i)} = 1 - \eta(2 + \alpha\lambda_j) & \text{if} \quad i = j. \end{cases}$$

For any $j \neq n$, it holds that $\alpha_j^{(n)} = 1 + \eta(\lambda_j - \lambda_i) > 1$. Clearly, the equilibrium points $\pm \boldsymbol{V}_j (j \neq n)$ are unstable. For the equilibrium points $\pm \boldsymbol{V}_n$, from $\eta\lambda_n < \eta\lambda_1 < 0.125$, and $\eta < 0.25$, it holds that

$$\begin{cases} \alpha_n^{(i)} = 1 + \eta(\lambda_n - \lambda_i) < 1 & \text{if} \quad i \neq n. \\ \alpha_n^{(i)} = 1 - \eta(2 + \alpha\lambda_n) < 1 & \text{if} \quad i = n. \end{cases} \tag{6.37}$$

Thus, $\pm \boldsymbol{V}_n$ are asymptotical stable.

This completes the proof.

From (6.37), we can easily see that the only fixed points where the MCA condition is fulfilled are the attractors, and all others are repellers or saddle points. We conclude that the Algorithm (6.17) converges toward the minor eigenvector $\pm \boldsymbol{V}_n$ associated with the minor eigenvalue λ_n.

6.3.3 Computer Simulations

In this section, we provide simulation results to illustrate the convergence and stability of the MCA Algorithm (6.17) in a stochastic case. Since OJAm [17], Moller [23], and Peng [11] are self-stabilizing algorithms and have better convergence performance than some existing MCA algorithms, we compare performance of Algorithm (6.17) with these algorithms. In order to measure the convergence speed and accuracy of these algorithms, we compute the norm of $\boldsymbol{W}(k)$ and the direction cosine at the kth update. In the simulation, the input data sequence, which is generated by [17], $\boldsymbol{X}(k) = \boldsymbol{C}\ \boldsymbol{h}(\mathrm{k})$, where $\mathbf{C}$ = randn(5, 5)/5 and $\boldsymbol{h}(k) \in \boldsymbol{R}^{5\times 1}$, is Gaussian and randomly generated with zero-mean and unitary standard deviation. The above-mentioned four MCA algorithms are used to extract minor component from the input data sequence $\{x(k)\}$. The following learning curves show the convergence of $\boldsymbol{W}(k)$ and direction cosine(k) with the same initial norm for the weight vector and constant learning rate, respectively. All the learning curves below are obtained by averaging over 30 independent experiments. Figures 6.2 and 6.3 investigate the case $\|\boldsymbol{W}(0)\| = 1$, and Figs. 6.4 and 6.5 show the simulation results for higher-dimensional data (D = 12), using different learning rates and maximal eigenvalues, which satisfy the conditions of Theorem 6.6.

From Fig. 6.3, we can see that for all these MCA algorithms, direction cosine (k) converge to 1 at approximately the same speeds. However, from Fig. 6.2 we can see that the Moller and OJAm algorithms have approximately the same convergence for the weight vector length and there appear to be a residual deviation from unity for the weight vector length, and the norm of the weight vector in Peng algorithm has larger oscillations, and the norm of the weight vector in Algorithm (6.17) has a faster convergence, a better numerical stability and higher precision than other algorithms. From Figs. 6.4 and 6.5, it is obvious that even for higher-dimensional data, only if the conditions of Theorems 6.2–6.6 are satisfied, Algorithm (6.17) can satisfactorily extract the minor component of the input data stream.

In this section, dynamics of a class of algorithms are analyzed by the DDT method. It has been proved that if some mild conditions about the learning rate and the initial weight vector are satisfied, these algorithms will converge to the minor

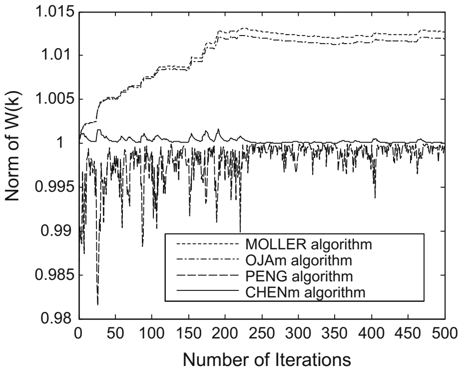

Fig. 6.2 Norm of $\boldsymbol{W}(k)$ ($\eta = 0.1$, $\|\boldsymbol{W}(0)\| = 1.0$, $\alpha = 2$)

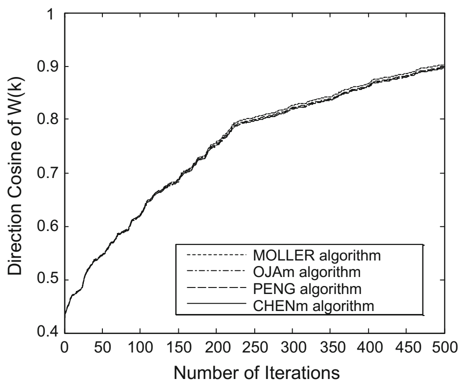

Fig. 6.3 Direction cosine of $\boldsymbol{W}(k)$ ($\eta = 0.1$, $\|\boldsymbol{W}(0)\| = 1.0$, $\alpha = 2$)

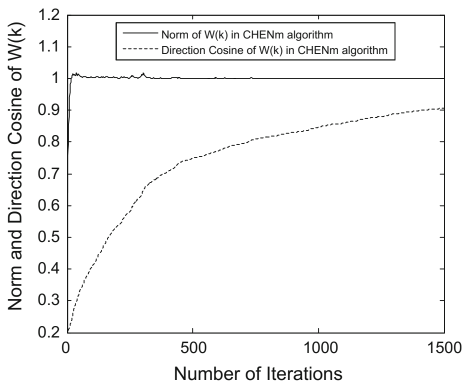

Fig. 6.4 $\eta = 0.1, \|\boldsymbol{W}(0)\| = 0.6$, $\lambda_1 = 1.1077$, $\alpha = 2$, and D = 12

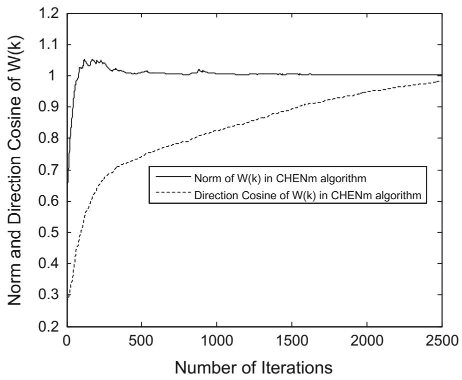

Fig. 6.5 $\eta = 0.01$, $\|\boldsymbol{W}(0)\| = 0.6$, $\lambda_1 = 12.1227$, $\alpha = 2$, and D = 12

component with unit norm. At the same time, stability analysis shows that the minor component is the asymptotical stable equilibrium point in these algorithms. Simulation results show that this class of self-stabilizing MCA algorithms outperforms some existing MCA algorithms.

6.4 DDT System of a Unified PCA and MCA Algorithm

In Sect. 6.3, the convergence of a MCA algorithm proposed by us is analyzed via DDT in details. However, in the above analysis, we made one assumption, i.e., the smallest eigenvalue of the correlation matrix of the input data is single. In this section, we will remove this assumption in the convergence analysis and analyze a unified PCA and MCA algorithm via the DDT method.

6.4.1 Introduction

Despite the large number of unified PCA and MCA algorithms proposed to date, there are few works that analyze these algorithms via the DDT method and derive the conditions to guarantee the convergence. Obviously, this is necessary from the point view of application. Among the unified PCA and MCA algorithms, Chen's algorithm [22] is regarded as a pioneering work. Other self-normalizing dual systems [24] or dual-purpose algorithms [19, 20] can be viewed as the generalizations of Chen's algorithm [22]. Chen's algorithm lays sound theoretical foundations for dual-purpose algorithms. However, no work has been done so far on the study of Chen's DDT system. In this section, the unified PCA and MCA algorithm proposed by Chen et al. [22] will be analyzed and some sufficient conditions to guarantee its convergence will be derived by the DDT method. These theoretical results will lay a solid foundation for the applications of this algorithm.

