Convergence Analysis of a Self-Stabilizing Algorithm for Minor Component Analysis

Haidi Dong, Yingbin Gao, and Gang Liu

***Abstract*—The Möller algorithm is a self-stabilizing minor component analysis algorithm. This research document involves the study of the convergence and dynamic characteristics of the Möller algorithm using the deterministic discrete time (DDT) methodology. Unlike other analysis methodologies, the DDT methodology is capable of serving the distinct time characteristic and having no constraint conditions. Through analyzing the dynamic characteristics of the weight vector, several convergence conditions are drawn, which are beneficial for its application. The performing computer simulations and real applications demonstrate the correctness of the analysis's conclusions.**

***Index Terms*—Convergence analysis, deterministic discrete time (DDT), dynamic characteristic, Möller algorithm.**

I. Introduction

MINOR component analysis (MCA) is termed as a robust instrument in several areas, such as instance frequency estimation [1], total least squares (TLS) [2], and filter design [3]. Hebbian neural network oriented MCA algorithms that are capable of adaptively estimating minor components (MCs) from input signals and quickly tracing moving signals have garnered considerable attention [4]. Recently, several MCA algorithms have been suggested, such as the Möller algorithm [5], the stable data projection method (SDPM) algorithm [6], the generalized orthogonal projection approximation and subspace tracking (GOPAST) algorithm [7], etc.

For neural network algorithms, one of the key research tasks involves the analysis of their convergence and dynamic characteristic during all iterations. To the best of our knowledge, three methodologies can be used to accomplish this task. They are the Lyapunov function methodology [2], the deterministic continuous time (DCT) methodology [8] and the deterministic discrete time (DDT) methodology [9]. The Lyapunov function methodology is capable of merely proving whether this algorithm possesses the potential to estimate the desired components or not while being unable to deliver convergence conditions. The DCT methodology requires the mandatory approximation of the learning rate to zero, which is difficult to satisfy in real applications. Contrary to the above two methodologies, the DDT methodology is capable of preserving the distinct time characteristic of the original algorithms and establishing constraint conditions with respect to MCA algorithms. Due to these benefits, the extensive use of the DDT methodology has been applied for analyzing the convergence characteristic of some algorithms [10], [11].

In [5], Möller proposed an MCA algorithm (denoted as the Möller algorithm as follows). Through analyzing the characteristics of its weight vector norm, Möller proved that this algorithm is self-stabilizing. By changing the weight vector into a weight matrix and adding a diagonal matrix to the Möller algorithm, Gao *et al*. [12] modified this algorithm into a multiple minor component extraction algorithm. The convergence analysis of the Möller algorithm was finished through the DCT methodology in [13]. As mentioned above, the DDT methodology is more advantageous than the DCT methodology. Therefore, we sought to analyze the Möller algorithm through the DDT methodology.

The organization of the remainder of this document is as follows. Section II presents several preliminaries and a concise description of the Möller algorithm. Section III analyzes the convergence and dynamic characteristics of the Möller algorithm and derives some convergence conditions. In Section IV, two computer simulations and two real application experiments are conducted to confirm the correctness of the obtained conclusions. Ultimately, we present the study's conclusions in Section V.

II. Preliminaries and Möller Algorithm

On the bases of the Hebbian neural network, Möller suggested a self-stabilizing MCA algorithm [5]. Its discrete time equation is given by

$$\boldsymbol{w}(k+1)=\boldsymbol{w}(k)+\eta y^2(k)\boldsymbol{w}(k)-\eta y(k)\left[2\boldsymbol{w}(k)^T\boldsymbol{w}(k)-1\right]\boldsymbol{x}(k) \quad (1)$$

where $\boldsymbol{w}(k)$ is the weight vector, $\boldsymbol{x}(k)$ is an input signal and $y(k)$ denotes the neural network output. In [5], Moller proved that after some iterations, the norm of the weight vector $\boldsymbol{w}(k)$ must converge to 1 and has no relationship with the initial weight vector. That is to say that Möller algorithm is as self-stabilizing.

The DDT system of the Möller algorithm can be formulated by the following steps [9]. By incorporating the conditional

Manuscript received March 28, 2018; accepted May 17, 2018. This work was supported by the National Natural Science Foundation of China (61903375, 61673387, 61374120) and Shaanxi Province Natural Science Foundation (2016JM6015). Recommended by Associate Editor Xin Luo. *(Corresponding author: Yingbin Gao.)*

Citation: H. D. Dong, Y. B. Gao, and G. Liu, "Convergence analysis of a self-stabilizing algorithm for minor component analysis," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 6, pp. 1585–1592, Nov. 2020.

H. D. Dong is with the College of Weaponry Engineering, Naval University of Engineering, Wuhan 430033, China (e-mail: donghaidi123@163.com).

Y. B. Gao is with the 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang 050081, China (e-mail: welcome 8793@sina.com).

G. Liu is with Xi'an Research Institute of High Technology, Xi'an 710025, China (e-mail: liugangepgc@163.com).

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

Digital Object Identifier 10.1109/JAS.2019.1911636

expectation operator $E\{\boldsymbol{w}(k+1)/\boldsymbol{w}(0), \boldsymbol{x}(i), i<k\}$ into (2) and identifying the conditional expected value as the next iteration, the DDT system is obtained and given by

$$\begin{aligned}\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \eta \boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k)\boldsymbol{w}(k)\\ - \eta(2\boldsymbol{w}^T(k)\boldsymbol{w}(k)-1)\boldsymbol{R}\boldsymbol{w}(k)\end{aligned} \tag{2}$$

where $\boldsymbol{R} = E[\boldsymbol{x}(k)\boldsymbol{x}^T(k)]$ represents the autocorrelation matrix of $\boldsymbol{x}(k)$. Comparing (2) with (1), we can find that (1) can be interpreted as stochastic approximation of its DDT system by replacing its correlation matrix $\boldsymbol{R}$ with its rank-1 instantaneous approximation. According to [9], this approximation operation does not change the convergence property of Moller algorithm. In other words, online algorithm and its DDT system has same convergence property.

By using the matrix theory [14], has non-negative eigenvalues ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{n-1} \geq \lambda_n \geq 0$) and corresponding eigenvectors $\boldsymbol{v}_i$ $(i=1,2,\ldots,n)$. Suppose that p $(1 \leq p \leq n)$ is the multiplicity of the smallest eigenvalue (denoted as τ herein). Thus, the eigenvectors corresponding to τ are defined as minor subspaces and are represented by $\boldsymbol{V}_\tau = \mathrm{span}(\boldsymbol{v}_{n-p+1}, \boldsymbol{v}_{n-p+2}, \ldots, \boldsymbol{v}_n)$.

