Analysis of Distributed Adaptive Filters Based on Diffusion Strategies Over Sensor Networks

Siyu Xie and Lei Guo, *Fellow, IEEE*

***Abstract*—In this paper, we will analyze a basic class of diffusion adaptive filters based on least mean squares algorithms. Both stability and performance analyses will be carried out under a general cooperative information condition, without such stringent conditions as statistical independence and stationarity that have been used in almost all the existing literature and, thus, makes our theory applicable to stochastic systems with feedback. In comparison with the existing work, a key theoretical difficulty that needs to be overcome in this paper is to analyze the product of asymmetric correlated nonstationary random matrices, which is inherent in the structure of the diffusion-type filtering algorithms. We will further demonstrate that the distributed adaptive filters can estimate a dynamic process of interest from noisy measurements by a set of sensors working in a cooperative way, in the natural scenario where none of the sensors can fulfill the estimation task individually due to insufficient information. Finally, the necessity of our cooperative information condition will also be discussed in this paper.**

***Index Terms*—Diffusion strategies, distributed adaptive filters, least mean squares, stochastic stability, tracking performance.**

I. Introduction

DISTRIBUTED adaptive filtering algorithms can estimate a macro unknown parameter process of interest cooperatively in sensor networks. This problem has recently attracted much attention in a number of research areas, e.g., signal processing and distributed control, see [1]–[11]. There are basically two features about adaptive filtering in sensor networks, i.e., distributed observations and distributed processing. For distributed observations, each sensor in the networks can only observe partial information of the unknown parameter process, and the sensor networks can fulfill the estimation task by sharing information among the sensors. In general, there are three different ways for processing in sensor networks, i.e., centralized, distributed, and the combination of both. In the centralized processing, the observations from all sensors are gathered and filtered at the fusion center, which may lack robustness and need strong communication capability over the sensor networks. Distributed processing may overcome these shortcomings, in which all the sensors could collect noisy observations and interact with their neighbors in a certain manner, based on the given network topology.

There are basically three different types of decentralized strategies for distributed adaptive filtering, namely, incremental [3], consensus [4]–[6], and diffusion [7]–[13] strategies, and the last two are fully decentralized. In our previous work [14], we have studied the consensus-type LMS algorithms and established the stability result under a general cooperative information condition. However, as shown by Tu and Sayed [15], the diffusion networks may converge faster and reach lower mean-square deviation than the consensus networks, because diffusion strategies allow information to diffuse more thoroughly through networks than the consensus strategies. Moreover, the diffusion algorithms are fundamentally different from the consensus algorithms in structure and hence in the analysis. Thus, it is necessary to investigate the diffusion strategies and to establish a theory.

The diffusion strategies were originally introduced for the solution of distributed estimation and adaptation problems, in which the sensors exchange estimates with their neighbors and fuse the collected estimates via a linear combination. According to the order of adaptation and combination, there are mainly two different types of diffusion strategies, i.e., the combine-then-adapt (CTA) diffusion strategy and the adapt-then-combine (ATC) diffusion strategy. We remark that almost all the existing related literature about diffusion adaptive filtering algorithms require certain statistical independence or stationarity conditions on the system signals in the stability and performance analyses. For example, in [7], the regressors and noises are independently and identically distributed (i.i.d.) in time and space, and in [8], they investigated the performance of the distributed LMS over sensor networks under i.i.d. regressors and Gaussian measurement noises. Moreover, Khalili *et al.* [9] analyzed the effects of noisy links on the steady-state performance of the diffusion LMS under temporal and spatial independence assumptions. Furthermore, Piggott and Solo [10] developed a theoretical performance analysis of the diffusion LMS under nodewise independence and temporal strict stationarity assumptions, Nosrati *et al.* [11] studied the tracking behavior of a wide range of adaptive networks and analyzed the mean-square-error performance under some time and spatial independence assumptions, Chen and Sayed[12] studied the learning behavior of adaptive networks under some conditions on the conditional expectation of the update vectors, and Gharehshiran *et al.* [13] studied weak convergence of the diffusion LMS for signals with decaying dependence [16], [17].

Manuscript received May 5, 2017; revised September 4, 2017; accepted November 28, 2017. Date of publication January 30, 2018; date of current version October 25, 2018. This work was supported by the National Natural Science Foundation of China under Grant 11688101 and Grant 61227902. Recommended by Associate Editor W. X. Zheng. *(Corresponding author: Lei Guo.)*

The authors are with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and also with School of Mathematical Science, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: xiesiyu7@gmail.com; Lguo@amss.ac.cn).

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/TAC.2018.2799567

To the best of our knowledge, the first step to relax the independence and stationarity conditions in diffusion adaptive filtering algorithms is made in [18], where a CTA diffusion normalized least-mean-squares (NLMS) algorithm is considered. It has been shown that the whole sensor networks can fulfil the estimation task under a cooperative stochastic information condition. Later, the authors further refined the stability result of [18] in [19] and gave a detailed performance analysis in [20]. However, when the sensor network degenerates to a single sensor, the information condition in [18]–[20] cannot include the weakest known information condition introduced by Guo [21] and used in [22], indicating that there is still much room for improvement.