6.4.2 A Unified Self-stabilizing Algorithm for PCA and MCA

Chen et al. proposed a unified stabilizing learning algorithm for principal components and minor components extraction [22], and the stochastic discrete form of the algorithm can be written as

$$\boldsymbol{W}(k+1) = \boldsymbol{W}(k) \pm \eta\left[\|\boldsymbol{W}(k)\|^2 y(k)\boldsymbol{X}(k) - y^2(k)\boldsymbol{W}(k)\right] + \eta(1 - \|\boldsymbol{W}(k)\|^2)\boldsymbol{W}(k), \tag{6.38}$$

where η $(0<\eta<1)$ is the learning rate. Algorithm (6.38) can extract principal component if "+" is used. If the sign is simply altered, (6.38) can also serve as a minor component extractor. It is interesting that the only difference between the PCA algorithm and the MCA algorithm is the sign on the right hand of (6.38).

In order to derive some sufficient conditions to guarantee the convergence of Algorithm (6.38), next we analyze the dynamics of (6.38) via the DDT approach. The DDT system associated with (6.38) can be formulated as follows. Taking the conditional expectation $E\{\boldsymbol{W}(k+1)/\boldsymbol{W}(0), \boldsymbol{X}(i), i<k\}$ to (6.38) and identifying the conditional expectation as the next iterate, a DDT system can be obtained and given as

$$\begin{aligned}\boldsymbol{W}(k+1) = \boldsymbol{W}(k) &\pm \eta\left[\|\boldsymbol{W}(k)\|^2\boldsymbol{R}\boldsymbol{W}(k) - \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)\boldsymbol{W}(k)\right] \\ &+ \eta(1 - \|\boldsymbol{W}(k)\|^2)\boldsymbol{W}(k),\end{aligned} \tag{6.39}$$

where $\boldsymbol{R} = E[\boldsymbol{X}(k)\boldsymbol{X}^{\mathrm{T}}(k)]$ is the correlation matrix of the input data. The main purpose of this section is to study the convergence of the weight vector $\boldsymbol{W}(k)$ of (6.39) subject to the learning rate η being some constant.

6.4.3 Convergence Analysis

Since $\boldsymbol{R}$ is a symmetric positive definite matrix, there exists an orthonormal basis of $\Re^n$ composed of the eigenvectors of $\boldsymbol{R}$. Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ to be all the eigenvalues of $\boldsymbol{R}$ ordered by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} \geq \lambda_n > 0$. Denote by σ, the largest eigenvalue of $\boldsymbol{R}$. Suppose that the multiplicity of σ is $m(1 \leq m \leq n)$. Then, $\sigma = \lambda_1 = \cdots = \lambda_m$. Suppose that $\{\boldsymbol{V}_i | i = 1, 2, \cdots, n\}$ is an orthogonal basis of $\Re^n$ such that each $\boldsymbol{V}_{\mathrm{i}}$ is a unitary eigenvector of $\boldsymbol{R}$ associated with the eigenvalue λ_i. Denote by $\boldsymbol{V}_\sigma$ the eigen-subspace of the largest eigenvalue σ, i.e., $\boldsymbol{V}_\sigma = \mathrm{span}\{\boldsymbol{V}_1, \cdots, \boldsymbol{V}_m\}$. Denote by $\boldsymbol{V}_\sigma^\perp$ the subspace which is perpendicular to $\boldsymbol{V}_\sigma$. Clearly, $\boldsymbol{V}_\sigma^\perp = \mathrm{span}\{\boldsymbol{V}_{m+1}, \cdots, \boldsymbol{V}_n\}$. Similarly, we can denote by $\boldsymbol{V}_\tau$ the eigen-subspace of the smallest eigenvalue τ. Suppose that the multiplicity of τ is $p(1 \leq p \leq n - m)$. Then, $\boldsymbol{V}_\tau = \mathrm{span}\{\boldsymbol{V}_{n-p}, \cdots, \boldsymbol{V}_n\}$ and $\boldsymbol{V}_\tau^\perp = \mathrm{span}\{\boldsymbol{V}_1, \cdots, \boldsymbol{V}_{n-p-1}\}$.

Since the vector set $\{\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_n\}$ is an orthonormal basis of $\Re^n$, for each $k \geq 0$, $\boldsymbol{W}(k)$ and $\boldsymbol{R}\boldsymbol{W}(k)$ can be represented, respectively, as

$$\boldsymbol{W}(k) = \sum_{i=1}^{n} z_i(k)\boldsymbol{V}_i, \quad \boldsymbol{R}\boldsymbol{W}(k) = \sum_{i=1}^{n} \lambda_i z_i(k)\boldsymbol{V}_i, \tag{6.40}$$

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

From (6.39) and (6.40), it holds that

$$z_i(k+1) = [1 \pm \eta(\lambda_i \|\boldsymbol{W}(k)\|^2 - \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)) + \eta(1 - \|\boldsymbol{W}(k)\|^2)]z_i(k), \tag{6.41}$$

$(i = 1, 2, \ldots, n)$, for all $k \geq 0$.

By letting $Q(\boldsymbol{R}, \boldsymbol{W}(k)) = \pm[\lambda_i \|\boldsymbol{W}(k)\|^2 - \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)]$,

(6.41) can be represented as

$$z_i(k+1) = [1 + \eta Q(\boldsymbol{R}, \boldsymbol{W}(k)) + \eta(1 - \|\boldsymbol{W}(k)\|^2)]z_i(k), \tag{6.42}$$

$(i = 1, 2, \ldots, n)$, for all $k \geq 0$. According to the properties of the Rayleigh Quotient [7], it clearly holds that

$$\lambda_n \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{W}(k) \leq \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k) \leq \lambda_1 \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{W}(k), \tag{6.43}$$

for all $k \geq 0$. From (6.43), it holds that

$$Q_{max} = (\lambda_1 - \lambda_n)\|\boldsymbol{W}(k)\|^2, \quad Q_{\min} = (\lambda_n - \lambda_1)\|\boldsymbol{W}(k)\|^2. \tag{6.44}$$

Next, we will analyze the convergence of DDT system (6.39) via the following Theorems 6.8–6.11.

Theorem 6.8 *Suppose that* $\eta \leq 0.3$. *If* $\|\boldsymbol{W}(0)\| \leq 1$ *and* $(\lambda_1 - \lambda_n) < 1$, *then it holds that* $\|\boldsymbol{W}(k)\| < (1 + \eta\lambda_1)$, *for all* $k \geq 0$.

Proof From (6.40)–(6.44), it follows that

$$\begin{aligned}
\|\boldsymbol{W}(k+1)\|^2 &= \sum_{i=1}^{n} z_i^2(k+1) = \sum_{i=1}^{n} [1 + \eta Q(\boldsymbol{R}, \boldsymbol{W}(k)) + \eta(1 - \|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k) \\
&\leq \sum_{i=1}^{n} \left[1 + \eta Q_{\max} + \eta(1 - \|\boldsymbol{W}(k)\|^2)\right]^2 z_i^2(k) \\
&\leq \left[1 + \eta(\lambda_1 - \lambda_n)\|\boldsymbol{W}(k)\|^2 + \eta\left(1 - \|\boldsymbol{W}(k)\|^2\right)\right]^2 \sum_{i=1}^{n} z_i^2(k) \\
&\leq [1 + \eta(\lambda_1 - \lambda_n)\|\boldsymbol{W}(k)\|^2 + \eta(1 - \|\boldsymbol{W}(k)\|^2)]^2 \|\boldsymbol{W}(k)\|^2.
\end{aligned}$$

Thus, it holds that $\|\boldsymbol{W}(k+1)\|^2 \leq [1 + \eta(\lambda_1 - \lambda_n)\|\boldsymbol{W}(k)\|^2 + \eta(1 - \|\boldsymbol{W}(k)\|^2)]^2$ $\|\boldsymbol{W}(k)\|^2$.

