

Performance analysis of distributed adaptive filters

CHEN CHEN, ZHIXIN LIU AND LEI GUO

This paper investigates the tracking performance of the normalized least mean square (LMS) based distributed adaptive filters, where a set of filters are designed to estimate the unknown time-varying parameters or signals using noisy measurements in a cooperative way. We show that under a general connected topology, the tracking error covariances of the distributed filters can be approximately described and calculated by a simple, linear and deterministic matrix difference equation. Different from most of the existing results, we do not require the regression vectors to satisfy stationarity or independency assumptions, which makes our theory applicable to stochastic systems with feedback.

1. Introduction

With the rapid development of the miniature sensing technology and the communication technology, sensor networks have exhibited the advantages of flexibility, fault tolerance, and ease of deployment [1], and thus, have wide applications in many practical situations, such as industrial process monitoring and control, machine health monitoring, and so on. Parameter estimation or system identification is one of the most important topics in system and control areas, and the least mean square (LMS) method is a basic algorithm to track a sequence of unknown time-varying parameters; See [2]–[4] for algorithms and theoretical analysis. However, for many complex systems, a single sensor may not be able to accomplish the estimation or filtering tasks because of its limited sensing and observation capability. It is natural to cooperate multiple sensors to design the estimation algorithms. Limited by the communication ability, the distributed adaptive filters where the sensors use the local observations to design the estimation algorithms

* Research supported in part by the National Natural Science Foundation of China under grant 61273221, and the National Key Basic Research Program of China (973 program) under grant 2014CB845302.

are widely adopted. In comparison with the centralized algorithms where a fusion center is required to receive and process the information from all sensors, the distributed adaptive filters have some advantages, e. g., reduction of the communication and computation costs, robustness to partial node failures.

Over the last decade, a lot of efforts have been paid on the design and performance analysis of the distributed filtering algorithms. Since the consensus scheme usually embodies the cooperation of the sensors over the network, the consensus-based algorithms are widely used. For the case where the regression vectors used in the filtering algorithms are deterministic, some theoretical analyses have been carried out, see, e.g., [5]–[8]. For the case of stochastic regression vectors, some theoretical results are also obtained, see [9]–[13]. However, in almost all existing results, the stationarity and independency assumptions of the regression vectors are used.

It is worth pointing out that the stationarity and independency assumption are idealizations of the regression vectors for the complicated practical systems. For feedback closed-loop stochastic control systems (cf., [14]), the stationarity and independency assumptions of regression vectors are not satisfied because the input and output constitute the regression vectors, and the outputs are usually generated from a set of highly nonlinear and stochastic equations, and thus, are correlated.

In this paper, we focus on the normalized diffusion least mean-square algorithm (DLMS) introduced in [10], and provide the performance approximation of the algorithms without imposing stationary and independent assumptions on the stochastic regression vectors. The difficulty lies in how to deal with the product of random matrices which are temporally non-commutative, non-independent and spatially coupled. Under a general connected topology, we establish the performance approximation formula of the tracking error covariances of the normalized LMS based distributed filters, which can be approximately described and calculated by a simple, linear and deterministic matrix difference equation. Our theory shows that the distributed filters can work well cooperatively, even if any individual sensor does not have the ability of tracking due to lack of necessary information. This desirable property for the distributed filters does not seem to have been established except the authors' companion work of [15].

The remainder of this article is organized as follows. In Section 2, we introduce the DLMS algorithm as well as the main results concerning the approximation of the tracking performance. In Section 3, we provide the proof of the main theorems. The concluding remarks are made in Section 4.

2. Problem statement and main results

Let us first introduce some notations used throughout the paper. For a matrix A , $A \geq 0$ means that A is semi-positive definite, and $A \geq B$ means that $A - B \geq 0$. The notations $(A)^\tau$, $\lambda_{max}(A)$ and $\lambda_{min}(A)$, respectively, denote the transpose, the largest eigenvalue and the smallest eigenvalue of the matrix A . The notation \otimes denotes the Kronecker product of two matrices. I_n denotes the identity matrix with order n . The operator $diag(c_1, c_2, \dots, c_n)$ denotes the diagonal matrix with diagonal elements c_1, c_2, \dots, c_n . The notation $col(\dots)$ stands for a vector by stacking the specified vectors. For a vector x , $\|x\|$ represents the Euclidean norm of x , and for a matrix A , its norm is defined as $\|A\| = \{\lambda_{max}(AA^\tau)\}^{\frac{1}{2}}$. We refer to $\|A\|_{L_p}$ defined by $\|A\|_{L_p} \triangleq \{E\|A\|^p\}^{\frac{1}{p}}$ as the L_p -norm of a random matrix A , where E is the mathematical expectation on the basic probability space (Ω, \mathcal{F}, P) . Given a matrix sequence $\{A_k, k = 1, 2, \dots\}$ and a positive scalar sequence $\{b_k, k = 1, 2, \dots\}$, by $A_k = O(b_k)$ we mean that there exists a positive constant M free of k such that for any $k \geq 0$, we have $\|A_k\| \leq Mb_k$.

2.1. Diffusion least mean square (DLMS) algorithm

Consider a network composed of n sensors. Each sensor is aimed at estimating a sequence of m -dimensional unknown time-varying parameter process $\{\theta_k\}$. In this paper, we assume that the observation of each sensor obeys the following stochastic linear regression model,

$$(1) \quad y_k^i = (\varphi_k^i)^\tau \theta_k + v_k^i, \quad k \geq 1, \quad i = 1, 2, \dots, n$$

where y_k^i and v_k^i are, respectively, scalar observation and noise signal of the sensor i at discrete-time instant k , and φ_k^i is an m -dimensional stochastic regression vector. In order to give a performance analysis of the filtering algorithm to be given in this paper, it is natural to introduce some assumptions on the unknown parameters. In general, we may write θ_k as the following equation,

$$(2) \quad \theta_k = \theta_{k-1} + \gamma \omega_k,$$

where γ is a scaling constant and ω_k is an unknown vector.

The communications between sensors are represented by a weighted graph $\mathcal{G} = \{V, \mathcal{E}, A\}$, where V is the set of sensors, $\mathcal{E} \subseteq V \times V$ is the edge set, and the matrix $A = [a_{ij}]_{n \times n}$ defines the communication weights between

sensors where $a_{ij} > 0$ if $(i, j) \in \mathcal{E}$ and $a_{ij} = 0$ otherwise. In this paper, we assume that the elements of the weighted matrix A satisfy $a_{ij} = a_{ji}$ for any i and j , and $\sum_{j=1}^n a_{ij} = 1$ for any i . For a sensor i , it can only communicate with its neighbors $\mathcal{N}_i = \{j : (i, j) \in \mathcal{E}\}$. The stochastic matrix provide some nice properties for the performance analysis of the distributed filtering algorithms. A matrix is called stochastic if all elements are non-negative and the sum of each row equals to 1. Furthermore, it is called doubly stochastic, if it is a stochastic matrix and the sum of each column also equals to 1. A stochastic matrix M is called ergodic if $\lim_{t \rightarrow \infty} M^t$ exists and all rows of $\lim_{t \rightarrow \infty} M^t$ are the same. It is clear that the weighted matrix A is doubly stochastic.

Denote the estimate of the sensor i at time k as $\hat{\theta}_k^i$. In this paper, we adopt the following normalized DLMS algorithm where the Combine-then-Adapt diffusion strategy is used to estimate the unknown parameter θ_k ,

$$(3) \quad \begin{cases} \hat{\theta}_{k+1}^i = \vartheta_k^i + \mu_i \frac{\varphi_k^i}{1 + \|\varphi_k^i\|^2} (y_k^i - (\varphi_k^i)^\tau \vartheta_k^i), \\ \vartheta_k^i = \sum_{j=1}^n a_{ij} \hat{\theta}_k^j, \end{cases}$$

where $\mu_i \in (0, 1)$ is a step-size, and the initial estimates $\hat{\theta}_0^i$ ($i = 1, 2, \dots, n$) can be chosen arbitrarily.