In this paper, we will consider both the CTA and the ATC diffusion strategies. The exponential stability of the homogenous part of the filtering error equation will be established, under a much more general cooperative information condition than that previously used in [12] and [18]–[20]. Compared with the stability analysis for the consensus LMS algorithms in the previous works [14], [23], here the random matrices in the homogenous part of the error equation is no longer symmetric, and we need to establish some new results on possibly asymmetric correlated nonstationary random matrices, and make them applicable to the analysis of the current diffusion algorithms. Under some additional mild conditions, the performance of the filtering algorithms measured by the tracking error covariance matrix will also be provided in the paper. Furthermore, we will show that our new cooperative information condition is also a necessary one for a wide class of stochastic signals with decaying dependence.

In the remainder of this paper, we will present the diffusion NLMS algorithms and introduce some useful definitions in Sections II and III, respectively. The main results will be stated in Section IV. Section V will present the proofs of the main results, Section VI will provide some simulation results, and Section VII will conclude the paper with some remarks.

II. Problem Formulation

Let us consider a sensor network consisting of n sensors. Assume that at each time instant k, each sensor $i = 1, \ldots, n$ in the sensor network receives a noisy scalar measurement y_k^i and an m-dimensional regressor $\boldsymbol{\varphi}_k^i \in \mathbb{R}^m$, where $\mathbb{R}^m$ denotes the set of $m \times 1$ column vectors with real entries. They are related by a stochastic time-varying linear regression model

$$y_k^i = (\boldsymbol{\varphi}_k^i)^T \boldsymbol{\theta}_k + v_k^i, \quad k \geq 0 \tag{1}$$

where $(\cdot)^T$ denotes the transpose operator, v_k^i is the scalar noise, and $\boldsymbol{\theta}_k \in \mathbb{R}^m$ is an unknown time-varying parameter (signal) vector whose variation at time k is denoted by $\Delta\boldsymbol{\theta}_k$, i.e.,

$$\Delta\boldsymbol{\theta}_k \triangleq \boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}, \quad k \geq 1. \tag{2}$$

Many problems from different application areas can be cast as (1), see, e.g., [24]–[27]. Note also that when $\Delta\boldsymbol{\theta}_k \equiv 0$, $\boldsymbol{\theta}_k$ reduces to a constant vector.

As usual, let the communication structure among sensors be represented by an undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ is the set of sensors and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges. The structure of the graph $\mathcal{G}$ is described by $\mathcal{A} = \{a_{ij}\}_{n\times n}$ which is called the weighted adjacency matrix, 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 $\mathcal{A}$ satisfy $a_{ij} = a_{ji}, \ \forall i, j = 1, \ldots, n$, and $\sum_{j=1}^n a_{ij} = 1, \ \forall i = 1, \ldots, n$. Thus, the matrix $\mathcal{A}$ is doubly stochastic.[1] Note that $(i,j) \in \mathcal{E} \Leftrightarrow a_{ij} > 0$. The set of neighbors of the sensor i is denoted as

$$\mathcal{N}_i = \{l \in V | (i,l) \in \mathcal{E}\}$$

and the sensor i shares information with its neighboring sensors in $\mathcal{N}_i$. The Laplacian matrix $\mathcal{L}$ of the graph $\mathcal{G}$ is defined by $\mathcal{L} = I_n - \mathcal{A}$, where I_n denotes the n-dimensional identity matrix.

The well-known LMS algorithm [26]–[29] can be used to estimate the unknown parameter of interest and it is a type of steepest descent algorithm that aims at minimizing the mean square prediction error recursively. The LMS algorithm has a number of well appreciated advantages (see in [29]): simplicity, efficiency, robustness, and numerical stability, and is the most basic adaptive algorithm in many areas, such as system identification, adaptive control, adaptive signal processing, etc.

In the following, we present the CTA and ATC diffusion strategies, both are based on NLMS algorithms.

Algorithm 1: CTA Diffusion NLMS Algorithm.

For any given sensor $i = 1 \ldots, n$**, begin with an initial estimate** $\widehat{\boldsymbol{\theta}}_0^{i,\text{CTA}}$.

The algorithm is recursively defined for iteration $k \geq 0$ as follows:

1: Combine local estimates:

$$\widehat{\boldsymbol{\beta}}_k^{i,\text{CTA}} = \sum_{l\in\mathcal{N}_i} a_{li} \widehat{\boldsymbol{\theta}}_k^{l,\text{CTA}}.$$

2: Adapt the local estimate:

$$\widehat{\boldsymbol{\theta}}_{k+1}^{i,\text{CTA}} = \widehat{\boldsymbol{\beta}}_k^{i,\text{CTA}} + \mu_i \frac{\boldsymbol{\varphi}_k^i}{1 + \| \boldsymbol{\varphi}_k^i \|^2} [y_k^i - (\boldsymbol{\varphi}_k^i)^T \widehat{\boldsymbol{\beta}}_k^{i,\text{CTA}}]$$

where $\mu_i \in (0,1)$ is a constant step-size of the sensor i.

Algorithm 2: ATC Diffusion NLMS Algorithm.