Define a differential function $f(s) = [1 + \eta(\lambda_1 - \lambda_n - 1)s + \eta]^2 s$, over the interval [0, 1], where $s = \|\boldsymbol{W}(k)\|^2$ and $f(s) = \|\boldsymbol{W}(k+1)\|^2$. It follows that

$$\dot{f}(s) = (1 + \eta - \eta s(\lambda_n + 1 - \lambda_1))(1 + \eta - 3\eta s(\lambda_n + 1 - \lambda_1)), \tag{6.45}$$

for all $0 < s < 1$. Clearly,

$$\dot{f}(s) = 0, \quad \textit{if} \quad s = \frac{1+\eta}{3\eta(\lambda_n + 1 - \lambda_1)} \quad \textit{or} \quad s = \frac{1+\eta}{\eta(\lambda_n + 1 - \lambda_1)}.$$

Denote $\theta = (1+\eta)/(3\eta(\lambda_n + 1 - \lambda_1))$. Then, we have

$$\dot{f}(s)\begin{cases} > 0, & \text{if } 0 < s < \theta \\ = 0, & \text{if } s = \theta \\ < 0, & \text{if } s > \theta. \end{cases} \tag{6.46}$$

By $\eta \leq 0.3$, clearly,

$$\theta = (1+\eta)/(3\eta(\lambda_n + 1 - \lambda_1)) = (1 + 1/\eta)/(3[1 - (\lambda_1 - \lambda_n)]) > 1. \tag{6.47}$$

From (6.46) and (6.47), it holds that $\dot{f}(s) > 0$ for all $0 < s < 1$. This means that $f(s)$ is monotonically increasing over the interval [0,1]. Then, for all $0 < s < 1$, it follows that

$$f(s) \leq f(1) = [1 + \eta(\lambda_1 - \lambda_n)]^2 < (1 + \eta\lambda_1)^2.$$

Thus, we have $\|\boldsymbol{W}(k)\| < (1 + \eta\lambda_1)$ for all $k \geq 0$.

This completes the proof.

Theorem 6.8 shows that there exists an upper bound for $\|\boldsymbol{W}(k)\|$ in the DDT system (6.39), for all $k \geq 0$.

Theorem 6.9 *Suppose that* $\eta \leq 0.3$. *If* $\|\boldsymbol{W}(0)\| \leq 1$, *then it holds that* $\|\boldsymbol{W}(k)\| > c$ *for all* $k \geq 0$, *where* $c = \min\Big\{[1 - \eta\lambda_1]\|\boldsymbol{W}(0)\|, [1 - \eta\lambda_1(1 + \eta\lambda_1)^2 - \eta(2\eta\lambda_1 + \eta^2\lambda_1^2)]\Big\}$.

Proof From Theorem 6.8, we have $\|\boldsymbol{W}(k)\| < (1 + \eta\lambda_1)$ for all $k \geq 0$ under the conditions of Theorem 6.9. Next, two cases will be considered.

Case 1: $0 < \|\boldsymbol{W}(k)\| \leq 1$.

From (6.40)–(6.44), it follows that

$$\begin{aligned}
\|\boldsymbol{W}(k+1)\|^2 &\geq \sum_{i=1}^{n} \Big[1 + \eta Q_{\min} + \eta(1 - \|\boldsymbol{W}(k)\|^2)\Big]^2 z_i^2(k) \\
&\geq \Big[1 + \eta(\lambda_n - \lambda_1)\|\boldsymbol{W}(k)\|^2 + \eta\Big(1 - \|\boldsymbol{W}(k)\|^2\Big)\Big]^2 \sum_{i=1}^{n} z_i^2(k) \\
&\geq \Big[1 + \eta(\lambda_n - \lambda_1)\|\boldsymbol{W}(k)\|^2\Big]^2 \|\boldsymbol{W}(k)\|^2 \geq \Big[1 - \eta\lambda_1\|\boldsymbol{W}(k)\|^2\Big]^2 \|\boldsymbol{W}(k)\|^2 \\
&\geq [1 - \eta\lambda_1]^2 \|\boldsymbol{W}(k)\|^2.
\end{aligned}$$

Case 2: $1 < \|\boldsymbol{W}(k)\| \leq (1 + \eta\lambda_1)$.

From (6.40)–(6.44), it follows that

$$\begin{aligned}
\|\boldsymbol{W}(k+1)\|^2 &\geq \sum_{i=1}^{n} [1 + \eta Q_{\min} + \eta(1 - \|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k) \\
&= [1 + \eta(\lambda_n - \lambda_1)\|\boldsymbol{W}(k)\|^2 + \eta(1 - \|\boldsymbol{W}(k)\|^2)]^2 \sum_{i=1}^{n} z_i^2(k) \\
&\geq \Big[1 - \eta\lambda_1\|\boldsymbol{W}(k)\|^2 + \eta\big(-2\eta\lambda_1 - \eta^2\lambda_1^2\big)\Big]^2 \|\boldsymbol{W}(k)\|^2 \\
&\geq \Big[1 - \eta\lambda_1(1 + \eta\lambda_1)^2 - \eta\big(2\eta\lambda_1 + \eta^2\lambda_1^2\big)\Big]^2.
\end{aligned}$$

From the above analysis, clearly,

$$\|\boldsymbol{W}(k)\| > c = \min\{[1 - \eta\lambda_1]\|\boldsymbol{W}(0)\|, [1 - \eta\lambda_1(1 + \eta\lambda_1)^2 - \eta(2\eta\lambda_1 + \eta^2\lambda_1^2)]\},$$

for all $k \geq 0$. From the conditions of Theorem 6.2, clearly, it holds that $c > 0$.

This completes the proof.

At this point, the boundness of DDT system (6.39) has been proved. Next, we will prove that under some mild conditions, $\lim_{k\to+\infty} \boldsymbol{W}(k) = \sum_{i=1}^{m} z_i^* \boldsymbol{V}_i \in \boldsymbol{V}$ for PCA and $\lim_{k\to+\infty} \boldsymbol{W}(k) = \sum_{i=n-p}^{n} z_i^* \boldsymbol{V}_i \in \boldsymbol{V}_\tau$ for MCA.

In order to analyze the convergence of DDT (6.39), we need to prove some preliminary results.

From (6.40), for each $k \geq 0$, $\boldsymbol{W}(k)$ can be represented as

$$\begin{cases} \boldsymbol{W}(k) = \sum_{i=1}^{m} z_i(k)\boldsymbol{V}_i + \sum_{j=m+1}^{n} z_j(k)\boldsymbol{V}_j & \textit{for} \quad \text{PCA} \\ \boldsymbol{W}(k) = \sum_{i=1}^{n-p} z_i(k)\boldsymbol{V}_i + \sum_{j=n-p+1}^{n} z_j(k)\boldsymbol{V}_j & \textit{for} \quad \text{MCA.} \end{cases}$$

Clearly, the convergence of $\boldsymbol{W}(k)$ can be determined by the convergence of $z_i(k)$ $(i = 1,2,\ldots,n)$. The following Lemmas 6.2–6.4 provide the convergence of $z_i(k)$ $(i = 1,2,\ldots,n)$ for PCA, and Lemmas 6.5–6.7 provide the convergence of $z_i(k)$ $(i = 1,2,\ldots,n)$ for MCA.

In the following Lemmas 6.2–6.4, we will prove that all $z_i(k)$ $(i = 2,3,\ldots,n)$ will converge to zero under some mild conditions.

Lemma 6.2 Suppose that $\eta \leq 0.3$. If $\boldsymbol{W}(0) \notin \boldsymbol{V}_\sigma^\perp$ and $\|\boldsymbol{W}(0)\| \leq 1$, then for PCA algorithm of (6.39) there exist constants $\theta_1 > 0$ and $\Pi_1 \geq 0$ such that $\sum_{j=m+1}^{n} z_j^2(k) \leq \prod_1 \cdot e^{-\theta_1 k}$ for all $k \geq 0$, where $\theta_1 = -\ln\beta > 0$ and $\beta = [1 - \eta(\sigma - \lambda_{m+1})/(1/c^2 + \eta(\sigma - \tau) + \eta(1/c^2 - 1))]^2$. Clearly, β is a constant and $0 < \beta < 1$.