Clearly, $\{\boldsymbol{v}_i | i=1,2,\ldots,n\}$ merely composes an orthogonal foundation of $_{n\times n}$. Through the use of this basis, $\boldsymbol{w}(k)$ and $\boldsymbol{R}\boldsymbol{w}(k)$ are decomposed into

$$\begin{cases}\boldsymbol{w}(k) = \displaystyle\sum_{i=1}^{n} z_i(k)\boldsymbol{v}_i\\ \boldsymbol{R}\boldsymbol{w}(k) = \displaystyle\sum_{i=1}^{n} \lambda_i z_i(k)\boldsymbol{v}_i\end{cases} \tag{3}$$

where $z_i(k) = \boldsymbol{w}^T(k)v_i$, $i=1,2,\ldots,n$ is the projection of the weight vector onto the eigenvector $\boldsymbol{v}_i$.

From (2) and (3), we can derive the following equation

$$z_i(k+1) = \left\{1+\eta \boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k) - \eta(2\|\boldsymbol{w}(k)\|^2-1)\lambda_i\right\} z_i(k) \tag{4}$$

where $i=1,2,\ldots,n$.

In accordance with the Rayleigh quotient [15], it holds that

$$0 \leq \lambda_n \boldsymbol{w}^T(k)\boldsymbol{w}(k) \leq \boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k) \leq \lambda_1 \boldsymbol{w}^T(k)\boldsymbol{w}(k) \tag{5}$$

where $\boldsymbol{w}(k) \neq 0$ and $k \geq 0$.

III. Convergence and Dynamic Characteristic Analysis of Möller Algorithm

In this section, we will study the convergence and dynamic characteristic of the Möller algorithm through the DDT methodology. In detail, it is necessary to prove that if certain sufficient conditions are given, any weight vector can tend to the direction of the eigenvector with respect to τ. To finish this proof, we first need to prove that the DDT system of the Möller algorithm is bounded, which is given by Theorem 1 and Theorem 2 as follows.

Theorem 1: Let us assume that $\eta\lambda_1 \leq 0.25$. If $\boldsymbol{w}(0) \notin \boldsymbol{V}_\tau^\perp$ and $\|\boldsymbol{w}(0)\| \leq 1$, it stands valid that $\|\boldsymbol{w}(k)\| < 1+2\eta\lambda_1$.

Proof: See Appendix A. ■

Theorem 2: Let us assume that $\eta\lambda_1 \leq 0.25$. If $\boldsymbol{w}(0) \notin \boldsymbol{V}_\tau^\perp$ and $\|\boldsymbol{w}(0)\| \leq 1$, it stands valid that $\|\boldsymbol{w}(k)\| > c^k\|\boldsymbol{w}(0)\|$, where $c = \{1-\eta\lambda_1[2(1+2\eta\lambda_1)^2-1]\}$.

Proof: See Appendix B. ■

Theorems 1 and 2 suggest that during the iterations, the DDT system of the Möller algorithm has both upper and lower bounds. The two conclusions lay the foundation for the future proof. From (3), it is easy to see that the weight vector can be expressed by its coefficients (i.e., the characteristic of $\boldsymbol{w}(k)$ is decided by $z_i(k)$). Next, we will analyze the dynamic behaviors of $z_i(k)$, which will be provided by the following three lemmas.

Lemma 1: Let us assume that $\eta\lambda_1 \leq 0.25$. If $\boldsymbol{w}(0) \notin \boldsymbol{V}_\tau^\perp$ and $\|\boldsymbol{w}(0)\| \leq 1$, there must exist two constants $\theta_1 > 0$ and $\Pi_1 > 0$ in a way that $\sum_{j=1}^{n-p} z_j^2(k) \leq \Pi_1 e^{-\theta_1 k}$.

Proof: See Appendix C. ■

Lemma 2: Let us assume that $\eta\lambda_1 \leq 0.25$. If $\boldsymbol{w}(0) \notin \boldsymbol{V}_\tau^\perp$ and $\|\boldsymbol{w}(0)\| \leq 1$, there must exist two constants $\theta_2 > 0$ and $\Pi_2 > 0$ in a way that

$$\begin{aligned}&\left|\boldsymbol{w}^T(k+1)\boldsymbol{R}\boldsymbol{w}(k+1) - (2\|\boldsymbol{w}(k+1)\|^2-1)\tau\right|\\ &\quad \leq (k+1)\Pi_2\left[e^{-\theta_2(k+1)} + \max\left\{e^{-\theta_2 k}, e^{-\theta_1 k}\right\}\right].\end{aligned}$$

Proof: See Appendix. D. ■

Lemma 3: Let us assume that there exist two constants $\theta > 0$ and $\Pi > 0$ in a way that

$$\eta\left|\left[\boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k) - (2\|\boldsymbol{w}(k)\|^2-1)\tau\right]z_i(k)\right| \leq k\Pi e^{-\theta k}.$$

Then, it holds that

$$\lim_{k\to\infty} z_i(k) = z_i^*, \;\; i = n-p+1, n-p+2, \ldots, n$$

where z_i^* $(i=n-p+1, n-p+2, \ldots, n)$ are some constants and represent the convergence values.

Proof: See Appendix. E. ■

To date, we have established the dynamic trajectory $z_i(k)$ through all of the iterations. Through the use of these conclusions, we are able to present the convergence analysis of the weight vector in the Möller algorithm, which will be presented by Theorem 3.

Theorem 3: Let us assume that $\eta\lambda_1 \leq 0.25$. If $\boldsymbol{w}(0) \notin \boldsymbol{V}_\tau^\perp$ and $\|\boldsymbol{w}(0)\| \leq 1$, the weight vector in the Möller algorithm must converge to the direction of the eigenvector corresponding to τ.