For any given sensor $i = 1 \ldots, n$**, begin with an initial estimate** $\widehat{\boldsymbol{\theta}}_0^{i,\text{CTA}}$.

The algorithm is recursively defined for iteration $k \geq 0$ as follows:

1: Adapt the local estimate:

$$\widehat{\boldsymbol{\beta}}_{k+1}^{i,\text{ATC}} = \widehat{\boldsymbol{\theta}}_k^{i,\text{ATC}} + \mu_i \frac{\boldsymbol{\varphi}_k^i}{1 + \| \boldsymbol{\varphi}_k^i \|^2} [y_k^i - (\boldsymbol{\varphi}_k^i)^T \widehat{\boldsymbol{\theta}}_k^{i,\text{ATC}}].$$

2: Combine local estimates:

$$\widehat{\boldsymbol{\theta}}_{k+1}^{i,\text{ATC}} = \sum_{l\in\mathcal{N}_i} a_{li} \widehat{\boldsymbol{\beta}}_{k+1}^{l,\text{ATC}}$$

where $\mu_i \in (0,1)$ is a constant step-size of the sensor i.

For convenience of analysis, we introduce the following notations:

$$\boldsymbol{Y}_k \triangleq \text{col}\{y_k^1, \ldots, y_k^n\}, \qquad (n \times 1)$$

$$\boldsymbol{\Phi}_k \triangleq \text{diag}\{\boldsymbol{\varphi}_k^1, \ldots, \boldsymbol{\varphi}_k^n\}, \qquad (mn \times n)$$

[1] A matrix is called doubly stochastic, if all elements are nonnegative, both the sum of each row and the sum of each column equal to 1.

$$\boldsymbol{V}_k \triangleq \text{col}\{v_k^1, \ldots, v_k^n\}, \qquad (n \times 1)$$

$$\Delta\boldsymbol{\Theta}_k \triangleq \text{col}\{\underbrace{\Delta\boldsymbol{\theta}_k, \ldots, \Delta\boldsymbol{\theta}_k}_{n}\}, \qquad (mn \times 1)$$

$$\boldsymbol{\Theta}_k \triangleq \text{col}\{\underbrace{\boldsymbol{\theta}_k, \ldots, \boldsymbol{\theta}_k}_{n}\}, \qquad (mn \times 1)$$

$$\widehat{\boldsymbol{\Theta}}_k^{\text{CTA}} \triangleq \text{col}\{\widehat{\boldsymbol{\theta}}_k^{1,\text{CTA}}, \ldots, \widehat{\boldsymbol{\theta}}_k^{n,\text{CTA}}\}, \qquad (mn \times 1)$$

$$\widetilde{\boldsymbol{\Theta}}_k^{\text{CTA}} \triangleq \text{col}\{\widetilde{\boldsymbol{\theta}}_k^{1,\text{CTA}}, \ldots, \widetilde{\boldsymbol{\theta}}_k^{n,\text{CTA}}\}, \qquad (mn \times 1)$$

$$\text{where } \widetilde{\boldsymbol{\theta}}_k^{i,\text{CTA}} = \widehat{\boldsymbol{\theta}}_k^{i,\text{CTA}} - \boldsymbol{\theta}_k,$$

$$\widehat{\boldsymbol{\Theta}}_k^{\text{ATC}} \triangleq \text{col}\{\widehat{\boldsymbol{\theta}}_k^{1,\text{ATC}}, \ldots, \widehat{\boldsymbol{\theta}}_k^{n,\text{ATC}}\}, \qquad (mn \times 1)$$

$$\widetilde{\boldsymbol{\Theta}}_k^{\text{ATC}} \triangleq \text{col}\{\widetilde{\boldsymbol{\theta}}_k^{1,\text{ATC}}, \ldots, \widetilde{\boldsymbol{\theta}}_k^{n,\text{ATC}}\}, \qquad (mn \times 1)$$

$$\text{where } \widetilde{\boldsymbol{\theta}}_k^{i,\text{ATC}} = \widehat{\boldsymbol{\theta}}_k^{i,\text{ATC}} - \boldsymbol{\theta}_k,$$

$$\widehat{\boldsymbol{\mathcal{B}}}_k^{\text{CTA}} \triangleq \text{col}\{\widehat{\boldsymbol{\beta}}_k^{1,\text{CTA}}, \ldots, \widehat{\boldsymbol{\beta}}_k^{n,\text{CTA}}\}, \qquad (mn \times 1)$$

$$\widehat{\boldsymbol{\mathcal{B}}}_k^{\text{ATC}} \triangleq \text{col}\{\widehat{\boldsymbol{\beta}}_k^{1,\text{ATC}}, \ldots, \widehat{\boldsymbol{\beta}}_k^{n,\text{ATC}}\}, \qquad (mn \times 1)$$

$$\boldsymbol{L}_k \triangleq \text{diag}\left\{\frac{\boldsymbol{\varphi}_k^1}{1+\|\boldsymbol{\varphi}_k^1\|^2}, \ldots, \frac{\boldsymbol{\varphi}_k^n}{1+\|\boldsymbol{\varphi}_k^n\|^2}\right\}, \qquad (mn \times n)$$