Proof Since $\boldsymbol{W}(0) \notin \boldsymbol{V}_\sigma^\perp$, there must exist some $i(1 \leq i \leq m)$ such that $z_i(0) \neq 0$. Without loss of generality, assume $z_1(0) \neq 0$. For PCA, it follows from (6.41) that

$$\begin{aligned} z_i(k+1) = {} & [1 + \eta(\sigma\|\mathbf{W}(k)\|^2 - \mathbf{W}^{\mathrm{T}}(k)\mathbf{R}\mathbf{W}(k)) \\ & + \eta(1 - \|\mathbf{W}(k)\|^2)]z_i(k), \quad (1 \leq i \leq m) \end{aligned} \tag{6.48}$$

and

$$\begin{aligned} z_j(k+1) = {} & [1 + \eta(\lambda_j\|\mathbf{W}(k)\|^2 - \mathbf{W}^{\mathrm{T}}(k)\mathbf{R}\mathbf{W}(k)) \\ & + \eta(1 - \|\mathbf{W}(k)\|^2)]z_j(k), \quad m+1 \leq j \leq n \end{aligned} \tag{6.49}$$

for $k \geq 0$.

Using Theorem 6.9, it holds that $\|\boldsymbol{W}(k)\| > c$ for all $k \geq 0$. Then, from (6.48) and (6.49), for each $j(m+1 \leq j \leq n)$, we have

$$\begin{aligned}\left[\frac{z_j(k+1)}{z_1(k+1)}\right]^2 &= \left[\frac{1+\eta(\lambda_j\|\boldsymbol{W}(k)\|^2-(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)}{1+\eta(\sigma\|\boldsymbol{W}(k)\|^2-(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)}\right]^2\cdot\left[\frac{z_j(k)}{z_1(k)}\right]^2\\ &= \left[1-\frac{\eta(\sigma-\lambda_j)\|\boldsymbol{W}(k)\|^2}{1+\eta(\sigma\|\boldsymbol{W}(k)\|^2-(\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)}\right]^2\cdot\left[\frac{z_j(k)}{z_1(k)}\right]^2\\ &= \left[1-\frac{\eta(\sigma-\lambda_j)}{1/\|\boldsymbol{W}(k)\|^2+\eta(\sigma-\tau)+\eta(1/\|\boldsymbol{W}(k)\|^2-1)}\right]^2\cdot\left[\frac{z_j(k)}{z_1(k)}\right]^2\\ &\le \left[1-\frac{\eta(\sigma-\lambda_{m+1})}{1/c^2+\eta(\sigma-\tau)+\eta(1/c^2-1)}\right]^2\cdot\left[\frac{z_j(k)}{z_1(k)}\right]^2\\ &= \beta\frac{z_j^2(k)}{z_1^2(k)}\le\beta^{k+1}\frac{z_j^2(0)}{z_1^2(0)}=\frac{z_j^2(0)}{z_1^2(0)}\mathrm{e}^{-\theta_1(k+1)},\end{aligned} \tag{6.50}$$

for all $k\ge 0$, where $\theta_1=-\ln\beta>0$. Since $\|\boldsymbol{W}(k)\|<(1+\eta\lambda_1)$, $z_1(k)$ must be bounded, i.e., there exists a constant $d>0$ such that $z_1^2(k)\le d$ for all $k\ge 0$. Then,

$$\sum_{j=m+1}^{n} z_j^2(k)=\sum_{j=m+1}^{n}\left[\frac{z_j(k)}{z_1(k)}\right]^2\cdot z_1^2(k)\le\prod_1 \mathrm{e}^{-\theta_1 k},$$

for $k\ge 0$ where $\prod_1 = d\sum\limits_{j=m+1}^{n}\left[\frac{z_j(0)}{z_1(0)}\right]^2\ge 0$.

This completes the proof.

Based on the Lemma, we have Lemma 6.3.

Lemma 6.3 *Suppose that $\eta\lambda_1<0.25$ and $\eta\le 0.3$. Then for PCA algorithm of* (6.39) *there exist constants $\theta_2>0$ and $\prod_2>0$ such that*

$$\begin{aligned}&\left|1-(1-\sigma)\|\boldsymbol{W}(k+1)\|^2-\boldsymbol{W}^{\mathrm{T}}(k+1)\boldsymbol{R}\boldsymbol{W}(k+1)\right|\le(k+1)\\&\quad\cdot\prod_2\cdot[\mathrm{e}^{-\theta_2(k+1)}+\max\{\mathrm{e}^{-\theta_2 k},\mathrm{e}^{-\theta_1 k}\}],\end{aligned}$$

for all $k\ge 0$.

Proof For PCA, it follows from (6.41) that

$$\begin{aligned}\|\boldsymbol{W}(k+1)\|^2 &= \sum_{i=1}^{n}[1+\eta(\lambda_i\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k)\\ &= \sum_{i=1}^{n}[1+\eta(\sigma\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k)\\ &\quad+\sum_{i=m+1}^{n}[1+\eta(\lambda_i\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k)\\ &\quad-\sum_{i=m+1}^{n}[1+\eta(\sigma\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k)\\ &= [1+\eta(\sigma\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2\|\boldsymbol{W}(k)\|^2+H(k),\end{aligned} \tag{6.51}$$

for any $k\geq 0$,
where $H(k)=\sum_{i=m+1}^{n}\Big[(2+\eta(\lambda_i+\sigma)\|\boldsymbol{W}(k)\|^2+2\eta(1-\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)))\;\cdot\eta(\lambda_i-\sigma)\|\boldsymbol{W}(k)\|^2\cdot z_i^2(k)\Big]$.

Clearly,

$$\begin{aligned}\boldsymbol{W}^{\mathrm{T}}(k+1)\boldsymbol{R}\boldsymbol{W}(k+1) &= \sum_{i=1}^{n}\lambda_i[1+\eta(\lambda_i\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k)\\ &= \sum_{i=1}^{n}\lambda_i[1+\eta(\sigma\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k)\\ &\quad+\sum_{i=m+1}^{n}\lambda_i[1+\eta(\lambda_i\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k)\\ &\quad-\sum_{i=m+1}^{n}\lambda_i[1+\eta(\sigma\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k)\\ &= [1+\eta(\sigma\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)+H'(k),\end{aligned} \tag{6.52}$$

for any $k\geq 0$,
where $H'(k)=\sum_{i=m+1}^{n}\Big[(2+\eta(\lambda_i+\sigma)\|\boldsymbol{W}(k)\|^2+2\eta(1-\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)))\;\cdot\eta(\lambda_i-\sigma)\|\boldsymbol{W}(k)\|^2\cdot$ $\lambda_i z_i^2(k)]$. Then, it follows from (6.51) and (6.52) that

$$\begin{aligned}&1-(1-\sigma)\|\boldsymbol{W}(k+1)\|^2-\boldsymbol{W}^{\mathrm{T}}(k+1)\boldsymbol{R}\boldsymbol{W}(k+1)\\ &\quad=(1-(1-\sigma)\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))\{1-[2\eta+\eta^2(1-(1-\sigma)\|\boldsymbol{W}(k)\|^2\\ &\qquad-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))]((1-\sigma)\|\boldsymbol{W}(k)\|^2+\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))\}-(1-\sigma)H(k)-H'(k)\end{aligned}$$

for all $k\geq 0$.