Proof: From Lemma 1, the constants $\theta_1 > 0$ and $\Pi_1 \geq 0$ exist such that

$$\sum_{i=1}^{n-p} z_i^2(k) \leq \Pi_1 e^{-\theta_1 k}, \; k \geq 0. \tag{6}$$

Through Lemma 2, the constants $\theta_2 > 0$ and $\Pi_2 \geq 0$ exist such that

$$\begin{aligned}&\left|\boldsymbol{w}^T(k+1)\boldsymbol{R}\boldsymbol{w}(k+1) - (2\|\boldsymbol{w}(k+1)\|^2-1)\tau\right|\\ &\quad \leq (k+1)\Pi_2\left[e^{-\theta_2(k+1)} + \max\left\{e^{-\theta_2 k}, e^{-\theta_1 k}\right\}\right].\end{aligned} \tag{7}$$

Clearly, it is easy to find two constants $\theta > 0$ and $\Pi > 0$ such that

$$\eta\left|\left[\boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k) - (2\|\boldsymbol{w}(k)\|^2-1)\tau\right]z_i(k)\right| \leq k\Pi e^{-\theta k} \tag{8}$$

where $i = n-p+1, n-p+2, \ldots, n$. Through the use of Lemmas 1 and 3, it holds valid that

$$\begin{cases}\displaystyle\lim_{k\to+\infty} z_i(k) = 0, & i=1,2,\ldots,n-p\\ \displaystyle\lim_{k\to+\infty} z_i(k) = z_i^*, & i=n-p+1, n-p+2, \ldots, n.\end{cases} \tag{9}$$

By using (3) and (9), it is easy to obtain that

$$\lim_{k\to+\infty} \boldsymbol{w}(k) = \lim_{k\to+\infty} \sum_{i=1}^{n} z_i(k)\boldsymbol{v}_i$$
$$= \lim_{k\to+\infty} \sum_{i=1}^{n-p} z_i(k)\boldsymbol{v}_i + \lim_{k\to+\infty} \sum_{i=n-p+1}^{n} z_i(k)\boldsymbol{v}_i = \sum_{i=n-p+1}^{n} z_i^{*}\boldsymbol{v}_i. \tag{10}$$

Clearly, the convergence values of $\boldsymbol{w}(k)$ is combined by the minor components $\boldsymbol{v}_i$ $(i = n-p+1,\ldots,n)$. In other words, the weight vector converge to the direction of the eigenvector corresponding to τ. ■

Next, we make several comments on the obtained conditions. From Theorem 3, we know that there are two limiting conditions. The first one requires that $\eta\lambda_1 \le 0.25$. In real applications, if the largest eigenvalue λ_1 is explicitly provided, it is very easy to calculate the suitable learning rate. If the exact value of λ_1 is unknown, under this case, Zhang *et al.* [16] proposed a new method to estimate it. By using the estimated value, the first condition can be satisfied. The second condition requires that $\boldsymbol{w}(0) \notin \boldsymbol{V}_\tau{}^{\perp}$, which is reached with a probability of 1 when the initial weight vector is randomly generated. The two easily stratified conditions mean that they are very useful for the applications of the Möller algorithm.

IV. Computer Simulations and Real Applications

In this section, we apply the Möller algorithm on four experiments for the purpose of illustrating the correctness of the obtained conclusions. The first experiment is meant to exhibit the dynamic characteristics of the Möller algorithm, whereas the second one presents an example in which the Möller algorithm exhibits the divergence phenomenon because of the discontentment of the attained conditions. In the third one, we investigate the performance of the Möller algorithm on addressing one data fitting issue. In the fourth one, the Möller algorithm is used the estimate the parameters of one filter.

A. Dynamic Trajectories of Möller Algorithm

Consider the following 5×5 symmetric positive definite matrix, which is created in a random manner.

$$\boldsymbol{R}_1 = \begin{bmatrix} 0.5684 & -0.0960 & -0.1307 & 0.1588 & 0.2084 \\ -0.0960 & 0.4461 & -0.0434 & 0.2871 & -0.1961 \\ -0.1307 & -0.0434 & 0.5386 & -0.0336 & 0.1738 \\ 0.1588 & 0.2871 & -0.0336 & 0.7409 & -0.0259 \\ 0.2084 & -0.1961 & 0.1738 & -0.0259 & 0.5166 \end{bmatrix}. \tag{11}$$

The eigenvalues of $\boldsymbol{R}_1$ are $\lambda_1 = 0.9825, \lambda_2 = 0.8391, \lambda_3 = 0.6264, \lambda_4 = \lambda_5 = 0.1813$. Next, the Möller algorithm is implemented for estimating the MC of $\boldsymbol{R}_1$. The preliminary parameters of the Möller algorithm have been established hereunder. The learning rate is fixed to be $\eta = 0.25$ and the random generated preliminary weight vector is adopted. Clearly, all of the convergence conditions can be catered to through these settings.

Fig. 1 illustrates the convergence of the component $z_i(k)$, $(i = 1,2,3,4,5)$. As is evident from the Fig. 1, $z_i(k)$ $(i = 1,2,3)$ has the tendency to be zero following approximately 50 iterations, whereas $z_i(k)$ $(i = 4,5)$ converges to a constant following approximately 30 iterations. The trajectories of $z_i(k)$ are consistent with the conclusion evidenced in Theorem 3. As suggested by this experiment, we conclude that, subject to the attained convergence conditions, the Möller algorithm possesses a reasonable convergence characteristic.

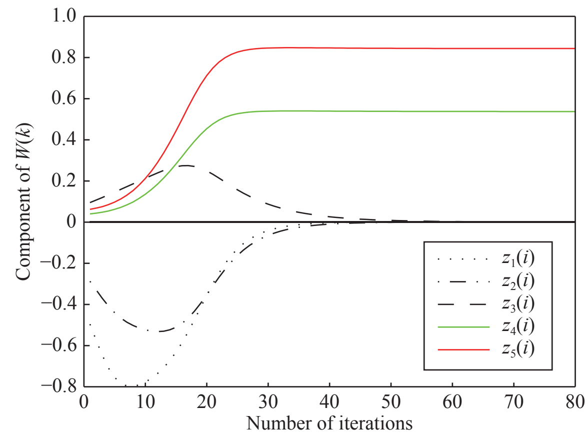


Fig. 1. Convergence of the component $z_i(k)$.

B. Divergency Experiment

Next, we provide an example through this experiment wherein the Möller algorithm diverges due to the fact that the conditions applied in Theorem 3 have not been addressed. Matrix (11) continues to be utilized in this experiment. Herein, we establish that $\eta = 0.4$, which clearly does not meet the condition $\eta\lambda_1 \le 0.25$. Fig. 2 illustrates the simulation findings as the initial vector that is provided by $\boldsymbol{w} = [1.0391, -1.1176, 1.2607, 0.6601, -0.0679]^T$. As is evident from this figure, we obtain that the Möller algorithm may diverge based on the dissatisfaction of the convergence conditions. Conversely, this experiment confirms the precision of the conclusions in Theorem 3.

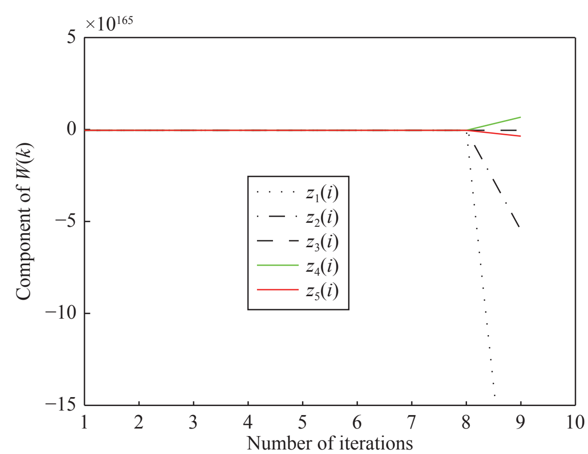


Fig. 2. Divergence of the component $z_i(k)$.