The DLMS algorithm (3) without normalization appears to be first introduced in [10]. The advantage of the DLMS over the centralized algorithms is discussed in [12]. The stability theorem for the above DLMS algorithm has been established, and an upper bound for tracking error covariance has also been given; see [15].

The objective of this paper is to refine the performance analysis of the above DLMS given in [15], by providing a performance approximation formula for the estimation error covariance under a class of general regression signals. In order to proceed with our analysis, we introduce the following quantities which are similar to those used in [15],

$$\begin{aligned} \Theta_k &\triangleq \text{col}\{\underbrace{\theta_k^1, \dots, \theta_k^n}_n\}, & W_k &\triangleq \text{col}\{\underbrace{\omega_k^1, \dots, \omega_k^n}_n\}, \\ Y_k &\triangleq \text{col}\{y_k^1, \dots, y_k^n\}, & \vartheta_k &\triangleq \text{col}\{\vartheta_k^1, \dots, \vartheta_k^n\}, \\ \Psi_k &\triangleq \text{diag}\{\varphi_k^1, \dots, \varphi_k^n\}, & \hat{\Theta}_k &\triangleq \text{col}\{\hat{\theta}_k^1, \dots, \hat{\theta}_k^n\}, \\ \tilde{\Theta}_k &\triangleq \text{col}\{\tilde{\theta}_k^1, \dots, \tilde{\theta}_k^n\} & \text{with } \tilde{\theta}_k^i &= \hat{\theta}_k^i - \theta_k^i \\ V_k &= \text{col}\{v_k^1, \dots, v_k^n\}, \end{aligned}$$

$$L_k \triangleq \text{diag} \left\{ \frac{\varphi_k^1}{1 + \|\varphi_k^1\|^2}, \dots, \frac{\varphi_k^n}{1 + \|\varphi_k^n\|^2} \right\},$$

$$\Lambda \triangleq \text{diag}\{\mu_1 I_m, \dots, \mu_n I_m\}, \quad \mathbf{A} = A \otimes I_m.$$

Using the Kronecker product \otimes , the Equation (3) can be written as

$$(4) \quad \hat{\Theta}_{k+1} = \mathbf{A}\hat{\Theta}_k + \Lambda L_k(Y_k - \Psi_k^T \mathbf{A}\hat{\Theta}_k),$$

where by (1), the observation Y_k evolves according to the following equation

$$(5) \quad Y_k = \Psi_k^T \Theta_k + V_k.$$

Subtracting Θ_k from both sides of (4) and using the fact that $\mathbf{A}\Theta_k = \Theta_k$, we obtain that the estimation error evolves according to the following equation,

$$(6) \quad \tilde{\Theta}_{k+1} = (I_{mn} - \Lambda F_k)\mathbf{A}\tilde{\Theta}_k + \Lambda L_k V_k - \gamma W_{k+1},$$

where $F_k = L_k \Psi_k^T$.

2.2. Definitions and assumptions

In order to analyze the tracking performance of Θ_k , some definitions and assumptions on the regression signals and observation noises are introduced, see [16]- [19] for more details. Throughout the sequel, the notation \mathcal{F}_k is employed to denote the σ -algebra generated by $\{\varphi_i^j, \omega_i, v_{i-1}^j, j = 1, \dots, n, i \leq k\}$.

The exponential stability of the homogeneous part of the error equation (6) plays a key role for our analysis. We first present the definitions for exponential stability of random matrices.

Definition 1. A sequence of $d \times d$ random matrices $A = \{A_k, k \geq 0\}$ is called L_p -exponentially stable ($p > 0$) with parameter $\lambda \in [0, 1)$, if it belongs to the following set

$$S_p(\lambda) = \left\{ A : \left\| \prod_{j=i+1}^k A_j \right\|_{L_p} \leq M \lambda^{k-i}, \forall k \geq i \geq 0, \text{ for some } M > 0 \right\}.$$

Definition 2. A scalar random sequence $a = \{a_i, i \geq 0\}$ is called stably exciting if it belongs to the following set:

$$S^o(\lambda) = \left\{ a_i \in [0, 1] : E \prod_{j=i+1}^k (1 - a_j) \leq M\lambda^{k-i}, \right. \\ \left. \forall k \geq i \geq 0, \text{ for some } M > 0 \right\}.$$

We may denote

$$S^o \triangleq \bigcup_{\lambda \in (0,1)} S^o(\lambda).$$

Definition 3. We say that a random sequence $x \triangleq \{x_k\} \in \mathcal{M}_p (p \geq 1)$ if there exists a constant $C_p(x)$ only depending on p and $\{x_k\}$ such that for any $j \geq 0$,

$$(7) \quad \left\| \sum_{i=j+1}^{j+h} x_i \right\|_{L_p} \leq C_p(x) h^{\frac{1}{2}}, \quad \forall h \geq 1, .$$

Remark 4. The widely used random sequences in the investigation of estimation or filtering algorithms, such as martingale difference sequence, the ϕ -mixing and α -mixing sequences, and the linear process driven by white noises, all belong to the set \mathcal{M}_p ; See [16] for details.

Assumption 5. (*Connectivity*) The graph \mathcal{G} is undirected and connected, and contains a self-loop at each vertex, i.e., $i \in \mathcal{N}_i$ for any $i \in V$.

By Assumption 5, it is easy to see that the weighted matrix A is ergodic. Furthermore, by the symmetry and stochastic property of the matrix A , we see that A has n real eigenvalues and 1 is one of the eigenvalues. Thus, we can arrange the eigenvalues of A in a non-decreasing order $-1 < \lambda_n(A) \leq \lambda_{n-1}(A) \leq \dots \leq \lambda_2(A) < \lambda_1(A) = 1$. Denote $\lambda_{gap} = \max\{|\lambda_2(A)|, |\lambda_n(A)|\} < 1$.

Assumption 6. (*Cooperative Excitation Condition*) Let $\{\varphi_k^i, \mathcal{F}_k, k \geq 0\}$, $i = 1, \dots, n$ be n adapted sequences, and $\{\rho_k, k \geq 0\} \in S^o(\lambda)$ for some $\lambda \in$

$(0, 1)$, where

$$(8) \quad \rho_k \triangleq \lambda_{\min} \left\{ E \left[\frac{1 - \lambda_{gap}}{n} \sum_{i=1}^n \frac{\varphi_{k+1}^i (\varphi_{k+1}^i)^\tau}{1 + \|\varphi_{k+1}^i\|^2} \mid \mathcal{F}_k \right] \right\},$$

and $\lambda_{gap} \triangleq \max\{|\lambda_2(A)|, |\lambda_n(A)|\}$ with A being the weighted matrix.

Remark 7. In order to analyze the tracking performance of Θ_k , some assumptions on the regressors are needed. It is clear that Assumption 6 is more practical and more general than the previously used stationarity and independency assumptions; See [15] for more details. In this paper, we aim at providing an asymptotic formula for the estimation error covariance when the regression signals satisfy Assumption 6.

Assumption 8. For some $p \geq 1$, we have $\|\tilde{\Theta}_0\|_{L_{2p}} < \infty$, $\{L_k V_k\} \in \mathcal{M}_{2p}$ and $\{W_k\} \in \mathcal{M}_{2p}$.

Under Assumptions 5, 6 and 8, we have derived an upper bound for the tracking performance of the DLMS algorithm; See [15]. To get more accurate results on the estimation error covariance than just upper bounds, we need more assumptions on the regression signals, observation noises, and the parameter drift.