Denote

$$V(k) = \left|1 - (1-\sigma)\|\boldsymbol{W}(k)\|^2 - \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)\right|,$$

for any $k \geq 0$. Clearly,

$$\begin{aligned} V(k+1) \leq V(k)\big|\{1 - [2\eta + \eta^2(1-(1-\sigma)\|\mathbf{W}(k)\|^2 - \mathbf{W}^{\mathrm{T}}(k)\mathbf{R}\mathbf{W}(k))]((1-\sigma)\|\mathbf{W}(k)\|^2 \\ + \mathbf{W}^{\mathrm{T}}(k)\mathbf{R}\mathbf{W}(k))\}\big| + |(1-\sigma)H(k) + H'(k)|. \end{aligned}$$

Denote

$$\delta = \left|\{1 - [2\eta + \eta^2(1-(1-\sigma)\|\boldsymbol{W}(k)\|^2 - \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))]((1-\sigma)\|\boldsymbol{W}(k)\|^2 + \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))\}\right|.$$

From Theorem 6.8, $\eta\lambda_1 < 0.25$, $\eta \leq 0.3$ and (6.43), it holds that

$$\begin{aligned} &[2\eta + \eta^2(1-(1-\sigma)\|\boldsymbol{W}(k)\|^2 - \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))]((1-\sigma)\|\boldsymbol{W}(k)\|^2 + \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)) \\ &\quad < [2\eta + \eta^2(1-(1-\sigma)\|\boldsymbol{W}(k)\|^2 - \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))] \\ &\quad < [2\eta + \eta^2(1-(1-\sigma)\|\boldsymbol{W}(k)\|^2 - \lambda_n\|\boldsymbol{W}(k)\|^2)] \\ &\quad < 2\eta + \eta^2(1+\sigma\|\boldsymbol{W}(k)\|^2) \leq 2\eta + \eta(\eta + \eta\lambda_1(1+\eta\lambda_1)^2) \leq 0.8071, \end{aligned}$$

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

$$V(k+1) \leq \delta V(k) + |(1-\sigma)H(k) + H'(k)|, k \geq 0.$$

Since

$$\begin{aligned} |(1-\sigma)H(k) + H'(k)| &\leq (2 + 2\eta\sigma\|\boldsymbol{W}(k)\|^2 + 2\eta)(\eta\sigma\|\boldsymbol{W}(k)\|^2)\sum_{i=m+1}^{n} z_i^2(k)[(1-\sigma)+\lambda_i] \\ &\leq (2 + 2\eta\sigma\|\boldsymbol{W}(k)\|^2 + 2\eta)(\eta\sigma\|\boldsymbol{W}(k)\|^2)\sum_{i=m+1}^{n} z_i^2(k) \leq \phi\prod_1 \mathrm{e}^{-\theta_1 k}, \end{aligned}$$

for any $k \geq 0$, where $\phi = (2 + 2\eta\sigma(1+\eta\lambda_1)^2 + 2\eta)\cdot\eta\sigma(1+\eta\lambda_1)^2$, then

$$\begin{aligned} V(k+1) &\leq \delta^{k+1}V(0) + \phi\prod\nolimits_1\sum_{r=0}^{k}(\delta\mathrm{e}^{\theta_1})^r\mathrm{e}^{-\theta_1 k} \\ &\leq \delta^{k+1}V(0) + (k+1)\phi\prod\nolimits_1\max\{\delta^k, \mathrm{e}^{-\theta_1 k}\} \\ &\leq (k+1)\prod\nolimits_2\left[\mathrm{e}^{-\theta_2(k+1)} + \max\{\mathrm{e}^{-\theta_2 k}, \mathrm{e}^{-\theta_1 k}\}\right], \end{aligned}$$

where $\theta_2 = -\ln\delta > 0$ and $\prod_2 = \max\{\left|1-(1-\sigma)\|\boldsymbol{W}(0)\|^2 - \boldsymbol{W}^{\mathrm{T}}(0)\boldsymbol{R}\boldsymbol{W}(0)\right|, \phi\prod_1\} > 0$.

This completes the proof.

Based on Lemmas 6.2 and 6.3, we have Lemma 6.4.

Lemma 6.4 *For PCA algorithm of* (6.39), *suppose there exist constants* $\theta > 0$ *and* $\prod > 0$ *such that*

$$\eta\left|(1-(1-\sigma)\|\boldsymbol{W}(k+1)\|^2 - \boldsymbol{W}^{\mathrm{T}}(k+1)\boldsymbol{R}\boldsymbol{W}(k+1))z_i(k+1)\right| \le (k+1)\prod \mathrm{e}^{-\theta(k+1)},$$
$$(i = 1,\ldots,m)$$

for all $k \ge 0$. *Then,* $\lim_{k\to\infty} z_i(k) = z_i^*, (i=1,\ldots,m)$, *where* $z_i^*, (i=1,\ldots,m)$ *are constants.*

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

$$\frac{\prod K\mathrm{e}^{-\theta K}}{(1-\mathrm{e}^{-\theta})^2} \le \varepsilon.$$

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

$$\begin{aligned}
|z_i(k_1) - z_i(k_2)| &= \left|\sum_{r=k_2}^{k_1-1}[z_i(r+1) - z_i(r)]\right| \le \eta\sum_{r=k_2}^{k_1-1}\left|(\sigma\|\boldsymbol{W}(r)\|^2 - \boldsymbol{W}(r)^{\mathrm{T}}\boldsymbol{R}\boldsymbol{W}(r) + 1 - \|\boldsymbol{W}(r)\|^2)z_i(r)\right| \\
&= \eta\sum_{r=k_2}^{k_1-1}\left|(1-(1-\sigma)\|\boldsymbol{W}(r)\|^2 - \boldsymbol{W}(r)^{\mathrm{T}}\boldsymbol{R}\boldsymbol{W}(r))z_i(r)\right| \\
&\le \prod\sum_{r=k_2}^{k_1-1} r\mathrm{e}^{-\theta r} \le \prod\sum_{r=K}^{+\infty} r\mathrm{e}^{-\theta r} \le \prod K\mathrm{e}^{-\theta K}\sum_{r=0}^{+\infty} r(\mathrm{e}^{-\theta})^{r-1} \\
&\le \frac{\prod K\mathrm{e}^{-\theta K}}{(1-\mathrm{e}^{-\theta})^2} \le \varepsilon, (i = 1,\ldots,m).
\end{aligned}$$

This means that the sequence $\{\ z_i(k)\ \}$ is a Cauchy sequence. By the Cauchy convergence principle, there must exist a constant $z^*(i=1,\ldots,m)$ such that $\lim_{k\to+\infty} z_i(k) = z^*, (i = 1,\ldots,m)$.

This completes the proof.

Using the above theorems and lemmas, the convergence of DDT system (6.39) for PCA can be proved as in Theorem 6.10 next.

Theorem 6.10 *Suppose that* $\eta\lambda_1 < 0.25$ *and* $\eta \le 0.3$. *If* $\boldsymbol{W}(0) \notin \boldsymbol{V}_\sigma^\perp$ *and* $\|\boldsymbol{W}(0)\| \le 1$, *then the weight vector of* (6.39) *for PCA will converge to a unitary eigenvector associated with the largest eigenvalue of the correlation matrix.*

Proof By *Lemma* 6.2, there exist constants $\theta_1 > 0$ and $\Pi_1 \ge 0$ such that $\sum_{j=m+1}^{n} z_j^2(k) \le \prod_1 \cdot \mathrm{e}^{-\theta_1 k}$, for all $k \ge 0$. By *Lemma* 6.3, there exist constants $\theta_2 > 0$ and $\prod_2 > 0$ such that

$$\left|(1-(1-\sigma)\|\boldsymbol{W}(k+1)\|^2-\boldsymbol{W}^{\mathrm{T}}(k+1)\boldsymbol{R}\boldsymbol{W}(k+1))\right| \leq (k+1)\cdot\prod\nolimits_2 \cdot[\mathrm{e}^{-\theta_2(k+1)}+\max\{\mathrm{e}^{-\theta_2 k},\mathrm{e}^{-\theta_1 k}\}],$$

for all $k \geq 0$. Obviously, there exist constants $\theta > 0$ and $\prod > 0$ such that

$$\eta\left|(1-(1-\sigma)\|\boldsymbol{W}(k+1)\|^2-\boldsymbol{W}^{\mathrm{T}}(k+1)\boldsymbol{R}\boldsymbol{W}(k+1))z_i(k+1)\right| \leq (k+1)\prod \mathrm{e}^{-\theta(k+1)}, \quad (i=1,\ldots,m)$$

for $k \geq 0$. Using *Lemmas* 6.4 and 6.2, it follows that

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

Then, $\lim\limits_{k\to+\infty}\boldsymbol{W}(k)=\sum_{i=1}^m z_i^*\boldsymbol{V}_i \in \boldsymbol{V}_\sigma$. It can be easily seen that $\lim\limits_{k\to+\infty}\|\boldsymbol{W}(k)\|^2=\sum_{i=1}^m (z_i^*)^2=1$.