C. Data Fitting Experiment

There are extensive latent applications of MCA algorithms, and one of the significant ones solves the fitting issue. We have delivered one surface fitting example in this experiment for showing the efficacy of the Möller algorithm subject to the attained conditions. Let us take into consideration the following curved surface.

$$2x^2+0.5y^2-z^2=0. \tag{12}$$

Fig. 3 illustrates the surface stated by (12). Through the sampling of this surface in such a way that the sampling intervals are evenly distributed on the plane, we are able to attain a 3-dimensional data set $G=\{(x_i,y_i,z_i)\}$. Through adding Gaussian noises to this data set, the observation data set $\tilde{G}=\{(\tilde{x}_i,\tilde{y}_i,\tilde{z}_i)\}$ is acquired. Fig. 4 presents the observed data in $\tilde{G}$. The issue of surface fitting involves discovering a parameterized framework $a_1x^2+a_2y^2+a_3z^2=0$ for fitting $\tilde{G}$, which can be solved through calculating the MC of the data set $F=\left\{(u_i,v_i,w_i)|u_i=\tilde{x}_i^2, v_i=\tilde{y}_i^2, w_i=\tilde{z}_i^2\right\}$.

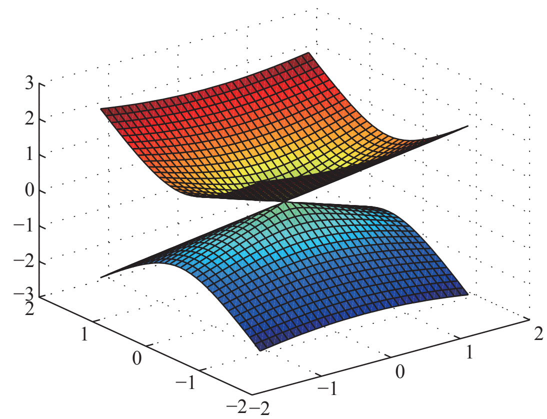


Fig. 3. Original surface.

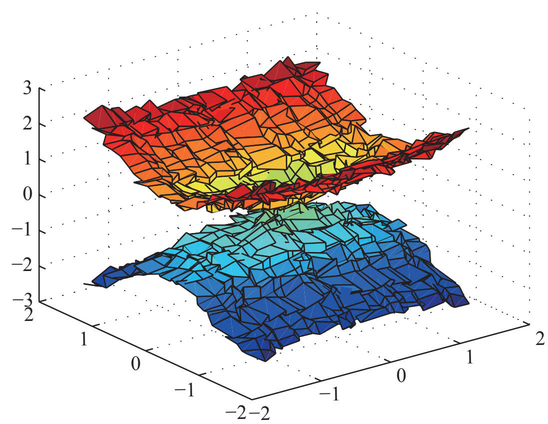


Fig. 4. Noise disturbed data set.

In this experiment, the input signal is produced through the random selection of vectors from the data set F. The eigenvalues of its autocorrelation matrix are $\lambda_1=27.4800$, $\lambda_2=21.3790$ and $\lambda_3=1.4554$. Then, Möller algorithm is applied for calculating the MC from this input signal. The preliminary learning rate is established as $\eta=0.008$, which stratifies the conditions in Theorem 3. Fig. 5 presents the convergence of the elements of the weight vector. After 3500 iterations, the ultimate convergence values of the weight vector are $\boldsymbol{w}*=[0.8227,0.2122,-0.4160]^T$, which possess the same direction as that of the coefficient vector $[2,0.5,-1]$ in (12). As suggested by this experiment, we are capable of observing that, subject to the extracted conditions, the Möller algorithm possesses a convergence fitting performance.

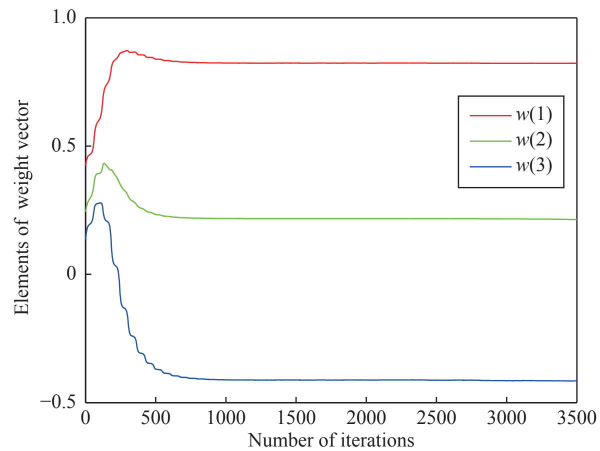


Fig. 5. Convergence of the weight vector elements.

D. Filter Coefficients Estimation Experiment

Another important application of the MCA algorithm is to estimate the filter coefficients. Fig. 6 shows the classical structure of this issue, where $\boldsymbol{x}(t)$ is the input signal, $\boldsymbol{y}(t)$ is the output of the unknown system and $\boldsymbol{d}(t)$ is the output of the adaptive filter. $\boldsymbol{n}_1(t)$ and $\boldsymbol{n}_2(t)$ are two additional Gaussian white noises. Suppose that the coefficient vector of the unknown system is $\boldsymbol{h}$. Then, the optimal estimation of $\boldsymbol{h}$ can be obtained by solving the following minimum problem, which has been proven in [17]

$$\boldsymbol{w}^*=\arg\min J(\boldsymbol{w}),\ \ J(\boldsymbol{w})=\frac{\boldsymbol{w}^T\boldsymbol{R}\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{w}} \tag{13}$$

where $\boldsymbol{w}=[\boldsymbol{h}^T,-1]^T$ and $\boldsymbol{R}$ is the autocorrelation matrix of the augmented vector $\boldsymbol{z}=[\boldsymbol{x}(k)+\boldsymbol{n}_1(k),\boldsymbol{y}(k)+\boldsymbol{n}_2(k)]$. Clearly, the solution of (13) justly composes the MC of $\boldsymbol{R}$.

In this experiment, let the coefficient vector of the unknown system be given by $\boldsymbol{h}=[-0.3, -0.9, 0.8, -0.7, -0.6, 0.1, 0.3, -0.5, 0.5, -0.4]$. The unknown system includes Gaussian white noise and the largest eigenvalue of $\boldsymbol{R}$ is $\lambda_1=11.7896$.