$$\boldsymbol{F}_k \triangleq \boldsymbol{L}_k \boldsymbol{\Phi}_k^T, \qquad (mn \times mn)$$

$$\boldsymbol{\Lambda} \triangleq \text{diag}\{\mu_1 I_m, \ldots, \mu_n I_m\}, \qquad (mn \times mn)$$

$$\mathscr{L} \triangleq \mathcal{L} \otimes I_m, \qquad (mn \times mn)$$

where $\text{col}\{\cdots\}$ denotes a vector by stacking the specified vectors, $\text{diag}\{\cdots\}$ is used in a nonstandard manner which means that $m \times 1$ column vectors are combined "in a diagonal manner" resulting in a $mn \times n$ matrix, and $\otimes$ is the Kronecker product. Note also that $\Delta\boldsymbol{\Theta}_k$ and $\boldsymbol{\Theta}_k$ mean just the n-times replication of vectors $\Delta\boldsymbol{\theta}_k$ and $\boldsymbol{\theta}_k$, respectively. By (1) and (2), we have

$$\boldsymbol{Y}_k = \boldsymbol{\Phi}_k^T \boldsymbol{\Theta}_k + \boldsymbol{V}_k \tag{3}$$

and

$$\Delta\boldsymbol{\Theta}_{k+1} = \boldsymbol{\Theta}_{k+1} - \boldsymbol{\Theta}_k. \tag{4}$$

For the CTA diffusion NLMS algorithm, we have

$$\begin{cases} \widehat{\boldsymbol{\mathcal{B}}}_k^{\text{CTA}} = (\mathcal{A} \otimes I_m)\widehat{\boldsymbol{\Theta}}_k^{\text{CTA}} \\ \widehat{\boldsymbol{\Theta}}_{k+1}^{\text{CTA}} = \widehat{\boldsymbol{\mathcal{B}}}_k^{\text{CTA}} + \boldsymbol{\Lambda}\boldsymbol{L}_k(\boldsymbol{Y}_k - \boldsymbol{\Phi}_k^T \widehat{\boldsymbol{\mathcal{B}}}_k^{\text{CTA}}) \end{cases} \tag{5}$$

where $\mathcal{A}$ is the adjacency matrix. Denoting $\widetilde{\boldsymbol{\Theta}}_k^{\text{CTA}} = \widehat{\boldsymbol{\Theta}}_k^{\text{CTA}} - \boldsymbol{\Theta}_k$, substituting (3) into (5), and noticing (4), we can get

$$\widetilde{\boldsymbol{\Theta}}_{k+1}^{\text{CTA}} = (\mathcal{A} \otimes I_m)\widehat{\boldsymbol{\Theta}}_k^{\text{CTA}} - \boldsymbol{\Theta}_k - \Delta\boldsymbol{\Theta}_k + \boldsymbol{\Lambda}\boldsymbol{L}_k[\boldsymbol{\Phi}_k^T \boldsymbol{\Theta}_k + \boldsymbol{V}_k - \boldsymbol{\Phi}_k^T (\mathcal{A} \otimes I_m)\widehat{\boldsymbol{\Theta}}_k^{\text{CTA}}].$$

Because $(\mathcal{A} \otimes I_m)\boldsymbol{\Theta}_k = \boldsymbol{\Theta}_k$, we have

$$\widetilde{\boldsymbol{\Theta}}_{k+1}^{\text{CTA}} = (I_{mn} - \boldsymbol{\Lambda}\boldsymbol{F}_k)(\mathcal{A} \otimes I_m)\widetilde{\boldsymbol{\Theta}}_k^{\text{CTA}} + \boldsymbol{\Lambda}\boldsymbol{L}_k\boldsymbol{V}_k - \Delta\boldsymbol{\Theta}_{k+1}.$$

Since $\mathcal{A} \otimes I_m = I_{mn} - (\mathcal{L} \otimes I_m) = I_{mn} - \mathscr{L}$, we can obtain the CTA diffusion NLMS error equation as follows:

$$\widetilde{\boldsymbol{\Theta}}_{k+1}^{\text{CTA}} = (I_{mn} - [\boldsymbol{\Lambda}\boldsymbol{F}_k + \mathscr{L} - \boldsymbol{\Lambda}\boldsymbol{F}_k\mathscr{L}])\widetilde{\boldsymbol{\Theta}}_k^{\text{CTA}} + \boldsymbol{\Lambda}\boldsymbol{L}_k\boldsymbol{V}_k - \Delta\boldsymbol{\Theta}_{k+1}. \tag{6}$$

Obviously, its homogeneous equation is

$$\boldsymbol{X}_{k+1} = (I_{mn} - [\boldsymbol{\Lambda}\boldsymbol{F}_k + \mathscr{L} - \boldsymbol{\Lambda}\boldsymbol{F}_k\mathscr{L}])\boldsymbol{X}_k, \ k \geq 0 \tag{7}$$

which will be analyzed in the following sections.