This completes the proof.

After proving the convergence of DDT system (6.39) for PCA, we can also prove the convergence of DDT system (6.39) for MCA using similar method. In order to prove the convergence of the weight vector of (6.39) for MCA, we can use the following Lemmas 6.5–6.7 and Theorem 6.11, the proofs of which are similar to those of Lemmas 6.2–6.4 and Theorem 6.10. Here, only these lemmas and theorem will be given and their proofs are omitted.

Lemma 6.5 *Suppose that $\eta \leq 0.3$ .If $\boldsymbol{W}(0) \notin \boldsymbol{V}_\tau^\perp$ and $\|\boldsymbol{W}(0)\| \leq 1$, then for MCA algorithm of* (6.39) *there exist constants $\theta_1' > 0$ and $\Pi_1' \geq 0$ such that $\sum_{j=1}^{n-p} z_j^2(k) \leq \prod_1' \cdot \mathrm{e}^{-\theta_1 k}$, for all $k \geq 0$, where $\theta_1' = -\ln\beta' > 0$, and $\beta' = [1-\eta(\lambda_{n-p-1}-\tau)/(1/c^2-\eta(\tau-\sigma)+\eta(1/c^2-1))]^2$. Clearly, β' is a constant and $0<\beta'<1$.*

Proof For MCA, it follows from (6.41) that

$$\begin{aligned}\|\boldsymbol{W}(k+1)\|^2 &= \sum_{i=1}^n [1-\eta(\lambda_i\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k) \\ &= [1-\eta(\tau\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2\|\boldsymbol{W}(k)\|^2+\bar{H}(k),\end{aligned} \tag{6.53}$$

for any $k \geq 0$ where

$$\begin{aligned}\bar{H}(k) = \sum_{i=1}^{n-p}\Big[&(2-\eta(\lambda_i+\tau)\|\boldsymbol{W}(k)\|^2+2\eta(1 \\ &-\|\boldsymbol{W}(k)\|^2+\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)))\cdot\eta(\tau-\lambda_i)\|\boldsymbol{W}(k)\|^2\cdot z_i^2(k)\Big].\end{aligned}$$

and,

$$\begin{aligned}\boldsymbol{W}^{\mathrm{T}}(k+1)\boldsymbol{R}\boldsymbol{W}(k+1) &= \sum_{i=1}^{n}\lambda_i[1-\eta(\lambda_i\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2 z_i^2(k)\\ &= [1-\eta(\tau\|\boldsymbol{W}(k)\|^2-\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))+\eta(1-\|\boldsymbol{W}(k)\|^2)]^2\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)+\overline{H}'(k),\end{aligned} \tag{6.54}$$

for any $k \geq 0$ where

$$\bar{H}'(k) = \sum_{i=1}^{n-p}\Big[(2-\eta(\lambda_i+\tau)\|\boldsymbol{W}(k)\|^2+2\eta(1-\|\boldsymbol{W}(k)\|^2+\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)))\cdot\eta(\tau-\lambda_i)\|\boldsymbol{W}(k)\|^2\cdot\lambda_i z_i^2(k)\Big].$$

Then, it follows from (6.53) and (6.54) that

$$\begin{aligned}&1-(1+\tau)\|\boldsymbol{W}(k+1)\|^2+\boldsymbol{W}^{\mathrm{T}}(k+1)\boldsymbol{R}\boldsymbol{W}(k+1)\\ &\quad=(1-(1+\tau)\|\boldsymbol{W}(k)\|^2+\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))\{1+[-2\eta+\eta^2(1-(1+\tau)\|\boldsymbol{W}(k)\|^2\\ &\qquad+\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))][\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)-(1+\tau)\|\boldsymbol{W}(k)\|^2]\}-(1+\tau)\bar{H}(k)+\bar{H}'(k)\end{aligned}$$

for all $k \geq 0$.

Denote

$$\bar{V}(k) = \left|1-(1+\tau)\|\boldsymbol{W}(k)\|^2+\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)\right|,$$

for any $k \geq 0$. Clearly,

$$\begin{aligned}\bar{V}(k+1) \leq \bar{V}(k)\big|\{1-[2\eta-\eta^2(1-(1+\tau)\|\mathbf{W}(k)\|^2+\mathbf{W}^T(k)\mathbf{R}\mathbf{W}(k))][\mathbf{W}^T(k)\mathbf{R}\mathbf{W}(k)\\ -(1+\tau)\|\mathbf{W}(k)\|^2]\}\big|+|H'(k)-(1+\tau)H(k)|.\end{aligned}$$

Denote

$$\delta' = \left|\{1-[2\eta-\eta^2(1-(1+\tau)\|\boldsymbol{W}(k)\|^2+\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))][\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k)-(1+\tau)\|\boldsymbol{W}(k)\|^2]\}\right|.$$

From Theorem 6.8, $\eta\lambda_1 < 0.25$, $\eta \leq 0.3$, and (6.43), it holds that

$$\begin{aligned}&[2\eta-\eta^2(1-(1+\tau)\|\boldsymbol{W}(k)\|^2+\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))][\sigma\|\boldsymbol{W}(k)\|^2-(1+\tau)\|\boldsymbol{W}(k)\|^2]\\ &\quad=[2\eta-\eta^2(1-(1+\tau)\|\boldsymbol{W}(k)\|^2+\boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{R}\boldsymbol{W}(k))][(\sigma-(1+\tau))\|\boldsymbol{W}(k)\|^2]\\ &\quad<[2\eta-\eta^2(1-\|\boldsymbol{W}(k)\|^2)][\sigma\|\boldsymbol{W}(k)\|^2]<(\eta\lambda_1)[2-\eta(1-\|\boldsymbol{W}(k)\|^2)][\|\boldsymbol{W}(k)\|^2]\\ &\quad\leq 0.25*[2-0.3+0.3*(1+0.25)^2](1+0.25)^2\\ &\quad=0.8471.\end{aligned}$$

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

$$\bar{V}(k+1) \leq \delta \bar{V}(k) + |H'(k) - (1+\tau)H(k)|, k \geq 0.$$

Since

$$\begin{aligned}
&|H'(k) - (1+\tau)H(k)| \\
&\leq \left| (2 - 2\eta\tau\|\boldsymbol{W}(k)\|^2 + 2\eta(1+\sigma\|\boldsymbol{W}(k)\|^2))(\eta\sigma\|\boldsymbol{W}(k)\|^2)\sum_{i=1}^{n-p} z_i^2(k)[\lambda_i - (1+\tau)] \right| \\
&\leq (2 + 2\eta(1+\sigma\|\boldsymbol{W}(k)\|^2)) \cdot (\eta\sigma\|\boldsymbol{W}(k)\|^2) \cdot |\sigma - (1+\tau)| \cdot \sum_{i=1}^{n-p} z_i^2(k) \\
&\leq \phi' {\prod}_1' \mathrm{e}^{-\theta_1' k},
\end{aligned}$$

for any $k \geq 0$, where $\phi' = (2 + 2\eta(1+\sigma(1+\eta\lambda_1)^2)) \cdot (\eta\sigma(1+\eta\lambda_1)^2) \cdot |\sigma - (1+\tau)|$, we have

$$\begin{aligned}
\bar{V}(k+1) &\leq \delta'^{k+1}\bar{V}(0) + \phi' \prod_1' \sum_{r=0}^{k} (\delta' \mathrm{e}^{\theta_1'})^r \mathrm{e}^{-\theta_1' k} \\
&\leq \delta'^{k+1}\bar{V}(0) + (k+1)\phi' \prod_1' \max\{\delta'^k, \mathrm{e}^{-\theta_1' k}\} \\
&\leq (k+1) \prod_2' \left[\mathrm{e}^{-\theta_2'(k+1)} + \max\{\mathrm{e}^{-\theta_2' k}, \mathrm{e}^{-\theta_1' k}\} \right],
\end{aligned}$$

where $\theta_2' = -\ln \delta' > 0$ and ${\prod}_2' = \max\Big\{ \left|1 - (1+\tau)\|\boldsymbol{W}(0)\|^2 + \boldsymbol{W}^{\mathrm{T}}(0)\boldsymbol{R}\boldsymbol{W}(0)\right|, \phi' {\prod}_1' \Big\} > 0$.