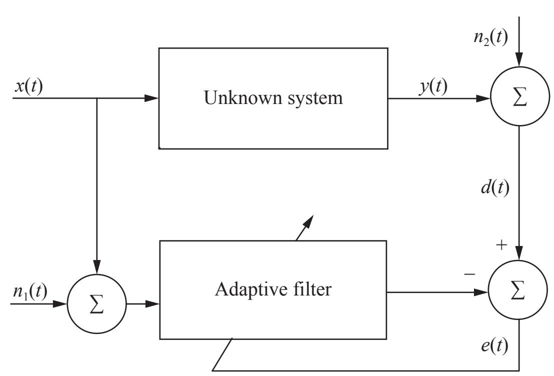


Fig. 6. Structure of the finite impulse response (FIR) filter.

For the Möller algorithm, the learning rate is set as $\eta = 0.02$. The estimation results of $\boldsymbol{h}$ are presented in Fig. 7. From this figure, we see that after approximately 1300 iterations, all the components of $\boldsymbol{h}$ tend to some constants. The final convergence values of $\boldsymbol{h}$ are [−0.3, −0.9, 0.8, −0.7, −0.6, 0.1, 0.3, −0.5, 0.5, −0.4], which are the same as the exact ones. From this experiment, we confirm again that under the derived conditions, the Möller algorithm has good convergence performance.

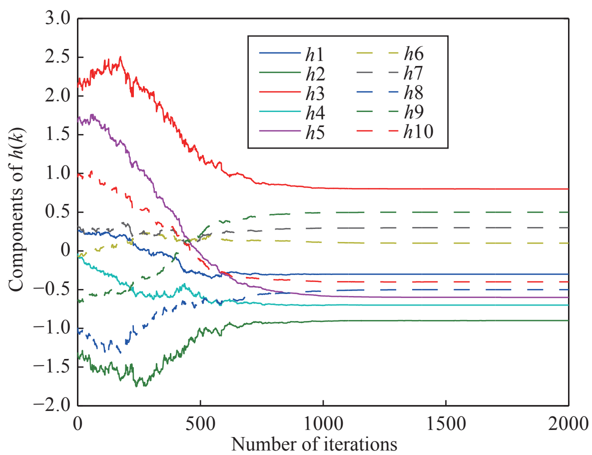


Fig. 7. Dynamic trajectories of the elements of $\boldsymbol{h}$.

V. Conclusions

As a self-stabilizing MCA algorithm, the Möller algorithm has been applied in various areas. In this research document, we have established the dynamic characteristic and convergence analysis of the Möller algorithm using the DDT methodology. First, it is verified that the DDT system of the Möller algorithm is bounded. Second, the dynamic trajectories of the weight vector projection are depicted. Third, based on the derived lemmas, we prove that after sufficient iterations, the weight vector must converge to the direction of the desired MC. Finally, computer simulations and real applications validate the precision of the attained conclusions.

Appendix A
Proof of Theorem 1

Proof: From (3) and (4), we have

$$\begin{aligned}\|\boldsymbol{w}(k+1)\|^2 &= \sum_{i=1}^{n} z_i^2(k+1)\\ &= \sum_{i=1}^{n}\left\{1-\eta[\lambda_i(2\|\boldsymbol{w}(k)\|^2-1)-\boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k)]\right\}^2 z_i^2(k)\\ &\le \sum_{i=1}^{n}\left[1-\lambda_i\eta(2\|\boldsymbol{w}(k)\|^2-1)+\eta\lambda_1\|\boldsymbol{w}(k)\|^2\right]^2 z_i^2(k)\\ &< (1+\eta\lambda_1+\eta\lambda_1\|\boldsymbol{w}(k)\|^2)^2\sum_{i=1}^{n} z_i^2(k)\\ &= (1+\eta\lambda_1+\eta\lambda_1\|\boldsymbol{w}(k)\|^2)^2\|\boldsymbol{w}(k)\|^2. \end{aligned}\tag{14}$$

Thereafter, it stands valid that

$$\|\boldsymbol{w}(k+1)\|^2 < (1+\eta\lambda_1+\eta\lambda_1\|\boldsymbol{w}(k)\|^2)^2\|\boldsymbol{w}(k)\|^2 \tag{15}$$

where $s = \|\boldsymbol{w}(k)\|^2$ and $f(s) = \|\boldsymbol{w}(k+1)\|^2$. By substituting them into (15), we get $f(s) = (1+\eta\lambda_1+\eta\lambda_1 s)^2 s$, which is consistent in the interval $[0, 1]$. The gradient of $f(s)$ with respect to s is represented by

$$\dot{f}(s) = (1+\eta\lambda_1+\eta\lambda_1 s)(1+\eta\lambda_1+3\eta\lambda_1 s). \tag{16}$$

If $s_1 = -(1+\eta\lambda_1)/\eta\lambda_1$ or $s_2 = -(1+\eta\lambda_1)/3\eta\lambda_1$, then $\dot{f}(s) = 0$. Through the use of $\eta > 0$ and $\lambda_1 > 0$, we obtain $s_1 < s_2 < 0$. Accordingly, it stands valid that $\dot{f}(s) > 0$ with respect to all $0 \le s \le 1$, which implies that $f(s)$ exhibits a monotonic increase over the interval $[0, 1]$. Thereafter, it holds that

$$f(s) \le f(1) < (1+2\eta\lambda_1)^2. \tag{17}$$

Thus, we obtain $\|\boldsymbol{w}(k)\| < 1+2\eta\lambda_1$. ■

Appendix B
Proof of Theorem 2

Proof: Through the use of (2)–(4), we obtain

$$\begin{aligned}\|\boldsymbol{w}(k+1)\|^2 &= \sum_{i=1}^{n}\left\{1-\eta\lambda_i(2\|\boldsymbol{w}(k)\|^2-1)+\eta\boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k)\right\}^2 z_i^2(k)\\ &\ge \sum_{i=1}^{n}\left\{1-\eta\lambda_1(2\|\boldsymbol{w}(k)\|^2-1)+\eta\lambda_n\|\boldsymbol{w}(k)\|^2\right\}^2 z_i^2(k)\\ &> [1-\eta\lambda_1(2\|\boldsymbol{w}(k)\|^2-1)]^2\sum_{i=1}^{n} z_i^2(k)\\ &> [1-\eta\lambda_1(2(1+2\eta\lambda_1)^2-1)]^2\|\boldsymbol{w}(k)\|^2. \end{aligned}\tag{18}$$

Denote $c = \{1-\eta\lambda_1(2(1+2\eta\lambda_1)^2-1)\}$. Through the use of $\eta\lambda_1 \le 0.25$, we obtain

$$\begin{aligned} c &= \left[1-\eta\lambda_1(2(1+2\eta\lambda_1)^2-1)\right]\\ &> \left\{1-0.25*\left[2*(1+2*0.25)^2-1\right]\right\}\\ &= 0.125 > 0. \end{aligned}\tag{19}$$