Assumption 9. (Weak dependence) There is a bounded function $\phi(\tilde{m}, \mu) \geq 0$ with

$$\lim_{\tilde{m} \rightarrow \infty, \mu \rightarrow 0} \phi(\tilde{m}, \mu) \log \frac{1}{\mu} = 0,$$

such that for any \tilde{m}, k and $\mu \in (0, 1)$,

$$\|E[F_k | \mathcal{F}_{k-\tilde{m}}] - E[F_k]\|_{L_4} \leq \phi(\tilde{m}, \mu),$$

where $\{F_k\}$ is defined in (6).

Assumption 10. For all $k \geq 1$ and $r \geq 1$, there exists a constant $M > 0$ such that

$$\begin{aligned} E[V_k | \mathcal{F}_k] &= 0, & E[W_{k+1} | \mathcal{F}_k] &= 0, & E[W_{k+1} V_k^\tau | \mathcal{F}_k] &= 0, \\ E[V_k V_k^\tau | \mathcal{F}_k] &= R_v(k) \geq 0, & E[W_{k+1} W_{k+1}^\tau] &= Q_\omega(k+1) \geq 0, \\ \sup_{k \geq 0} (\|V_k\|_{L_8} + \|W_k\|_{L_8}) &\leq M, \end{aligned}$$

where $\{V_k\}$ and $\{W_k\}$ are defined in (6).

Remark 11. Assumption 10 requires that the measurement noise V_k and the parameter drift W_k are of white noise character. We note that the martingale difference property of W_k essentially means that the true parameters are assumed to be a random walk. As pointed out in [16], in the context of slow adaptation (small μ), a random walk model captures the tracking behavior of the algorithm quite well. Furthermore, we note that time-varying covariances $Q_\omega(k)$ and $R_v(k)$ are allowed, which may cover some special model drifts of interests. Finally, we remark that higher order moments are assumed, since more accurate performance than upper bounds are to be obtained in the paper (see Theorem 13 in the following).

2.3. The main theorems

Under some conditions on the regression vectors and measurement noises, a preliminary performance analysis is given for the estimation error covariance.

Lemma 12. [15] *Assume that $\Lambda = \mu I_{mn}$, where $\mu \in (0, 1/e)$, and that Assumptions 5-8 are satisfied. Then we have for all small $\mu > 0$,*

$$(9) \quad \|\tilde{\Theta}_{k+1}\|_{L_p} = O\left(\left[\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right] \log \frac{1}{\mu} + (1 - \mu\alpha)^k \|\tilde{\Theta}_0\|_{L_{2p}}\right).$$

The above theorem provides an upper bound for the estimation error covariance. Based on this theorem and following the ideas of [16], we further proceed to give an asymptotic formula for the tracking error covariance. To this end, we recursively define a linear and deterministic difference equation as follows:

$$(10) \quad \begin{aligned} \hat{\Pi}_{k+1} &= (I_{mn} - \mu E[F_k]) \mathbf{A} \hat{\Pi}_k \mathbf{A} (I_{mn} - \mu E[F_k])^\tau \\ &\quad + \mu^2 R_{v,\varphi}(k) + \gamma^2 Q_\omega(k+1), \end{aligned}$$

where $\hat{\Pi}_0 = E[\tilde{\Theta}_0 \tilde{\Theta}_0^\tau]$, $R_{v,\varphi}(k) \triangleq E[L_k V_k V_k^\tau L_k^\tau]$ and $Q_\omega(k+1)$ is defined in Assumption 10. The deterministic matrix sequence $\{\hat{\Pi}_k\}$ is used to approximate $\Pi_k \triangleq E[\tilde{\Theta}_k \tilde{\Theta}_k^\tau]$.

Theorem 13. *Suppose that $\Lambda = \mu I_{mn}$, and that Assumptions 5-6 and 9-10 hold. Then for all small $\mu \in (0, 1/e)$ and $k \geq 1$, we have*

$$\|E[\tilde{\Theta}_{k+1} \tilde{\Theta}_{k+1}^\tau] - \hat{\Pi}_{k+1}\| \leq c\sigma(\mu) \left[\mu + \frac{\gamma^2}{\mu} + (1 - \alpha\mu)^k \right],$$

where $c > 0$ and $\alpha \in (0, 1)$ are constants, and $\sigma(\mu)$ is defined by

$$(11) \quad \sigma(\mu) \triangleq \min_{\tilde{m} \geq 1} \left\{ \tilde{m} \sqrt{\mu} \log^3 \frac{1}{\mu} + \phi(\tilde{m}, \mu) \log \frac{1}{\mu} \right\}.$$

which tends to zero as μ approaches to zero.

The proof of Theorem 13 is given in Section 3.

Remark 14. Theorem 13 implies that in a certain sense, the tracking error covariance of DLMS can be approximated by the solution of a simple linear deterministic difference equation (10), and the degree of the approximation can be characterized by $\sigma(\mu)$: the faster it tends to zero, the better the approximation can be obtained.

In general, the “dominant term” of $\{E[\tilde{\Theta}_k \tilde{\Theta}_k^T]\}$, i.e., Π_k may not be easily comprehended and analyzed. However, under some further conditions on the signals and noises, we can obtain a nice expression of $\{E[\tilde{\Theta}_k \tilde{\Theta}_k^T]\}$; See the following theorem.

Theorem 15. *If in addition to the conditions in Theorem 13, the following weak stationary conditions hold,*

$$\begin{aligned} R_{v,\varphi}(k) &\equiv R_{v,\varphi}, & R_\omega(k) &\equiv R_\omega, \\ F &= E[F_k] = \text{diag}(F^1, \dots, F^n) \text{ with } F^i = F^j, \end{aligned}$$

then we have for all small $\mu \in (0, 1/e)$,

$$(12) \quad E[\tilde{\Theta}_{k+1} \tilde{\Theta}_{k+1}^T] = \mu \bar{R}_{v,\varphi} + \frac{\gamma^2}{\mu} \bar{R}_\omega + O\left(\sigma(\mu) \left(\mu + \frac{\gamma^2}{\mu}\right)\right) + o(1).$$

where the term $o(1)$ tends to 0 exponentially fast as $k \rightarrow \infty$, and $\bar{R}_{v,\varphi}, \bar{R}_\omega$ are two constant matrices depending on $F, R_{v,\varphi}$ and R_ω . The explicit expressions can be found in Section 3.

Remark 16. Note that $\sigma(\mu)$ tends to zero as μ approaches to zero. Thus, it follows from (12) that for all small μ and large k

$$E[\tilde{\Theta}_{k+1} \tilde{\Theta}_{k+1}^T] \sim \mu \bar{R}_{v,\varphi} + \frac{\gamma^2}{\mu} \bar{R}_\omega.$$

Consequently, the “best” choice for μ is a tradeoff between the noise sensitivity and the parameter variation. In particular, we can choose $\mu = \mu^* =$

$\gamma\sqrt{\text{tr}\bar{R}_\omega/\text{tr}\bar{R}_{v,\varphi}}$, which minimizes the trace of the right-hand-side, and gives

$$\sum_{i=1}^n E\|\tilde{\theta}_k^i\|^2 \sim \mu\text{tr}(\bar{R}_{v,\varphi}) + \frac{\gamma^2}{\mu}\text{tr}(\bar{R}_\omega).$$

3. Proof of Theorems 13 and 15.

In order to prove Theorems 13 and 15, we first present some preliminary results.

The following lemma plays a key role in the stability analysis of the homogeneous part of the DLMS algorithm, and will be used in the proof of the main theorems in this paper.

Lemma 17. [15] *Suppose that Assumption 5 holds, and that $\{\Phi_k = (\Phi_k^{ij}) \in \mathbb{R}^{m \times m}, k = 1, \dots, n\}$ is a sequence of symmetric matrices satisfying $0 \leq \Phi_k \leq I_m, k = 1 \dots n$. Then we have*

$$\lambda_{max}[\mathbf{A} \cdot \text{diag}(I_m - \Phi_1, \dots, I_m - \Phi_n) \cdot \mathbf{A}] \leq 1 - (1 - \lambda_{gap})\delta,$$

where $\delta = \frac{1}{n}\lambda_{min}(\Phi_1 + \dots + \Phi_n)$.