This completes the proof.

Lemma 6.6 *Suppose that* $\eta\lambda_1 < 0.25$ *and* $\eta \leq 0.3$. *Then for MCA algorithm of* (6.39) *there exist constants* $\theta_2' > 0$ *and* ${\prod}_2' > 0$ *such that*

$$\begin{aligned}
&\left|1 - (1+\tau)\|\boldsymbol{W}(k+1)\|^2 + \boldsymbol{W}^{\mathrm{T}}(k+1)\boldsymbol{R}\boldsymbol{W}(k+1)\right| \leq (k+1) \cdot {\prod}_2' \\
&\quad \cdot [\mathrm{e}^{-\theta_2'(k+1)} + \max\{\mathrm{e}^{-\theta_2' k}, \mathrm{e}^{-\theta_1' k}\}],
\end{aligned}$$

for all $k \geq 0$.

For the proof of this lemma, refer to Lemma 6.3.

Lemma 6.7 *For MCA algorithm of* (6.39), *suppose there exists constants* $\theta' > 0$ *and* ${\prod}' > 0$ *such that*

$$\eta\left|(1-(1+\tau)\|\boldsymbol{W}(k+1)\|^2+\boldsymbol{W}^{\mathrm{T}}(k+1)\boldsymbol{R}\boldsymbol{W}(k+1))z_i(k+1)\right| \le (k+1)\prod{}' \mathrm{e}^{-\theta'(k+1)}, \\ (i=n-p+1,\ldots,n)$$

for $k \ge 0$. *Then,* $\lim_{k\to\infty} z_i(k) = z_i^*, (i=n-p+1,\ldots,n)$, *where* $z_i^*, (i=n-p+1,\ldots,n)$ *are constants.*

For the proof of this lemma, refer to Lemma 6.4.

Theorem 6.11 *Suppose that* $\eta\lambda_1 < 0.25$ *and* $\eta \le 0.3$. *If* $\boldsymbol{W}(0) \notin \boldsymbol{V}_\tau^{\perp}$ *and* $\|\boldsymbol{W}(0)\| \le 1$, *then the weight vector of* (6.39) *for MCA will converge to a unitary eigenvector associated with the smallest eigenvalue of the correlation matrix.*

From Lemmas 6.5–6.7, clearly Theorem 6.11 holds.

At this point, we have completed the proof of the convergence of DDT system (6.39). From Theorems 6.8 and 6.9, we can see that the weight norm of PCA algorithm and MCA algorithm of DDT system (6.39) have the same bounds, and from Theorems 6.8–6.11, it is obvious that the sufficient conditions to guarantee the convergence of the two algorithms are also same, which is in favored in practical applications.

6.4.4 Computer Simulations

In this section, we provide simulation results to illustrate the performance of Chen's algorithm. This experiment mainly shows the convergence of Algorithm (6.39) under the condition of Theorems 6.10 and 6.11.

In this simulation, we randomly generate a 12×12 correlation matrix and its eigenvalues are $\lambda_1 = 0.2733$, $\lambda_2 = 0.2116$, $\lambda_3 = 0.1543$, …and $\lambda_{12} = 0.0001$. The initial weight vector is Gaussian and randomly generated with zero-mean and unitary standard deviation, and its norm is less than 1. In the following experiments, the learning rate for PCA is $\eta = 0.05$ and the learning rate for MCA is $\eta = 0.20$, which satisfies the condition of $\eta\lambda_1 \le 0.25$ and $\eta \le 0.3$. Figure 6.6 shows that the convergence of the component $z_i(k)$ of $\boldsymbol{W}(k)$ in (6.39) for PCA where $z_i(k) = \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{V}_i$ is the coordinate of $\boldsymbol{W}(k)$ in the direction of the eigenvector $\boldsymbol{V}_i (i = 1, 2, 3, 4, \ldots, 12)$. In the simulation result, $z_i(k) (i = 2, 3, 4, \ldots, 12)$ converges to zero and $z_1(k)$ converges to a constant 1, as $k \to \infty$, which is consistent with the convergence results in Theorem 6.10. Figure 6.7 shows the convergence of the component $z_i(k)$ of $\boldsymbol{W}(k)$ in (6.39) for MCA. In the simulation result, $z_i(k)(i = 1, 2, 3, \ldots, 11)$ converges to zero and $z_{12}(k)$ converges to a constant 1, as $k \to \infty$, which is consistent with the convergence results in Theorem 6.11.

From the simulation results shown in Figs. 6.6 and 6.7, we can see that on conditions of $\eta\lambda_1 \le 0.25$, $\eta \le 0.3$, and $\|\boldsymbol{W}(0)\| \le 1$, Algorithm (6.39) for PCA converge to the direction of the principal component of the correlation matrix. And if we simply switch the sign in the same learning rule, Algorithm (6.39) for MCA also converge to the direction of minor component of the correlation matrix. Besides, further simulations with high dimensions, e.g., 16, 20, and 30, also show

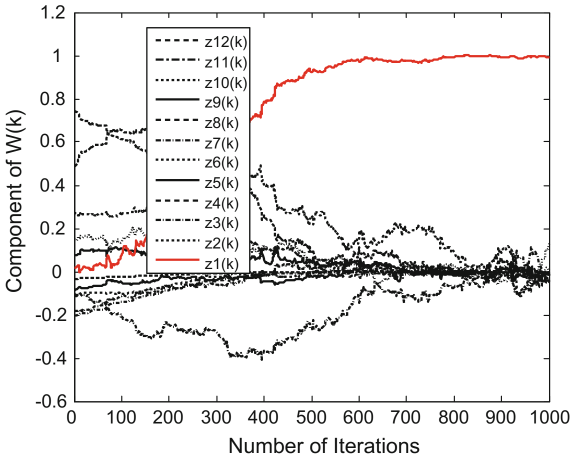

Fig. 6.6 Convergence of $\boldsymbol{W}(k)$ for PCA

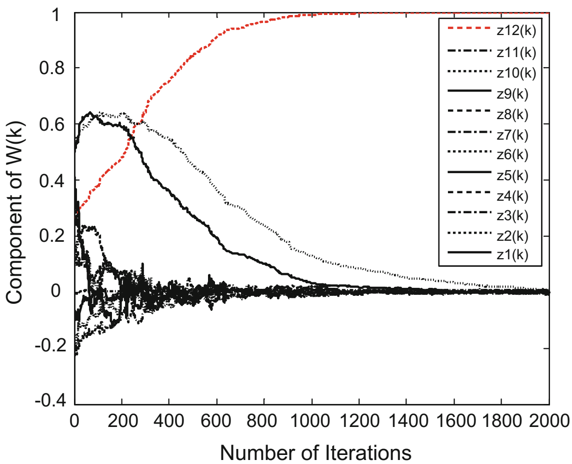

Fig. 6.7 Convergence of $\boldsymbol{W}(k)$ for MCA

that Algorithm (6.39) has satisfactory convergence under the conditions of Theorems 6.10 and 6.11. Figures 6.8 and 6.9 show the simulation results of Chen's PCA and MCA algorithm with dimension 20, respectively, where the learning rate for PCA is $\eta = 0.05$ and the learning rate for MCA is $\eta = 0.20$, which satisfy the condition of $\eta\lambda_1 \leq 0.25$ and $\eta \leq 0.3$.

In this section, dynamics of a unified self-stability learning algorithm for principal and minor components extraction are analyzed by the DDT method. The learning rate is assumed to be constant and thus not required to approach zero as required by the DCT method. Some sufficient conditions to guarantee the convergence are derived.