In a similar way, we can obtain the ATC diffusion NLMS error equation

$$\widetilde{\boldsymbol{\Theta}}_{k+1}^{\text{ATC}} = (I_{mn} - [\boldsymbol{\Lambda}\boldsymbol{F}_k + \mathscr{L} - \boldsymbol{\Lambda}\mathscr{L}\boldsymbol{F}_k])\widetilde{\boldsymbol{\Theta}}_k^{\text{ATC}} + \boldsymbol{\Lambda}(\mathcal{A} \otimes I_m)\boldsymbol{L}_k\boldsymbol{V}_k - \Delta\boldsymbol{\Theta}_{k+1} \tag{8}$$

together with its homogeneous equation

$$\boldsymbol{Y}_{k+1} = (I_{mn} - [\boldsymbol{\Lambda}\boldsymbol{F}_k + \mathscr{L} - \boldsymbol{\Lambda}\mathscr{L}\boldsymbol{F}_k])\boldsymbol{Y}_k, \ k \geq 0 \tag{9}$$

which can be analyzed in a similar way as (7). In the sequel, we will mainly focus on the analysis of the CTA diffusion NLMS algorithms.

Note that by the stochastic internal–external stability results in [21] (see Propositions 2.1 and 2.2 there), it is easy to see from the above distributed filtering error (6) that the tracking error hinges on the exponential stability of (7), which depends essentially on the properties of product of random matrices. We remark that, compared with the consensus NLMS error equation in [14], here the random matrices $\boldsymbol{\Lambda}\boldsymbol{F}_k + \mathscr{L} - \boldsymbol{\Lambda}\boldsymbol{F}_k\mathscr{L}$ and $\boldsymbol{\Lambda}\boldsymbol{F}_k + \mathscr{L} - \boldsymbol{\Lambda}\mathscr{L}\boldsymbol{F}_k$ are asymmetric, and hence, the method used in [14] is no longer applicable. In order to obtain the exponential stability of (7) and (9), we need to generalize the results in [21] for symmetric random matrices to possible asymmetric random matrices, and make them applicable to the current proofs of the diffusion algorithms. Before that, we first give some definitions in the following section.

III. Some Definitions

In the sequel, the set of $m \times n$ matrices with real entries is denoted by $\mathbb{R}^{m\times n}$. Let $X \in \mathbb{R}^{n\times n}$ and $Y \in \mathbb{R}^{n\times n}$ be two symmetric matrices, then $X \geq Y$ means that $X - Y$ is a positive semidefinite matrix and $X > Y$ means that $X - Y$ is a positive definite matrix. Also, let $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ denote the largest and the smallest eigenvalues of a matrix $(\cdot)$, respectively. For any random matrix $X \in \mathbb{R}^{m\times n}$, its Euclidean norm is defined as its maximum singular value, i.e., $\| X \| = \{\lambda_{\max}(XX^T)\}^{\frac{1}{2}}$, and its L_p-norm is defined as $\| X \|_{L_p} = \{\mathbb{E}[\| X \|^p]\}^{\frac{1}{p}}$, where $\mathbb{E}[\cdot]$ denotes the expectation operator. Also, we use $\mathcal{F}_k = \sigma\{\varphi_i^j, \omega_i, v_{i-1}^j, j = 1, \ldots, n, i \leq k\}$ to denote the σ-algebra generated by $\{\varphi_i^j, \omega_i, v_{i-1}^j, j = 1, \ldots, n, i \leq k\}$, where the definition of σ-algebra together with that of conditional mathematical expectation operator $\mathbb{E}[\cdot|\mathcal{F}_k]$ to be used later can be found in [30]. To proceed with further discussions, we need the following definitions introduced in [21].

Definition 3.1: For a random matrix sequence $\{A_k, k \geq 0\}$ defined on the basic probability space $(\Omega, \mathcal{F}, P)$, if

$$\sup_{k\geq 0} \mathbb{E}[\| A_k \|^p] < \infty$$

holds for some $p > 0$, then $\{A_k\}$ is called L_p-bounded. Furthermore, if $\{A_k\}$ is a solution of a random difference equation, then $\{A_k\}$ is called L_p-stable.

Definition 3.2: For a sequence of $d \times d$ random matrices $A = \{A_k, k \geq 0\}$, if it belongs to the following set with $p \geq 0$,

$$S_p(\lambda) = \left\{ A : \left\| \prod_{j=i+1}^{k} (I - A_j) \right\|_{L_p} \leq M\lambda^{k-i} \quad \forall k \geq i+1 \quad \forall i \geq 0, \text{for some } M > 0 \right\} \tag{10}$$

then $\{I - A_k, k \geq 0\}$ is called L_p-exponentially stable with parameter $\lambda \in [0, 1)$.

As pointed out in [21], (10) is in some sense the necessary and sufficient condition for stability of random linear equations of the form $x_k = (I - A_k)x_k + \xi_{k+1}, k \geq 0$, and it is well known that the analysis of such a random matrix product is a mathematically difficult problem. However, as demonstrated by Guo [21], for linear random equations arising from adaptive filtering algorithms, it is possible to transfer the product of the random matrices to that of a certain class of scalar sequences, and the later can be further analyzed based on some excitation or information conditions on the regressors. This paper will follow a similar line of arguments for distributed algorithms. To this end, we introduce the following subclass of $S_1(\lambda)$ for a scalar sequence $a = \{a_k, k \geq 0\}$

$$S^0(\lambda) = \left\{ a : a_k \in [0, 1], \mathbb{E}\left[\prod_{j=i+1}^{k} (1 - a_j) \right] \leq M\lambda^{k-i} \quad \forall k \geq i+1 \quad \forall i \geq 0, \text{for some } M > 0 \right\} \tag{11}$$

where $\lambda \in [0, 1)$. This definition will be used when we transfer the product of random matrices to that of a scalar sequence.