In order to prove Theorem 13, we define a new sequence $\bar{\Theta}_k = \text{col}\{\bar{\theta}_k^1, \dots, \bar{\theta}_k^n\}$ by

$$(13) \quad \bar{\Theta}_{k+1} = (I_{mn} - \mu E[F_k])\mathbf{A}\bar{\Theta}_k + \mu L_k V_k - \gamma W_{k+1},$$

with $\bar{\Theta}_0 = \tilde{\Theta}_0$. It is clear that $E[\bar{\Theta}_k \bar{\Theta}_k^T] = \hat{\Pi}_k$, which is recursively defined by Equation (10). We will show that the homogeneous part of (13) is exponentially stable; See the following lemma.

Lemma 18. *Let Assumptions 5 and 6 hold, then there exist positive constants M and $\beta \in (0, 1)$ only depending on $\{F_k\}$, such that for any $\mu \in (0, 1/e)$ and any $k \geq i \geq 0$, we have*

$$(14) \quad \left\| \prod_{j=i+1}^k (I_{mn} - \mu E[F_j])\mathbf{A} \right\| \leq M(1 - \mu\beta)^{k-i}.$$

To prove Lemma 18, we first introduce a lemma.

Lemma 19. [19] For a sequence of scalars $\{a_i \in (0, 1), i \geq 1\}$, the following two assertions are equivalent:

i) There exist positive constants M and $\lambda \in (0, 1)$ such that for any i and j , $\prod_{k=i+1}^j (1 - a_k) \leq M\lambda^{j-i}$.

ii) There exists an integer $h > 0$ such that

$$(15) \quad \inf_j \sum_{i=j+1}^{j+h} a_i \neq 0.$$

Proof of Lemma 18. Denote

$$(16) \quad A_k \triangleq E \left[(1 - \lambda_{gap}) \cdot \frac{1}{n} \sum_{i=1}^n \frac{\varphi_{k+1}^i (\varphi_{k+1}^i)^\tau}{1 + \|\varphi_{k+1}^i\|^2} \middle| \mathcal{F}_k \right].$$

By the definition of ρ_k , we have $\rho_k = \lambda_{min}(A_k) \in (0, 1)$. Using Assumption 6, we know that there exist positive constants M and $\lambda \in (0, 1)$ such that

$$(17) \quad E \prod_{j=i+1}^k (1 - \rho_j) \leq M\lambda^{k-i}.$$

Choosing an integer h_0 large enough such that the left hand side of (17) is smaller than $\frac{1}{2}$, i.e., for any $i \geq 0$,

$$(18) \quad \begin{aligned} & 1 - (2^{h_0} - 1) \left(\sum_{j=i+1}^{i+h_0} E\rho_j \right) \\ & < 1 - \sum_{j=i+1}^{i+h_0} E[\rho_j] + \sum_{i+1 \leq j_1, j_2 \leq i+h_0} E[\rho_{j_1} \rho_{j_2}] \\ & \quad + \dots + (-1)^{h_0} E[\rho_{i+1} \dots \rho_{i+h_0}] < \frac{1}{2}, \end{aligned}$$

where the fact that $\rho_j \in (0, 1)$ is used. By (18), we have

$$(19) \quad \sum_{j=i+1}^{i+h_0} E\rho_j > \sigma, \forall i \geq 0,$$

where $\sigma = 1/(2(2^{h_0} - 1)) > 0$. Note that

$$(20) \quad \lambda_{min}(E(A_j)) \geq E\lambda_{min}(A_j),$$

Combining this with (19), we have

$$\sum_{j=i+1}^{i+h_0} \lambda_{\min}(E(A_j)) > \sigma.$$

By Lemma 17, we have

$$(21) \quad \left\| \prod_{j=i+1}^k (I_{mn} - \mu E[F_j]) \mathbf{A} \right\| \leq \prod_{j=i+1}^k (1 - \mu \lambda_{\min}(EA_{j-1}))^{1/2}.$$

By (21) and Lemma 19, we see that there exist positive constants M and $\beta \in (0, 1)$ such that (14) holds. This completes the proof of the lemma. \square

In the the following lemma, we provide an estimation for the upper bound of the L_2 -norm of $\bar{\Theta}_k$ defined by (13).

Lemma 20. *Let Assumptions 5, 6 and 10 be satisfied. Then for $\mu \in (0, 1/e)$,*

$$(22) \quad \|\bar{\Theta}_{k+1}\|_{L_2} = O\left(\left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right) \log \frac{1}{\mu} + (1 - \alpha\mu)^k\right),$$

where $\alpha \in (0, 1)$ is a constant only depending on $\{\mathcal{F}_k, k \geq 0\}$.

The proof of Lemma 20 is similar to that of Theorem 2.2 given in [15]. We omit it to save space.

The following property concerning the contraction property of the ergodic matrix will be used in the proof of Theorem 13.

Lemma 21. *Assume that $P = [P_{ij}] \in \mathbb{R}^{n \times n}$ is an ergodic matrix, and $Y = \text{col}\{y_1, \dots, y_n\} \in \mathbb{R}^{nh}$ with y_i being an h -dimensional vector for $1 \leq i \leq n$. If $Z \triangleq \text{col}\{z_1, \dots, z_n\} = (P \otimes I_h)Y$ with $z_i \in \mathbb{R}^h$, then*

$$(23) \quad \max_{s,s'} \|z_s - z_{s'}\| \leq \tau(P) \max_{j,j'} \|y_j - y_{j'}\|,$$

where $0 < \tau(P) < 1$ is a constant only depending on P .

The above lemma can be obtained by following the proof of Theorem 3.1 in [20], and the details are omitted.

By the above lemma, we can derive the following result.

Lemma 22. *Let Assumptions 5, 6 and 10 be satisfied. Then for $\mu \in (0, 1/e)$ and $k \geq 0$,*

$$(24) \quad \max_{i,j} \|\tilde{\theta}_{k+1}^i - \tilde{\theta}_{k+1}^j\|_{L_2} = O\left(\mu + \sqrt{\mu}\gamma \log \frac{1}{\mu}\right) + O(\tau^{k+1}(A)),$$

where $\tau(A) \in (0, 1)$ is determined by the matrix A .

Proof. By (6), we have for $i \neq j$,

$$(25) \quad \tilde{\theta}_{k+1}^i - \tilde{\theta}_{k+1}^j = \sum_{l=1}^n a_{il}\tilde{\theta}_k^l - \sum_{l=1}^n a_{jl}\tilde{\theta}_k^l + O(\mu\|\tilde{\Theta}_k\| + \mu\|L_k V_k\|).$$

Using Lemma 21, we have

$$(26) \quad \max_{i,j} \left\| \sum_{l=1}^n a_{il}\tilde{\theta}_k^l - \sum_{l=1}^n a_{jl}\tilde{\theta}_k^l \right\| \leq \tau(A) \max_{i,j} \|\tilde{\theta}_k^i - \tilde{\theta}_k^j\|.$$

Thus, we have

$$\begin{aligned} \max_{i,j} \|\tilde{\theta}_{k+1}^i - \tilde{\theta}_{k+1}^j\| &\leq \tau(A) \max_{i,j} \|\tilde{\theta}_k^i - \tilde{\theta}_k^j\| + O(\mu\|\tilde{\Theta}_k\| + \mu\|L_k V_k\|) \\ &= \tau^{k+1}(A) \max_{i,j} \|\tilde{\theta}_0^i - \tilde{\theta}_0^j\| \\ &\quad + \sum_{i=0}^k \tau^{k-i}(A) O(\mu\|\tilde{\Theta}_i\| + \mu\|L_i V_i\|). \end{aligned}$$

By Assumptions 5 we have $\tau(A) < 1$. Combining this with Lemma 12, it can be seen that

$$\max_{i,j} \|\tilde{\theta}_{k+1}^i - \tilde{\theta}_{k+1}^j\|_{L_2} \leq \tau^{k+1}(A) \max_{i,j} \|\tilde{\theta}_0^i - \tilde{\theta}_0^j\|_{L_2} + O\left(\mu + \sqrt{\mu}\gamma \log \frac{1}{\mu}\right).$$

□

We are now in a position to estimate the difference between $\tilde{\Theta}_{k+1}$ and $\bar{\Theta}_{k+1}$ defined by (6) and (13), respectively.