From (18) and (19), we obtain

$$\|\boldsymbol{w}(k)\|^2 > c^2\|\boldsymbol{w}(k-1)\|^2 > \cdots > c^{2k}\|\boldsymbol{w}(0)\|^2. \tag{20}$$

As is evident from (20), it stands valid that $\|\boldsymbol{w}(k)\| > c^k\|\boldsymbol{w}(0)\|$. ■

Appendix C
Proof of Lemma 1

Proof: Since $\boldsymbol{w}^T(0) \notin \boldsymbol{V}_\tau{}^{\perp}$, there must be some constant i $((n-p+1) \le i \le n)$ such that $z_i(0) \ne 0$. Without the loss of generality, let us suppose that $z_n(0) \ne 0$. As is evident from (4), for each j $(1 \le j \le n-p)$, it follows that

$$\left[\frac{z_j(k+1)}{z_n(k+1)}\right]^2 = \left\{\frac{1-\eta\left[\lambda_j(2\|\boldsymbol{w}(k)\|^2-1)-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]}{1-\eta\left[\tau(2\|\boldsymbol{w}(k)\|^2-1)-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]}\right\}^2\left[\frac{z_j(k)}{z_n(k)}\right]^2$$

$$= \left\{1-\frac{\eta(\lambda_j-\tau)(2\|\boldsymbol{w}(k)\|^2-1)}{1+\eta\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)-\eta\tau(2\|\boldsymbol{w}(k)\|^2-1)}\right\}^2\left[\frac{z_j(k)}{z_n(k)}\right]^2$$

$$\le \left\{1-\frac{\eta(\lambda_j-\tau)(2\|\boldsymbol{w}(k)\|^2-1)}{1+\eta\lambda_1\|\boldsymbol{w}(k)\|^2-\eta\tau(2\|\boldsymbol{w}(k)\|^2-1)}\right\}^2\left[\frac{z_j(k)}{z_n(k)}\right]^2$$

$$= \left\{1-\frac{\eta(\lambda_j-\tau)(2\|\boldsymbol{w}(k)\|^2-1)}{1+\eta\lambda_1\left(1-\|\boldsymbol{w}(k)\|^2\right)+\eta\left(\lambda_1-\tau\right)(2\|\boldsymbol{w}(k)\|^2-1)}\right\}^2\left[\frac{z_j(k)}{z_n(k)}\right]^2$$

$$\le \beta_k\left[\frac{z_j(k)}{z_n(k)}\right]^2 \le \beta_k\beta_{k-1}\cdots\beta_0\left[\frac{z_j(0)}{z_n(0)}\right]^2$$

$$\le \beta^{k+1}\left[\frac{z_j(0)}{z_n(0)}\right]^2 = \frac{z_j^2(0)}{z_n^2(0)}e^{-\theta_1(k+1)} \tag{21}$$

where $\theta_1 = -\ln\beta$, $\beta = \max(\beta_0, \beta_1, \ldots, \beta_k)$ and

$$\beta_k = \left\{1-\frac{\eta(\lambda_j-\tau)(2\|\boldsymbol{w}(k)\|^2-1)}{1+\eta\lambda_1\left(1-\|\boldsymbol{w}(k)\|^2\right)+\eta\left(\lambda_1-\tau\right)(2\|\boldsymbol{w}(k)\|^2-1)}\right\}^2.$$

Through the use of $\eta\lambda_1 \le 0.25$, we obtain

$$\begin{aligned}
&1+\eta\lambda_1\left(1-\|\boldsymbol{w}(k)\|^2\right)\\
&\quad\ge 1+\eta\lambda_1\left(1-(1+2\eta\lambda_1)^2\right)\\
&\quad\ge 1+0.25*(1-(1+2*0.25)^2)\\
&\quad= 0.6875 > 0.
\end{aligned} \tag{22}$$

This implies that $0 < \beta_k < 1$. Thus, we are able to obtain that $\theta_1 > 0$. Through the use of $c^{2k}\|\boldsymbol{w}(0)\| < \|\boldsymbol{w}(k)\| \le 1+2\eta\lambda_1$, we have that $z_n(k)$ is bounded (i.e., a constant $d > 0$ such that $z_n^2(k) \le d$ with respect to all $k \ge 0$). Subsequent to that, we are able to obtain

$$\sum_{i=1}^{n-p} z_i^2(k) = \sum_{i=1}^{n-p}\left[\frac{z_i(k)}{z_n(k)}\right]^2 z_n^2(k) \le \Pi_1 \mathrm{e}^{-\theta_1 k}. \tag{23}$$

With respect to all $k \ge 0$, where

$$\Pi_1 = d\sum_{i=1}^{n-p}\left[\frac{z_i(0)}{z_n(0)}\right]^2 \ge 0.$$

■

Appendix D
Proof of Lemma 2

Proof: From (2), we are able to obtain that

$$\begin{aligned}
&\|\boldsymbol{w}(k+1)\|^2\\
&= \sum_{i=1}^{n}\left\{1-\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\lambda_i-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]\right\}^2 z_i^2(k)\\
&= \sum_{i=1}^{n}\left\{1-\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\tau-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]\right\}^2 z_i^2(k)\\
&\quad+\sum_{i=1}^{n-p}\left\{1-\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\lambda_i-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]\right\}^2 z_i^2(k)\\
&\quad-\sum_{i=1}^{n-p}\left\{1-\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\tau-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]\right\}^2 z_i^2(k)\\
&= \left\{1-\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\tau-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]\right\}^2\|\boldsymbol{w}(k)\|^2+H(k)
\end{aligned} \tag{24}$$

where

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

From (2), we obtain

$$\begin{aligned}
&\boldsymbol{w}(k+1)^T\boldsymbol{R}\boldsymbol{w}(k+1)\\
&= \sum_{i=1}^{n}\lambda_i\left\{1-\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\lambda_i-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]\right\}^2 z_i^2(k)\\
&= \sum_{i=1}^{n}\lambda_i\left\{1-\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\tau-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]\right\}^2 z_i\\
&\quad+\sum_{i=1}^{n-p}\lambda_i\left\{1-\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\lambda_i-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]\right\}^2 z_i^2(k)\\
&\quad-\sum_{i=1}^{n-p}\lambda_i\left\{1-\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\tau-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]\right\}^2 z_i^2(k)\\
&= \left\{1-\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\tau-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]\right\}^2\\
&\quad\times\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)+H'(k)
\end{aligned} \tag{25}$$

where

$$\begin{aligned}
H'(k) = \sum_{i=1}^{n-p}&\lambda_i\left\{\eta(2\|\boldsymbol{w}(k)\|^2-1)\left[\eta(\lambda_i+\tau)(2\|\boldsymbol{w}(k)\|^2-1)\right.\right.\\
&\left.\left.-2(1+\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k))\right]\right\}(\lambda_i-\tau)z_i^2(k).
\end{aligned}$$