Definition 3.3: A random sequence $x = \{x_k\}$ is called an element of the weakly dependent set $\mathcal{M}_p (p \geq 1)$, if there exists a constant C_p^x depending only on p and the distribution of $\{x_k\}$ such that for any $k \geq 0$ and $h \geq 1$,

$$\left\| \sum_{i=k+1}^{k+h} x_i \right\|_{L_p} \leq C_p^x h^{\frac{1}{2}}. \tag{12}$$

Remark 3.1: It is known that many typical random sequences, such as the martingale difference, zero mean ϕ- and α-mixing sequences, and the linear process driven by white noises, all belong to $\mathcal{M}_p$ (see [22]).

Definition 3.4: Let $\{A_k, k \geq 0\}$ be a matrix sequence and $\{b_k, k \geq 0\}$ be a positive scalar sequence. Then, by $A_k = O(b_k)$ we mean that there exists a constant $M > 0$ such that

$$\|A_k\| \leq M b_k \quad \forall k \geq 0. \tag{13}$$

IV. Main Results

A. Stability and Performance Results

In this section, we study the stability of the error (6) and first give an exponential stability result for the homogeneous part (7) in the following theorem. For that, we need the following conditions.

Condition 4.1 (Network Topology): The graph $\mathcal{G}$ is connected and contains a self-loop at each node.

Remark 4.1: It is known that the eigenvalues of the Laplacian matrix $\mathcal{L}$ of the graph $\mathcal{G}$ can be arranged in a nondecreasing order $0 = \lambda_1(\mathcal{L}) \leq \lambda_2(\mathcal{L}) \leq \cdots \leq \lambda_n(\mathcal{L}) \leq 2$. The smallest eigenvalue $\lambda_1(\mathcal{L})$ always equals to zero, with $\frac{1}{\sqrt{n}}(1, \ldots, 1)^T$ being the corresponding unit eigenvector. Under *Condition 4.1*, we know that $\lambda_2(\mathcal{L}) > 0$ and $\lambda_n(\mathcal{L}) < 2$, see [31].

Condition 4.2 (Cooperative Information Condition): For the adapted sequences $\{\boldsymbol{\varphi}_k^i, \mathcal{F}_k, k \geq 0\}(i = 1, \ldots, n)$, there exists an integer $h > 0$ such that $\{\lambda_k, k \geq 0\} \in S^0(\lambda)$ for some $\lambda \in (0, 1)$, where λ_k is defined by

$$\lambda_k \triangleq \lambda_{\min}\left\{ \mathbb{E}\left[\frac{1}{n(h+1)} \sum_{i=1}^{n} \sum_{j=k+1}^{k+h} \frac{\boldsymbol{\varphi}_j^i (\boldsymbol{\varphi}_j^i)^T}{1 + \|\boldsymbol{\varphi}_j^i\|^2} \Big| \mathcal{F}_k \right] \right\} \tag{14}$$

where $\mathbb{E}[\cdot|\mathcal{F}_k]$ is the conditional mathematical expectation operator and $\mathcal{F}_k = \sigma\{\boldsymbol{\varphi}_j^i, \boldsymbol{\omega}_j, v_{j-1}^i, i = 1, \ldots, n, j \leq k\}$.

Remark 4.2: Most of the existing theories on distributed adaptive filters require that the regressors satisfy some statistical independence and stationarity conditions, which are rather stringent and cannot be satisfied for stochastic signals generated from feedback systems. This conditional mathematical expectation-based information condition was first introduced by Guo in [32] and then refined in [21] for the traditional single sensor case, which is quite general and even necessary for exponential stability (see [21]). Our *Condition 4.2* is a natural generalization of the information condition from single sensor to sensor networks. We remark that, by properties of the conditional mathematical expectation, *Condition 4.2* implies that the system signals will have some kind of "persistent excitations" since the prediction of the "future" is nondegenerate given the "past," which is required to track constantly changing unknown signals. We remark also that the information condition used in [18] is only a special case of *Condition 4.2* with $h = 1$. Moreover, under *Condition 4.2*, the distributed filtering network can be shown to fulfil the estimation task cooperatively even if any individual filter cannot.

Without loss of generality, we express the step-size at each node i as $\mu_i = \sigma_i \mu^*$, where $\mu^* = \max\{\mu_1, \ldots, \mu_n\} \in (0, 1), \sigma_i \in (0, 1]$. The main results of the paper are as follows.