Lemma 23. *Under conditions of Theorem 13, the following equality holds:*

$$(27) \quad \|\tilde{\Theta}_{k+1} - \bar{\Theta}_{k+1}\|_{L_2} = O\left(\left[\sqrt{\mu} \log \frac{1}{\mu} m^* + \phi(m^*, \mu)\right] \varepsilon_k\left(\frac{\alpha}{2}\right)\right),$$

where $m^* = m^*(\mu)$ is defined by

$$(28) \quad m^*(\mu) = \operatorname{argmin}_{\tilde{m} \geq 1} \left[\tilde{m} \sqrt{\mu} \log^3 \frac{1}{\mu} + \phi(\tilde{m}, \mu) \right].$$

and

$$(29) \quad \varepsilon_k(\alpha) = \left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}} \right) \log \frac{1}{\mu} + (1 - \alpha\mu)^k.$$

In order to prove Lemma 23, we need to introduce the following two lemmas.

Lemma 24 ([16]). *Let $\alpha \in (0, 1)$ be a constant. Then $\forall \mu \in (0, 1)$,*

$$\begin{aligned} i) \quad & \sup_{k \geq 0} (1 - \alpha\mu)^k \sqrt{k} = O(\mu^{-1/2}), \\ & \sup_{k \geq 0} (1 - \alpha\mu)^k k = O(\mu^{-1}), \\ ii) \quad & \sum_{k=0}^{\infty} (1 - \alpha\mu)^k k = O(\mu^{-2}), \\ iii) \quad & \sum_{k=0}^{\infty} (1 - \alpha\mu)^k \sqrt{k} = O(\mu^{-3/2}). \end{aligned}$$

Lemma 25. [22] *Assume that $\{A_t \in \mathbb{R}^{N \times N}, t = 1, 2, \dots, n\}$ is a sequence of stochastic matrices, with a common stationary distribution $\pi = (\pi_1, \pi_2, \dots, \pi_N)$. Set $A \triangleq A_n A_{n-1} \cdots A_1$, then for $i = 1, 2, \dots, N$,*

$$\sum_{j=1}^N \frac{1}{\pi_j} (A_{ij} - \pi_j)^2 \leq \left(\frac{1}{\pi_i} - 1 \right) \prod_{t=1}^n \sigma_2^2(A_t),$$

where $\sigma_2(A)$ is the second largest singular value of the matrix A and $\sigma_2(A) < 1$ if A is ergodic.

Proof of Lemma 23. By (6) and (13), it is seen that

$$\tilde{\Theta}_{k+1} - \bar{\Theta}_{k+1} = (I_{mn} - \mu E[F_k]) \mathbf{A} (\tilde{\Theta}_k - \bar{\Theta}_k) + \mu (E[F_k] - F_k) \mathbf{A} \tilde{\Theta}_k.$$

Define

$$\Phi^0(k+1, i) = (I_{mn} - \mu E[F_k]) \mathbf{A} \Phi^0(k, i),$$

$$\Phi^0(i, i) = I_{mn}, \quad \forall k \geq i.$$

Then for $k \geq 0$

$$\begin{aligned} (30) \quad \tilde{\Theta}_{k+1} - \bar{\Theta}_{k+1} &= \mu \sum_{i=0}^k \Phi^0(k+1, i+1)(E[F_i] - F_i)\mathbf{A}\tilde{\Theta}_i \\ &= \mu \sum_{i=0}^{m^*-1} \Phi^0(k+1, i+1)(E[F_i] - F_i)\mathbf{A}\tilde{\Theta}_i \\ &\quad + \mu \sum_{i=m^*}^k \Phi^0(k+1, i+1)(E[F_i] - F_i)\mathbf{A} \\ &\quad \cdot \{\tilde{\Theta}_{i-m^*} + (\tilde{\Theta}_i - \tilde{\Theta}_{i-m^*})\} \end{aligned}$$

where $m^* = m^*(\mu)$ has been defined in (28). Thus,

$$\sqrt{\mu}m^*(\mu) \leq \sqrt{\mu} \log^3 \frac{1}{\mu} + \phi(1, \mu),$$

which implies that for $\mu \in (0, 1/e)$,

$$m^*(\mu) \leq \log^3 \frac{1}{\mu} + \frac{\phi(1, \mu)}{\sqrt{\mu}} \leq \frac{c}{\sqrt{\mu}}$$

holds for some positive constant c . Consequently, for any $\alpha \in (0, 1)$

$$(1 - \alpha\mu)^{-m^*(\mu)} \leq (1 - \alpha\mu)^{-\frac{c}{\sqrt{\mu}}} \rightarrow 1, \quad \text{as } \mu \rightarrow 0.$$

Hence, the function $(1 - \alpha\mu)^{-m^*(\mu)}$ is bounded for any $\alpha \in (0, 1)$. In the sequel, we will frequently use this fact without further explanations and will drop the variable μ from $m^*(\mu)$ in what follows.

From Assumption 10, it can be deduced that $\{L_k V_k\}$ and $\{\gamma W_{k+1}\}$ are two martingale difference sequences and then by Remark 4, we have

$$(31) \quad \{L_k V_k\} \in \mathcal{M}_8 \subset \mathcal{M}_4, \quad \{\gamma W_{k+1}\} \in \mathcal{M}_8 \subset \mathcal{M}_4.$$

Using Lemma 12, we have $\|\tilde{\Theta}_i\|_{L_4} = O(\varepsilon_{i-1}(\alpha))$. Furthermore, using the inequality $\|E[F_i] - F_i\| \leq 2$ and Lemma 18, we obtain that

$$(32) \quad \mu \left\| \sum_{i=0}^{m^*-1} \Phi^0(k+1, i+1)(E[F_i] - F_i)\mathbf{A}\tilde{\Theta}_i \right\|_{L_2} = O(\mu m^* \varepsilon_k(\alpha)).$$

Now, we consider the last term of (30). Note that

$$\begin{aligned}
 (33) \quad & \tilde{\Theta}_i - \tilde{\Theta}_{i-m^*} = [\tilde{\Theta}_i - \tilde{\Theta}_{i-1}] + \cdots + [\tilde{\Theta}_{i-(m^*-1)} - \tilde{\Theta}_{i-m^*}] \\
 & = \sum_{j=i-m^*}^{i-1} \left([(\mathbf{A} - I_{mn}) - \mu F_j \mathbf{A}] \tilde{\Theta}_j + \mu L_j V_j - \gamma W_{j+1} \right).
 \end{aligned}$$