From (24) and (25), it holds that

$$\begin{aligned}
&\boldsymbol{w}^T(k+1)\boldsymbol{R}\boldsymbol{w}(k+1)-(2\|\boldsymbol{w}(k+1)\|^2-1)\tau\\
&= \left\{1-\left[\eta^2\left((2\|\boldsymbol{w}(k)\|^2-1)\tau-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right)-2\eta\right]\right.\\
&\quad\left.\times(\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)-2\tau\|\boldsymbol{w}(k)\|^2)\right\}\\
&\quad\times\left[\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)-(2\|\boldsymbol{w}(k)\|^2-1)\tau\right]+H'(k)-2\tau H(k).
\end{aligned} \tag{26}$$

Denote $V(k) = \left|\boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k)-(2\|\boldsymbol{w}(k)\|^2-1)\tau\right|$. Thus, it holds that

$$\begin{aligned}
V(k+1) \le V(k)&\left|\left\{1-\eta\left[\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)-2\tau\|\boldsymbol{w}(k)\|^2\right]\right.\right.\\
&\left.\left.\times\left[\eta\left((2\|\boldsymbol{w}(k)\|^2-1)\tau-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right)-2\right]\right\}\right|\\
&+\left|H'(k)-2\tau H(k)\right|.
\end{aligned} \tag{27}$$

Denote

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

Through the use of Theorem 1 and $\eta\lambda_1 \le 0.25$, it stands valid that

$$\begin{aligned}
&\eta\left\{\eta\left[(2\|\boldsymbol{w}(k)\|^2-1)\tau-\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right]-2\right\}\\
&\quad\times\left[\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)-2\tau\|\boldsymbol{w}(k)\|^2\right]\\
&=\eta\Big[2\eta\tau\|\boldsymbol{w}(k)\|^2\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)+2\eta\tau^2\|\boldsymbol{w}(k)\|^2\\
&\quad+4\tau\|\boldsymbol{w}(k)\|^2-\eta\tau\boldsymbol{w}(k)^T\\
&\boldsymbol{R}\boldsymbol{w}(k)-\eta\left(\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)\right)^2-2\boldsymbol{w}(k)^T\boldsymbol{R}\boldsymbol{w}(k)-4\eta\tau^2\|\boldsymbol{w}(k)\|^4\Big]\\
&<\eta\left[2\eta{\lambda_1}^2\|\boldsymbol{w}(k)\|^4+2\eta\tau^2\|\boldsymbol{w}(k)\|^2+4\tau\|\boldsymbol{w}(k)\|^2-\eta\tau^2\|\boldsymbol{w}(k)\|^2\right.\\
&\quad\left.-\eta\tau^2\|\boldsymbol{w}(k)\|^4-2\tau\|\boldsymbol{w}(k)\|^2-4\eta\tau^2\|\boldsymbol{w}(k)\|^4\right]\\
&<2\eta^2{\lambda_1}^2\|\boldsymbol{w}(k)\|^4+2\eta\lambda_1\|\boldsymbol{w}(k)\|^2\\
&\le 2\eta^2{\lambda_1}^2(1+2\eta\lambda_1)^4+2\eta\lambda_1(1+2\eta\lambda_1)^2\le 0.9141.
\end{aligned}\tag{28}$$

Therefore, if $0<\delta<1$, then it holds that $V(k+1)\le \delta V(k)+|H'(k)-2\tau H(k)|$ Because

$$\begin{aligned}
&\left|H'(k)-2\tau H(k)\right|\\
&=\left|\sum_{i=1}^{n-p}\lambda_i\left\{\eta(\lambda_i-\tau)(2\|\boldsymbol{w}(k)\|^2-1){z_i}^2(k)\right.\right.\\
&\quad\left.\times\left[\eta(\lambda_i+\tau)(2\|\boldsymbol{w}(k)\|^2-1)-2(1+\eta\boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k))\right]\right\}\\
&\quad-2\tau\sum_{i=1}^{n-p}\left\{\eta(\lambda_i-\tau)(2\|\boldsymbol{w}(k)\|^2-1){z_i}^2(k)\right.\\
&\quad\left.\left.\times\left[\eta(\lambda_i+\tau)(2\|\boldsymbol{w}(k)\|^2-1)-2(1+\eta\boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k))\right]\right\}\right|\\
&=\left|\sum_{i=1}^{n-p}\left\{\eta(\lambda_i-\tau)(2\|\boldsymbol{w}(k)\|^2-1)(\lambda_i-2\tau){z_i}^2(k)\right.\right.\\
&\quad\left.\left.\times\left[\eta(\lambda_i+\tau)(2\|\boldsymbol{w}(k)\|^2-1)-2(1+\eta\boldsymbol{w}^T(k)\boldsymbol{R}\boldsymbol{w}(k))\right]\right\}\right|\\
&<\eta\lambda_1(2\|\boldsymbol{w}(k)\|^2-1)\left(2\eta\lambda_1\|\boldsymbol{w}(k)\|^2-2\eta\lambda_1-2\right)|\lambda_i-2\tau|\sum_{i=1}^{n-p}{z_i}^2(k)\\
&\le\eta\lambda_1(2(1+2\eta\lambda_1)^2-1)\left(2\eta\lambda_1(1+2\eta\lambda_1)^2-2\eta\lambda_1-2\right)\\
&\quad\times|\lambda_i-2\tau|\sum_{i=1}^{n-p}{z_i}^2(k)\\
&\le\phi\Pi_1e^{-\theta_1k}
\end{aligned}\tag{29}$$

where

$$\phi=\eta\lambda_1(2(1+2\eta\lambda_1)^2-1)|\lambda_i-2\tau|$$
$$\times\left(2\eta\lambda_1(1+2\eta\lambda_1)^2-2\eta\lambda_1-2\right).$$

We have

$$\begin{aligned}
V(k+1)&\le\delta V(k)+\left|H'(k)-2\tau H(k)\right|\\
&\le\delta^{k+1}V(0)+\phi\Pi_1\sum_{r=0}^{k}(\delta e^{\theta_1})^re^{-\theta_1k}\\
&\le\delta^{k+1}V(0)+(k+1)\phi\Pi_1\max\left\{\delta^k,e^{-\theta_1k}\right\}\\
&\le(k+1)\Pi_2\left[e^{-\theta_2(k+1)}+\max\left\{e^{-\theta_2k},e^{-\theta_1k}\right\}\right]
\end{aligned}\tag{30}$$