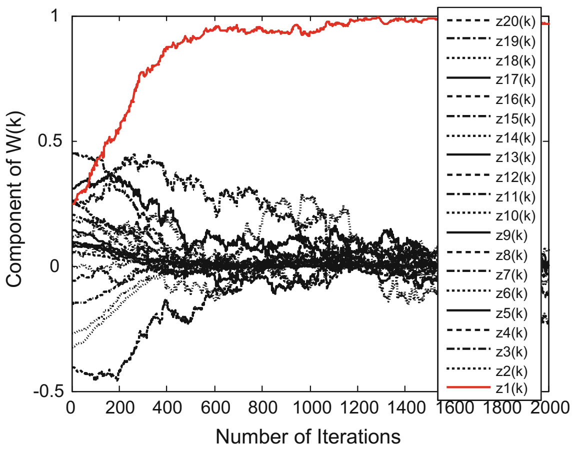

Fig. 6.8 Convergence of $\boldsymbol{W}(k)$ for PCA

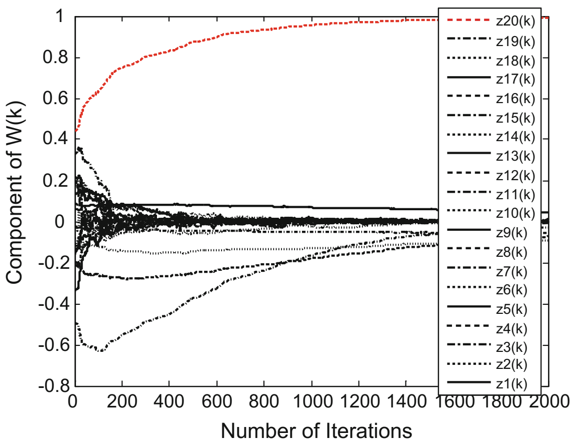

Fig. 6.9 Convergence of $\boldsymbol{W}(k)$ for MCA

6.5 Summary

In this chapter, we have analyzed the DDT systems of neural network principal/minor component analysis algorithms in details. First, we have reviewed several convergence or stability performance analysis methods for neural network-based PCA/MCA algorithms. Then, a DDT system of a novel MCA algorithm proposed by us has been analyzed. Finally, we have removed the assumption that the smallest eigenvalue of the correlation matrix of the input data is single, and a DDT system of a unified PCA and MCA algorithm has been analyzed.

References

1. Zhang, Q., & Leung, Y. W. (1995). Energy function for one-unit Oja algorithm. *IEEE Transactions on Neural Networks, 6*(9), 1291–1293.
2. Zhang, Q., & Bao, Z. (1995). Dynamical system for computing the eigenvectors with the largest eigenvalue of a positive definite matrix. *IEEE Transactions on Neural Networks, 6*(5), 790–791.
3. Yan, W. Y., Helmke, U., & Moore, J. B. (1994). Global analysis of Oja's flow for neural networks. *IEEE Transactions on Neural Networks, 5*(9), 674–683.
4. Plumbley, M. D. (1995). Lyapunov functions for convergence of principal component algorithms. *Neural Networks, 8*(1), 11–23.
5. Kushner, H. J., & Clark, D. S. (1976). *Stochastic approximation methods for constrained and unconstrained systems*. New York: Springer-Verlag.
6. Ljung, L. (1977). Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control, 22*(4), 551–575.
7. Cirrincione, G., Cirrincione, M., Herault, J., & Van Huffel, S. (2002). The MCA EXIN neuron for the minor component analysis. *IEEE Transactions on Neural Networks, 13*(1), 160–187.
8. Zufiria, P. J. (2002). On the discrete-time dynamics of the basic Hebbian neural network node. *IEEE Transactions on Neural Networks, 13*(6), 1342–1352.
9. Zhang, Q. (2003). On the discrete-time dynamics of a PCA learning algorithm. *Neurocomputing, 55*(10), 761–769.
10. PengD, Z., Zhang, Y., & Xiang, Y. (2008). On the discrete time dynamics of a self-stabilizing MCA learning algorithm. *Mathematical and Computer Modeling, 47*(9–10), 903–916.
11. Peng, D. Z., Zhang, Y., Lv, J. C., & Xiang, Y. (2008). A stable MCA learning algorithm. *Computers and Mathematics with Applications, 56*(4), 847–860.
12. Ye, M., Fan, X., & Li, X. (2006). A class of self-stabilizing MCA learning algorithms. *IEEE Transactions on Neural Networks, 17*(6), 1634–1638.
13. Zhang, Y., Ye, M., Lv, J. C., & Tan, K. K. (2005). Convergence analysis of a deterministic discrete time system of Oja's PCA learning algorithm. *IEEE Transactions on Neural Networks, 16*(6), 1318–1328.
14. Luo, F. L., & Unbehauen, R. (1997). A minor subspace analysis algorithm. *IEEE Transactions on Neural Networks, 8*(5), 1149–1155.
15. Luo F. L., & Unbehauen R. (1997). A generalized learning algorithm of minor component. In *Proceeding of International Conference on Acoustic, Speech, Signal Processing*, *4*(4), 3229–3232
16. Feng, D. Z., Bao, Z., & Jiao, L. C. (1998). Total least mean squares algorithm. *IEEE Transactions on Signal Processing, 46*(6), 2122–2130.
17. Feng, D. Z., Zheng, W. X., & Jia, Y. (2005). Neural network Learning algorithms for tracking minor subspace in high-dimensional data stream. *IEEE Transactions on Neural Network, 16* (3), 513–521.
18. Kong, X. Y., Hu, C. H., & Han, C. Z. (2010). On the discrete-time dynamics of a class of self-stabilizing MCA extraction algorithms. *IEEE Transactions on Neural Networks, 21*(1), 175–181.
19. Peng, D. Z., Zhang, Y., & Xiang, Y. (2009). A unified learning algorithm to extract principal and minor components. *Digital Signal Processing, 19*(4), 640–649.
20. Kong, X. Y., An, Q. S., Ma, H. G., Han, C. Z., & Zhang, Q. (2012). Convergence analysis of deterministic discrete time system of a unified self-stabilizing algorithm for PCA and MCA. *Neural Networks, 36*(8), 64–72.
21. Oja, E. (1992). Principal component, minor component and linear neural networks. *Neural Networks, 5*(6), 927–935.
22. Chen, T., & Amari, S. (2001). Unified stabilization approach to principal and minor components extraction algorithms. *Neural Networks, 14*(10), 1377–1387.

23. Möller, R. (2004). A self-stabilizing learning rule for minor component analysis. *International Journal of Neural Systems, 14*(1), 1–8.
24. Hasan M. A. (2007). Self-normalizing dual systems for minor and principal component extraction. In *Proceeding of International Conference on Acoustic, Speech, and Signal Processing*,*4*, 885–888.
25. Reddy, V. U., Egardt, B., & Kailath, T. (1982). Least squares type algorithm for adaptive implementation of Pisarenko's harmonic retrieval method. *IEEE Transactions on Acoustic, Speech, and Signal Processing, 30*(3), 399–405.
26. Xu, L., Oja, E., & Suen, C. (1992). Modified Hebbian learning for curve and surface fitting. *Neural Networks, 5*(3), 441–457.
27. Ouyang, S., Bao, Z., Liao, G. S., & Ching, P. C. (2001). Adaptive minor component extraction with modular structure. *IEEE Transactions on Signal Processing, 49*(9), 2127–2137.
28. Zhang, Q., & Leung, Y. W. (2000). A class of learning algorithms for principal component analysis and minor component analysis. *IEEE Transactions on Neural Networks, 11*(2), 529–533.
29. Douglas, S. C., Kung, S. Y., & Amari, S. (1998). A self-stabilized minor subspace rule. *IEEE Signal Processing Letter, 5*(12), 328–330.
30. LaSalle, J. P. (1976). *The stability of dynamical systems*. Philadelphia, PA: SIAM.
31. Magnus, J. R., & Neudecker, H. (1991). *Matrix differential calculus with applications in statistics and econometrics* (2nd ed.). New York: Wiley.
32. Kong, X. Y., Han, C. Z., & Wei, R. X. (2006). Modified gradient algorithm for total least squares filtering. *Neurocomputing, 70*(1), 568–576.