By the Hölder inequality and (31), we have for any $i \geq m^*$ and $\mu \in (0, 1)$,

$$\begin{aligned}
 & \|\tilde{\Theta}_i - \tilde{\Theta}_{i-m^*}\|_{L_2} \\
 & \leq \sum_{j=i-m^*}^{i-1} \left(\left\| (\mathbf{A} - I_{mn}) \tilde{\Theta}_j \right\|_{L_2} + \mu \|F_j \mathbf{A}\| \cdot \left\| \tilde{\Theta}_j \right\|_{L_2} \right) \\
 & \quad + \left\| \sum_{j=i-m^*}^{i-1} (\mu L_j V_j - \gamma W_{j+1}) \right\|_{L_2} \\
 & = \sum_{j=i-m^*}^{i-1} O \left(\max_{l_1, l_2} \|\tilde{\theta}_j^{l_1} - \tilde{\theta}_j^{l_2}\|_{L_2} + \mu \varepsilon_j(\alpha) \right) + O \left(\sqrt{m^*} [\mu + \gamma] \right) \\
 (34) \quad & = O \left(m^* \left(\mu + \sqrt{\mu} \gamma \log \frac{1}{\mu} + \tau^{i-m^*}(A) \right) + \mu \sum_{j=i-m^*}^{i-1} \varepsilon_j(\alpha) \right) \\
 & \quad + O(\sqrt{m^*} [\mu + \gamma]) \\
 (35) \quad & = O \left(\sqrt{\mu} m^* \left[\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}} \right] \log \frac{1}{\mu} + \mu m^* (1 - \alpha \mu)^{i-m^*} \right) \\
 & \quad + O(m^* \tau^{i-m^*}(A)),
 \end{aligned}$$

where Lemma 22 is used in the inequality (34). Hence, by (35) we have for $k \geq m^*$ and $\mu \in (0, 1/e)$

$$\begin{aligned}
 (36) \quad & \mu \left\| \sum_{i=m^*}^k \Phi^0(k+1, i+1) (E[F_i] - F_i) \mathbf{A} (\tilde{\Theta}_i - \tilde{\Theta}_{i-m^*}) \right\|_{L_2} \\
 & = O \left(\mu \sum_{i=m^*}^k \|\Phi^0(k+1, i+1)\| \cdot \|\tilde{\Theta}_i - \tilde{\Theta}_{i-m^*}\|_{L_2} \right) \\
 & = O \left(\sqrt{\mu} m^* \left[\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}} \right] \log \frac{1}{\mu} + \mu^2 m^* (k - m^*) (1 - \alpha \mu)^{k-m^*} \right) \\
 & \quad + O \left(\mu m^* \sum_{i=m^*}^k (1 - \alpha \mu)^{k-i} \tau^{i-m^*}(A) \right) \\
 & = O \left(\sqrt{\mu} m^* \left[\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}} \right] \log \frac{1}{\mu} \right),
 \end{aligned}$$

where i) of Lemma 24 is used in the last relationship.

Following the idea of [16], we set for $j \geq 0$ and $i \geq m^*$

$$(37) \quad \delta_j^i \triangleq E[F_i|\mathcal{F}_{i-j}] - E[F_i|\mathcal{F}_{i-j-1}].$$

Then for $i \geq m^*$

$$(38) \quad F_i - E[F_i] = \sum_{j=0}^{m^*-1} \delta_j^i + E[F_i|\mathcal{F}_{i-m^*}] - EF_i.$$

For any fixed $0 \leq j \leq m^* - 1$, denote $e_i = \delta_j^i \mathbf{A} \tilde{\Theta}_{i-m^*}$. It is clear that $\{e_i, \mathcal{F}_{i-j}, i \geq m^*\}$ is a martingale difference sequence and

$$(39) \quad \|e_i\|_{L_2} \leq 2\|\tilde{\Theta}_{i-m^*}\|_{L_2} = O(\varepsilon_{i-m^*}(\alpha)), \quad i \geq m^*.$$

Denote $S(k, i) \triangleq \sum_{j=i}^k e_j$. Using the properties of the martingale difference sequence, we have for $i \neq j$, $Ee_i e_j = 0$. Thus, for $k \geq i > m^*$,

$$(40) \quad \begin{aligned} \|S(k, i)\|_{L_2} &= \left\{ \sum_{j=i}^k Ee_j^2 \right\}^{1/2} \\ &= O\left((1 - \alpha\mu)^{i-m^*} \mu^{-1/2} + \sqrt{k-i+1} \left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}} \right) \log \frac{1}{\mu} \right). \end{aligned}$$

Denote $\mathbf{F}_i = I_{mn} - (I_{mn} - \mu EF_i)(A \otimes I_{m^*})$. Thus, we have

$$(41) \quad \begin{aligned} &\mu \left\| \sum_{i=m^*}^k \Phi^0(k+1, i+1) \sum_{j=0}^{m^*-1} \delta_j(i) \mathbf{A} \tilde{\Theta}_{i-m^*} \right\|_{L_2} \\ &\leq \mu \sum_{j=0}^{m^*-1} \left\| \sum_{i=m^*}^k \Phi^0(k+1, i+1) e_i \right\|_{L_2} \\ &\leq \mu \sum_{j=0}^{m^*-1} \|\Phi^0(k+1, m^*+1) S(k, m^*)\|_{L_2} \\ &\quad + \mu \sum_{j=0}^{m^*-1} \left\| \sum_{i=m^*+1}^{k-L} \Phi^0(k+1, i+1) \mathbf{F}_i S(k, i) \right\|_{L_2} \\ &\quad + \mu \sum_{j=0}^{m^*-1} \left\| \sum_{i=k-L+1}^k \Phi^0(k+1, i+1) \mathbf{F}_i S(k, i) \right\|_{L_2} \\ &\triangleq S_1 + S_2 + S_3, \end{aligned}$$

where L is an integer taken as $L = \lceil \frac{\log \mu}{\log \sigma_2(A)} \rceil + 1$. By Lemma 25, it is clear that $\max_{i,j} |(A^L)_{ij} - \frac{1}{n}| \leq \{ \sum_{j=1}^n ((A^L)_{ij} - \frac{1}{n})^2 \}^{1/2} \leq (\sigma_2(A))^L < \mu$.

For the first term of (41), using Lemma 24, we have

$$\begin{aligned}
 (42) \quad S_1 &\leq \mu \sum_{j=0}^{m^*-1} \|\Phi^0(k+1, m^*+1)\| \cdot \|S(k, m^*)\|_{L_2} \\
 &= \mu m^* (1 - \alpha\mu)^{k-m^*} \\
 &\quad \cdot O\left(\sqrt{k-m^*+1} \left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right) \log \frac{1}{\mu} + \mu^{-\frac{1}{2}}\right) \\
 &= O\left(\mu^{1/2} m^* \varepsilon_{k-m^*}(\alpha)\right).
 \end{aligned}$$

For the second term of (41), using the same way as (25) in [15], we have

$$\begin{aligned}
 (43) \quad S_2 &= \mu \sum_{j=0}^{m^*-1} \left\| \sum_{i=m^*+1}^{k-L} \Phi^0(k+1, i+L+1) \right. \\
 &\quad \left. \cdot \Phi^0(i+L+1, i+1) \mathbf{F}_i S(k, i) \right\|_{L_2} \\
 &= O\left(\mu^2 m^* \log \frac{1}{\mu}\right) \sum_{i=m^*+1}^{k-L} \|\Phi^0(k+1, i+L+1) S(k, i)\|_{L_2} \\
 &= O\left(\mu^2 m^* \log \frac{1}{\mu} \sum_{i=m^*+1}^{k-L} \left[(1 - \mu\alpha)^{k-L-m^*} \mu^{-\frac{1}{2}} \right. \right. \\
 &\quad \left. \left. + (1 - \mu\alpha)^{k-i-L} \sqrt{k-i+1} \cdot \left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right) \log \frac{1}{\mu} \right] \right) \\
 &= O\left(\mu^{3/2} m^* \log \frac{1}{\mu} (k-m^*) (1 - \mu\alpha)^{k-m^*} \right. \\
 &\quad \left. + \mu^{1/2} m^* \left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right) \log^2 \frac{1}{\mu} \right),
 \end{aligned}$$

where the boundedness of $(1 - \mu\alpha)^{-L}$ is used. By i) of Lemma 24, we have

$$\left(\frac{1 - \alpha\mu}{1 - \frac{\alpha}{2}\mu}\right)^{k-m^*} (k - m^*) = O(\mu^{-1}).$$