Theorem 4.1: Consider the model (1) and the estimation error (6). Suppose that *Conditions 4.1* and *4.2* are satisfied. Then for any $p \geq 1$, there exists a constant $\mu^* \in (0, 1)$, such that for any $0 < \boldsymbol{\Lambda} \leq \mu^* I_{mn}$,

$$\{I_{mn} - [\boldsymbol{\Lambda F}_k + \mathscr{L} - \boldsymbol{\Lambda F}_k \mathscr{L}], k \geq 1\}$$

is L_p-exponentially stable $(p \geq 1)$.

The detailed proof of *Theorem 4.1* is given in Section V, and the precise value of $\mu^* \in (0, 1)$ can be found in (41). By *Theorem 4.1*, we can obtain a preliminary tracking error bound in the following theorem.

Theorem 4.2: Consider the model (1) and the estimation error (6). Suppose that *Conditions 4.1* and 4.2 are satisfied. If for some $p \geq 1$ and $\beta > 1$,

$$\sigma_p \triangleq \sup_k \|\xi_k \log^\beta (e + \xi_k)\|_{L_p} < \infty$$

$$\|\widetilde{\boldsymbol{\Theta}}_0^{\text{CTA}}\|_{L_p} < \infty$$

hold, where $\xi_k = \|\boldsymbol{V}_k\| + \|\Delta\boldsymbol{\Theta}_{k+1}\|$, then there exists a constant $\mu^* \in (0, 1)$, such that for any $0 < \boldsymbol{\Lambda} \leq \mu^* I_{mn}$, $\{\widetilde{\boldsymbol{\Theta}}_k^{\text{CTA}},$

$k \geq 1\}$ is L_p-stable and

$$\limsup_{k\to\infty} \|\widetilde{\Theta}_k^{\mathrm{CTA}}\|_{L_p} \leq c[\sigma_p log(e+\sigma_p^{-1})] \tag{15}$$

where c is a positive constant.

Remark 4.3: Since the upper bound for the error can be derived in a similar way as that of [21, Th. 4.2], details will be omitted here. From this theorem, we know that when both the observation noises v_k^i and parameter variation $\Delta\boldsymbol{\theta}_k$ are small in the "L_p sense," σ_p will be small, and consequently, the tracking error will also be small in the L_p sense.

Remark 4.4: The result of *Theorem 4.1* also holds similarly for the ATC diffusion NLMS algorithm, namely, for any $p \geq 1$, there exists a constant $\mu^* \in (0,1)$, such that for any $0 < \boldsymbol{\Lambda} \leq \mu^* I_{mn}$,

$$\{I_{mn} - [\boldsymbol{\Lambda}\boldsymbol{F}_k + \mathscr{L} - \boldsymbol{\Lambda}\mathscr{L}\boldsymbol{F}_k], k \geq 1\}$$

is L_p-exponentially stable ($p \geq 1$). Also, *Theorem 4.2* holds for $\{\widetilde{\Theta}_k^{\mathrm{ATC}}, k \geq 1\}$.

To obtain a more accurate tracking error bound, we assume that the variation $\Delta\boldsymbol{\theta}_{k+1}$ has the following form:

$$\Delta\boldsymbol{\theta}_{k+1} \stackrel{\triangle}{=} \gamma\boldsymbol{\omega}_{k+1}, \quad k \geq 1 \tag{16}$$

where γ is a nonnegative number reflecting the speed of parameter variation and $\boldsymbol{\omega}_k$ is an as yet undefined vector. Henceforth, we denote $\boldsymbol{\Omega}_k \stackrel{\triangle}{=} \mathrm{col}\{\boldsymbol{\omega}_k, \ldots, \boldsymbol{\omega}_k\}$.

Condition 4.3: For some $p \geq 1$, the initial estimation error is bounded, i.e., $\|\widetilde{\Theta}_0\|_{L_{2p}} < \infty$. Furthermore, $\{\boldsymbol{L}_k\boldsymbol{V}_k\} \in \mathcal{M}_{2p}$ and $\{\boldsymbol{\Omega}_k\} \in \mathcal{M}_{2p}$.

Remark 4.5: This condition simply implies that both the noises and parameter variations are weakly dependent with certain bounded moments.

Theorem 4.3: Assume that *Conditions 4.1–4.3* are satisfied. For any $p \geq 1$, there exists a constant $\mu^* \in (0,1)$, such that for any $0 < \boldsymbol{\Lambda} \leq \mu^* I_{mn}$, we have for all $k \geq 0$,

$$\|\widetilde{\Theta}_{k+1}^{\mathrm{CTA}}\|_{L_p} = O\left(\left[\sqrt{\mu^*} + \frac{\gamma}{\sqrt{\mu^*}}\right]\log\frac{1}{\mu^*} + (1-\alpha_{2p}\mu^*)^{k+1}\right) \tag{17}$$

where $\alpha_{2p} \in (0,1)$ is a constant which is defined as in *Lemma 5.11*, "O" is a constant depends only on α.