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

$$\Pi_2=\max\left\{\boldsymbol{w}^T(0)\boldsymbol{R}\boldsymbol{w}(0)-(2\|\boldsymbol{w}(0)\|^2-1)\tau,\phi\Pi_1\right\}>0.$$

■

Appendix E
Proof of Lemma 3

Proof: Provided any $\varepsilon>0$, there is a $K\ge1$ such that

$$\frac{\Pi Ke^{-\theta K}}{(1-e^{-\theta})^2}\le\varepsilon.\tag{31}$$

With respect to any $k_1>k_2>K$, it holds that

$$\begin{aligned}
&|z_i(k_1)-z_i(k_2)|\\
&\quad=\left|\sum_{r=k_2}^{k_1-1}[z_i(r+1)-z_i(r)]\right|\\
&\quad=\sum_{r=k_2}^{k_1-1}\eta\left|\left[\boldsymbol{w}^T(r)\boldsymbol{R}\boldsymbol{w}(r)-(2\|\boldsymbol{w}(r)\|^2-1)\tau\right]z_i(r)\right|\\
&\quad\le\sum_{r=k_2}^{k_1-1}r\Pi e^{-\theta r}\le\Pi\sum_{r=K}^{+\infty}re^{-\theta r}\\
&\quad\le\Pi Ke^{-\theta K}\sum_{r=0}^{+\infty}r\left(e^{-\theta}\right)^{r-1}\\
&\quad\le\frac{\Pi Ke^{-\theta K}}{(1-e^{-\theta})^2}\le\varepsilon.
\end{aligned}\tag{32}$$

As suggested by (32), we can conclude that $\{z_i(k)\}$ is considered to be a Cauchy order. Through the application of the Cauchy convergence principle, it holds that $\lim_{k\to\infty}z_i(k)={z_i}^*$, $i=n-p+1,n-p+2,\ldots,n$, where ${z_i}^*$ is a constant. ■

References

[1] B. Y. Ye, M. Cirrincione, M. Pucci, and G. Cirrincione, "Speed sensorless control of induction motors based on MCA EXIN Pisarenko method," in *Proc. IEEE Conf. Energy Conversion Congress and Exposition*, Sep. 2015, pp. 2176–2183.

[2] Y. B. Gao, X. Y. Kong, H. H. Zhang, and L. A. Hou, "A weighted information criterion for multiple minor components and its adaptive extraction algorithms," *Neural Networks*, vol. 89, no. 5, pp. 1–10, 2017.

[3] Y. D. Jou, C. M. Sun, and F. K. Chen, "Eigenfilter design of FIR digital filters using minor component analysis," in *Proc. 9th Int. Conf. Communications and Signal Processing*, Dec. 2013, pp. 1–5.

[4] R. Wang, M. L. Yao, D. M. Zhang, and H. X. Zou, "A novel orthonormalization matrix based fast and stable DPM algorithm for principal and minor subspace tracking," *IEEE Trans. Signal Processing*, vol. 60, no. 1, pp. 466–472, 2011.

[5] R. Möller, "A self-stabilizing learning rule for minor component analysis," *Int. J. Neural Systems*, vol. 14, no. 1, pp. 1–8, 2004.

[6] Y. H. Shao, N. Y. Deng, W. J. Chen, and Z. Wang, "Improved generalized eigenvalue proximal support vector machine," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 213–216, 2013.

[7] M. Thameri, K. Abed-Meraim, and A. Belouchrani, "Low complexity adaptive algorithms for principal and minor component analysis," *Digital Signal Processing*, vol. 23, pp. 19–29, 2013.

[8] T. D. Nguyen and I. Yamada, "A unified convergence analysis of normalized PAST algorithms for estimating principal and minor components," *Signal Processing*, vol. 93, pp. 176–184, 2013.

[9] P. J. Zufiria, "On the discrete-time dynamics of the basic Hebbian neural-network node," *IEEE Trans. Neural Networks*, vol. 13, no. 6, pp. 1342–1352, 2002.

[10] D. Z. Peng, Y. Zhang, Y. Xiang, and H. X. Zhang, "A globally convergent MC algorithm with an adaptive learning rate," *IEEE Trans. Neural Networks and Learning Systems*, vol. 23, no. 2, pp. 359–365, 2012.

[11] T. D. Nguyen and I. Yamada, "Necessary and sufficient conditions for convergence of the DDT systems of the normalized PAST algorithms," *Signal Processing*, vol. 94, pp. 288–299, 2014.

[12] Y. B. Gao, X. Y. Kong, C. H. Hu, and L. A. Hou, "Modified Möller algorithm for multiple minor components extraction," *Control and Decision*, vol. 32, no. 3, pp. 493–497, 2017.

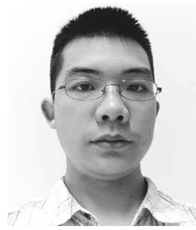
[13] R. Möller, *Derivation of Coupled PCA and SVD Learning Rules from a Newton Zero-finding Framewor*. Computer Engineering, Faculty of Technology, Bielefeld University, 2017.

[14] G. ORegan, *Matrix Theory*. Switzerland: Springer International Publishing, 2016.

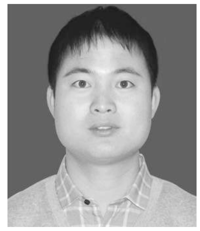
[15] K. L. Du, and M. N. S. Swamy, *Principal Component Analysis*. Springer London, 2014.

[16] Y. Zhang, Y. Mao, L. 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*, vol. 16, no. 6, pp. 1318–1328, 2005.

[17] X. Y. Kong, C. H. Hu, and Z. S. Duan, *Principal Component Analysis Networks and Algorithms*. Springer Singapore, 2017.



Haidi Dong received the bachelor, master, and Ph.D. degrees from Xi'an Research Institute of High Technology, in 2011, 2014, and 2018, respectively, all in control science and engineering. He is currently a Lecturer at Naval University of Engineering. His current research interests include adaptive signal processing and feature extract.



Yingbin Gao received the bachelor, master, and Ph.D. degrees from Xi'an Research Institute of High Technology, in 2009, 2011, and 2016, respectively, all in control science and engineering. He is currently an Engineer with the China Electronics Technology Group Corp 54th Research Institute, China. His research interests include adaptive signal processing and feature extraction.

Gang Liu received the bachelor and master degrees from Xi'an Research Institute of High Technology, in 1988 and 1991, respectively, and the Ph.D. degree from Northwestern Polytechnical University, in 1998. He is currently a Professor with Xi'an Research Institute of High Technology. His research interests include adaptive signal processing, system modeling, and fault diagnosis.