Thus,

$$(44) \quad S_2 = O\left(\mu^{1/2}m^* \log \frac{1}{\mu} \varepsilon_{k-m^*} \left(\frac{\alpha}{2}\right)\right).$$

For the third term, we have by (40)

$$(45) \quad \begin{aligned} S_3 &\leq \mu mL \sup_{k,i} \{\|\Phi^0(k+1, i+1)\| \|S(k, i)\|_{L_2}\} \\ &= O\left(\mu^{1/2}m^* \log \frac{1}{\mu} \varepsilon_{k-m^*}(\alpha)\right). \end{aligned}$$

Substituting (42), (44) and (45) into (41), we obtain that

$$(46) \quad \begin{aligned} &\mu \left\| \sum_{i=m^*}^k \Phi^0(k+1, i+1) \sum_{j=0}^{m^*-1} \delta_j(i) \mathbf{A} \tilde{\Theta}_{i-m^*} \right\|_{L_2} \\ &= O\left(\sqrt{\mu}m^* \log \frac{1}{\mu} \varepsilon_k \left(\frac{\alpha}{2}\right)\right). \end{aligned}$$

Moreover, by Assumption 9 and Lemma 12(with $p = 4$), we have

$$(47) \quad \begin{aligned} &\mu \left\| \sum_{i=m^*}^k \Phi^0(k+1, i+1) \{E[F_i | \mathcal{F}_{i-m^*}] - EF_i\} \cdot \mathbf{A} \tilde{\Theta}_{i-m^*} \right\|_{L_2} \\ &\leq \mu \sum_{i=m^*}^k \|\Phi^0(k+1, i+1)\| \cdot \|E[F_i | \mathcal{F}_{i-m^*}] - EF_i\|_{L_4} \cdot \|\tilde{\Theta}_{i-m^*}\|_{L_4} \\ &= O\left(\phi(m^*, \mu) \varepsilon_k \left(\frac{\alpha}{2}\right)\right), \end{aligned}$$

where the boundedness of $\|V_k\|_{L_8} + \|W_k\|_{L_8}$ is used to estimate $\|\tilde{\Theta}_{i-m^*}\|_{L_4}$.

By (46) and (47), we have

$$(48) \quad \begin{aligned} &\mu \left\| \sum_{i=0}^k \Phi^0(k+1, i+1) [F_i - EF_i] \mathbf{A} \tilde{\Theta}_{i-m^*} \right\|_{L_2} \\ &= O\left(\left[\sqrt{\mu} \log \frac{1}{\mu} m^* + \phi(m^*, \mu)\right] \varepsilon_k \left(\frac{\alpha}{2}\right)\right), \end{aligned}$$

which in conjunction with (36) and (32) yields Lemma 23. □

Proof of Theorem 13. By (13) and Assumption 10 it is evident that

$$(49) \quad \hat{\Pi}_{k+1} = E[\bar{\Theta}_{k+1} \bar{\Theta}_{k+1}^T], \forall k \geq 0.$$

Hence by Schwarz inequality

$$\begin{aligned}
 (50) \quad & \|E[\tilde{\Theta}_{k+1}\tilde{\Theta}_{k+1}^\tau] - \hat{\Pi}_{k+1}\| \\
 &= \|E[\tilde{\Theta}_{k+1}\tilde{\Theta}_{k+1}^\tau - \bar{\Theta}_{k+1}\bar{\Theta}_{k+1}^\tau]\| \\
 &= \|E[(\tilde{\Theta}_{k+1} - \bar{\Theta}_{k+1})\tilde{\Theta}_{k+1}^\tau] + \bar{\Theta}_{k+1}(\tilde{\Theta}_{k+1} - \bar{\Theta}_{k+1}^\tau)]\| \\
 &\leq \|\tilde{\Theta}_{k+1} - \bar{\Theta}_{k+1}\|_{L_2}(\|\tilde{\Theta}_{k+1}\|_{L_2} + \|\bar{\Theta}_{k+1}\|_{L_2}).
 \end{aligned}$$

By Lemma 12 and Lemma 20, we know that

$$(51) \quad \|\tilde{\Theta}_{k+1}\|_{L_2} + \|\bar{\Theta}_{k+1}\|_{L_2} = O(\varepsilon_k(\alpha)).$$

This completes the proof of Theorem 13. □

Proof of Theorem 15. By Lemma 18 and the conditions of Theorem 15, it can be deduced that both $(I - \mu F)(A \otimes I_m)$ and $(I - \mu F)$ are exponentially stable. Thus, $\hat{\Pi}_k$ converges to some matrix Π with the exponentially rate and Π satisfies

$$(52) \quad \Pi = (I - \mu F)(A \otimes I_m)\Pi(A \otimes I_m)(I - \mu F) + \mu^2 R_{v,\varphi} + \gamma^2 Q_\omega.$$

Then we have

$$\begin{aligned}
 (53) \quad \mathbf{A}_{ave}\Pi\mathbf{A}_{ave} &= \mathbf{A}_{ave}(I - \mu F)\mathbf{A}\Pi\mathbf{A}(I - \mu F)\mathbf{A}_{ave} \\
 &\quad + \mu^2\mathbf{A}_{ave}R_{v,\varphi}\mathbf{A}_{ave} + \gamma^2\mathbf{A}_{ave}Q_\omega\mathbf{A}_{ave},
 \end{aligned}$$

where $\mathbf{A}_{ave} \triangleq A_{ave} \otimes I_m$ and $A_{ave} = \lim_{k \rightarrow \infty} A^k$ with $(A_{ave})_{ij} = 1/n$. By the assumption that $F^i = F^j (i \neq j)$, we have $\mathbf{A}_{ave}(I - \mu F) = (I - \mu F)\mathbf{A}_{ave}$. Furthermore, using the fact that $\mathbf{A}_{ave}(A \otimes I_m) = (A \otimes I_m)\mathbf{A}_{ave} = \mathbf{A}_{ave}$, we can rewrite the Equation (53) into the following equivalent form,

$$\begin{aligned}
 (54) \quad \mathbf{A}_{ave}\Pi\mathbf{A}_{ave} &= (I - \mu F)\mathbf{A}_{ave}\Pi\mathbf{A}_{ave}(I - \mu F) \\
 &\quad + \mu^2\mathbf{A}_{ave}R_{v,\varphi}\mathbf{A}_{ave} + \gamma^2\mathbf{A}_{ave}Q_\omega\mathbf{A}_{ave},
 \end{aligned}$$

or another form

$$\begin{aligned}
 (55) \quad & F\mathbf{A}_{ave}\Pi\mathbf{A}_{ave} + \mathbf{A}_{ave}\Pi\mathbf{A}_{ave}F \\
 &= \mu F\mathbf{A}_{ave}\Pi\mathbf{A}_{ave}F + \mu\mathbf{A}_{ave}R_{v,\varphi}\mathbf{A}_{ave} + \frac{\gamma^2}{\mu}\mathbf{A}_{ave}Q_\omega\mathbf{A}_{ave}.
 \end{aligned}$$

According to Lemma 20, we have $\hat{\Pi}_k = O(\mu + \frac{\gamma^2}{\mu}) \log^2 \frac{1}{\mu} + (1 - \mu\alpha)^k$, which is followed by $\Pi = O(\mu + \frac{\gamma^2}{\mu}) \log^2 \frac{1}{\mu}$. Thus, the Equation (55) can be further

rewritten as

$$(56) \quad \begin{aligned} & F \mathbf{A}_{ave} \Pi \mathbf{A}_{ave} + \mathbf{A}_{ave} \Pi \mathbf{A}_{ave} F \\ &= \mu \mathbf{A}_{ave} R_{v,\varphi} \mathbf{A}_{ave} + \frac{\gamma^2}{\mu} \mathbf{A}_{ave} Q_\omega \mathbf{A}_{ave} + O\left((\mu^2 + \gamma^2) \log^2 \frac{1}{\mu}\right). \end{aligned}$$

Now, the solution of the above Lyapunov equation is

$$(57) \quad \mathbf{A}_{ave} \Pi \mathbf{A}_{ave} = \mu \bar{R}_{v,\varphi} + \frac{\gamma^2}{\mu} \bar{R}_\omega + O\left((\mu^2 + \gamma^2) \log^2 \frac{1}{\mu}\right) + o(1).$$

where

$$\begin{aligned} \bar{R}_{v,\varphi} &\triangleq \int_0^\infty e^{-Ft} \mathbf{A}_{ave} R_{v,\varphi} \mathbf{A}_{ave} e^{-Ft} dt \quad \text{and} \\ \bar{R}_\omega &\triangleq \int_0^\infty e^{-Ft} \mathbf{A}_{ave} Q_\omega \mathbf{A}_{ave} e^{-Ft} dt. \end{aligned}$$

Note that for any $i \neq j$, we have

$$\lim_{k \rightarrow \infty} E[\tilde{\theta}_k^i (\tilde{\theta}_k^j)^\tau] = \lim_{k \rightarrow \infty} E[(\tilde{\theta}_k^{i_1} + (\tilde{\theta}_k^i - \tilde{\theta}_k^{i_1})) (\tilde{\theta}_k^{j_1} + (\tilde{\theta}_k^j - \tilde{\theta}_k^{j_1}))^\tau],$$

where $1 \leq i_1, j_1 \leq n$ can be chosen arbitrarily. Using the proof of Lemma 22, we can deduce that $\|\tilde{\theta}_k^j - \tilde{\theta}_k^{j_1}\|_{L_2} = O(\mu + \gamma)$. By Lemma 20 and Hölder inequality, we can obtain that

$$(58) \quad \lim_{k \rightarrow \infty} E[\tilde{\theta}_k^i (\tilde{\theta}_k^j)^\tau] = \lim_{k \rightarrow \infty} E[\tilde{\theta}_k^{i_1} (\tilde{\theta}_k^{j_1})^\tau] + O\left(\left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right) (\mu + \gamma) \log \frac{1}{\mu}\right).$$

By the construction of Π and $\mathbf{A}_{ave} \Pi \mathbf{A}_{ave}$, it can be seen that the i, j th $m \times m$ block of Π is $\lim_{k \rightarrow \infty} E[\tilde{\theta}_k^i (\tilde{\theta}_k^j)^\tau]$, while the term in the corresponding position of $\mathbf{A}_{ave} \Pi \mathbf{A}_{ave}$ is the linear combination of $\{\lim_{k \rightarrow \infty} E[\tilde{\theta}_k^a (\tilde{\theta}_k^b)^\tau], 1 \leq a, b \leq n\}$ with the summation of coefficients equal to 1. Then by (58), it can be computed that

$$(59) \quad \mathbf{A}_{ave} \Pi \mathbf{A}_{ave} = \Pi + O\left(\left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}}\right)^2 \sqrt{\mu} \log \frac{1}{\mu}\right).$$

Combining (59) and (57) with Lemma 13, we complete the proof of Theorem 15. □

4. Concluding remarks

In this paper, we investigated a class of DLMS algorithms, and evaluated the tracking performance without requiring the regression signals to satisfy the previously used stationarity and independency assumptions. Compared with the existing work, we presented an approximate formula for the tracking error covariance matrices of the DLMS algorithm, which can be described by a simple, linear and deterministic matrix equation. Of course, there are still many problems need to be further investigated, for example, how to analyze the distributed Kalman filtering-based algorithms? How to study the distributed control problem using distributed estimation or filtering algorithms?

References

- [1] I. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, *Wireless sensor networks: A survey*, Comput. Netw., **38**(4), 393–422, 2002.
- [2] O. Macchi, *Adaptative processing: the least mean squares approach with applications in transmission*, Chichester, New York: John Wiley & Sons Ltd, 1995.
- [3] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Wiley-IEEE Press, 2003.
- [4] S. S. Haykin, *Adaptive filter theory*, Prentice Hall, 2008.
- [5] S. Kar, J. M. F. Moura and H. V. Poor, *Distributed Linear Parameter Estimation: Asymptotically Efficient Adaptive Strategies*, SIAM J. Control and Optimization, **51**(3), 2200–2229, 2013.
- [6] S. Kar and J. M. F. Moura, *Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs*, IEEE Journal on Selected Topics in Signal Processing, **5**(4), 674–690, Aug. 2011.
- [7] W. S. Chen, C. Y. Wen, S. Y. Hua and C. Y. Sun, *Distributed cooperative adaptive identification and control for a group of continuous-time systems with a cooperative PE condition via consensus*, IEEE Trans. on Automatic Control, **59**(1), 91–106, 2014.
- [8] Q. Zhang and J. F. Zhang, *Distributed parameter estimation over unreliable networks with Markovian switching topologies*, IEEE Trans. on Automatic Control, **57**(10), 2545–2560, 2012.

- [9] S. S. Stankovic, M. S. Stankovic and D. M. Stipanovic, *Decentralized parameter estimation by consensus based stochastic approximation*, IEEE Trans. on Automatic Control, **56**(3), 531–543, 2011.
- [10] F. S. Cattivelli and A. H. Sayed, *Diffusion LMS Strategies for Distributed Estimation*, IEEE Trans. Signal Process., **55**, 2069–2084, 2010.
- [11] I. D. Schizas, G. Mateos and G. B. Giannakis, *Distributed LMS for consensus-based in-network adaptive processing*, IEEE Trans. Signal Process., **8**(6), 2365–2381, 2009.
- [12] A. H. Sayed, *Adaptive networks*, Proceedings of the IEEE, **102**(4), 460–497, 2014.
- [13] A. H. Sayed, *Adaptation, learning, and optimization over networks*, Foundations and Trends in Machine Learning, vol. 7, no. 4–5, 311–801, NOW Publishers, July 2014.
- [14] L. Guo and H. F. Chen, *Identification and Stochastic Adaptive Control*, Birkhäuser, Boston, 1991.
- [15] C. Chen, Z. X. Liu and L. Guo, *Performance of Distributed Adaptive Filters with Cooperative Correlated Signals*, submitted to *Science China Information Sciences*, 2015.
- [16] L. Guo and L. Ljung, *Performance analysis of general tracking algorithms*, IEEE Trans. on Automatic Control, **40**(8), 1388–1402, 1995.
- [17] L. Guo, *Stability of recursive stochastic tracking algorithms*, SIAM J. Control Optim., **32**(5), 1195–1225, 1994.
- [18] L. Guo and L. Ljung, *Exponential stability of general tracking algorithms*, IEEE Trans. on Automatic Control, **40**(8), 1376–1387, 1995.
- [19] L. Guo, *Time Varying Stochastic Systems — Stability, Estimation and Control*, Jilin Science and Technology Press, 1993 (in Chinese).
- [20] E. Seneta, *Non-negative Matrices and Markov Chains*, 2nd ed. New York: Springer-Verlag, 1981.
- [21] A. H. Sayed and C. G. Lopes, *Adaptive processing over distributed networks*, IEICE Trans. Fund. Electron., Commun. Comput. Sci., vol. E90-A, no. 8, 1504–1510, 2007.
- [22] L. Saloff-Coste and J. & Zúñiga, *Convergence of some time inhomogeneous Markov chains via spectral techniques*, Stochastic Processes and Their Applications, **117**, 961–979, 2007.

KEY LABORATORY OF SYSTEMS AND CONTROL

INSTITUTE OF SYSTEMS SCIENCE, AMSS

CHINESE ACADEMY OF SCIENCES

BEIJING 100190, CHINA

E-mail address: chenchen@amss.ac.cn; Lzx@amss.ac.cn; Lguo@amss.ac.cn