The detailed proof of *Theorem 4.3* is given in Section V. Rather than giving upper bounds only, we may further get the approximate value of the mean square tracking error matrix by strengthening the conditions used in *Theorem 4.3*. Now, to approximate the true mean square tracking error matrix

$$\boldsymbol{\Pi}_k^{\mathrm{CTA}} = \mathbb{E}[\widetilde{\Theta}_k^{\mathrm{CTA}}(\widetilde{\Theta}_k^{\mathrm{CTA}})^T]$$

we define the following linear deterministic difference equation for $\widehat{\boldsymbol{\Pi}}_k^{\mathrm{CTA}}$ [22]:

$$\begin{aligned}\widehat{\boldsymbol{\Pi}}_{k+1}^{\mathrm{CTA}} &= (I_{mn} - \mathbb{E}[\boldsymbol{A}_k])\widehat{\boldsymbol{\Pi}}_k^{\mathrm{CTA}}(I_{mn} - \mathbb{E}[\boldsymbol{A}_k])^T \\ &\quad + \boldsymbol{\Lambda}^2\mathbb{E}[\boldsymbol{T}_k] + \gamma^2\boldsymbol{Q}_\omega(k+1)\end{aligned} \tag{18}$$

where

$$\begin{aligned}\boldsymbol{A}_k &= \boldsymbol{\Lambda}\boldsymbol{F}_k + \mathscr{L} - \boldsymbol{\Lambda}\boldsymbol{F}_k\mathscr{L} \\ \boldsymbol{T}_k &= \boldsymbol{L}_k\boldsymbol{V}_k\boldsymbol{V}_k^T\boldsymbol{L}_k^T \\ \boldsymbol{Q}_\omega(k+1) &= \mathbb{E}[\boldsymbol{\Omega}_{k+1}\boldsymbol{\Omega}_{k+1}^T] \\ \widehat{\boldsymbol{\Pi}}_0^{\mathrm{CTA}} &= \mathbb{E}[\widetilde{\Theta}_0^{\mathrm{CTA}}(\widetilde{\Theta}_0^{\mathrm{CTA}})^T].\end{aligned}$$

Note that $\widehat{\boldsymbol{\Pi}}_k^{\mathrm{CTA}}$ can easily be calculated and examined, which will be used to approximate the mean square tracking error matrix $\boldsymbol{\Pi}_k^{\mathrm{CTA}}$. For this, we need the following additional conditions.

Condition 4.4: Let $\mathcal{F}_k = \sigma\{\boldsymbol{\varphi}_j^i, \boldsymbol{\omega}_j, v_{j-1}^i, i = 1, \ldots, n, j \leq k\}$, and assume that $\boldsymbol{\Lambda} = \mu I_{mn}$ and for all $k \geq 1$,

$$\begin{gathered}\mathbb{E}[\boldsymbol{V}_k|\mathcal{F}_k] = 0, \mathbb{E}[\boldsymbol{\Omega}_{k+1}|\mathcal{F}_k] = 0, \mathbb{E}[\boldsymbol{\Omega}_{k+1}\boldsymbol{V}_k^T|\mathcal{F}_k] = 0 \\ \mathbb{E}[\boldsymbol{V}_k\boldsymbol{V}_k^T|\mathcal{F}_k] = \boldsymbol{P}_v(k) \geq 0 \\ \mathbb{E}[\boldsymbol{\Omega}_{k+1}\boldsymbol{\Omega}_{k+1}^T] = \boldsymbol{Q}_\omega(k+1) \geq 0 \\ \sup_k(\|\boldsymbol{V}_k\|_{L_8} + \|\boldsymbol{\Omega}_k\|_{L_8}) < \infty\end{gathered} \tag{19}$$

and that there exists a bounded function $\bar{\phi}(t,\mu) \geq 0$ with $\lim_{t\to\infty,\mu\to 0}\bar{\phi}(t,\mu)\log\frac{1}{\mu} = 0$ such that $\forall k \geq 0 \quad \forall t$ and $\mu \in (0,1)$,

$$\|\mathbb{E}[\boldsymbol{F}_k|\mathcal{F}_{k-t}] - \mathbb{E}[\boldsymbol{F}_k]\|_{L_4} \leq \bar{\phi}(t,\mu). \tag{20}$$

Remark 4.6: If we are only interested to get an upper bound of $\widetilde{\Theta}_k$, then some moment conditions on $\boldsymbol{V}_k$ and $\boldsymbol{\Omega}_k$ are sufficient, see *Theorem 4.1*. A refined upper bound can also be obtained under the additional *Conditions 4.3*, see Theorem 4.3. Moreover, the stronger *Condition 4.4* will be only used to obtain an approximate tracking performances as will be shown in the following Theorem 4.4. *Conditions 4.4* means that the measurement noise $\boldsymbol{V}_k$ and the parameter variation $\boldsymbol{\Omega}_{k+1}$ are of white noise characters, which are commonly used in many works [11] and is a worst-case analysis since the future behavior of the model is unpredictable, as mentioned in [33]. This condition also means that the observation noise and the parameter variations are uncorrelated given the past signals, but spatial correlations of the noises are allowed. Also, (20) describes the decaying correlation between $\boldsymbol{F}_k$ and $\mathcal{F}_{k-t}$, which can be guaranteed by imposing certain weak dependence conditions on the regressor $\{\boldsymbol{\varphi}_k^i\}$, e.g., ϕ-mixing property (see [22]).

Theorem 4.4: Let *Conditions 4.1–4.4* be satisfied. Then, there exists a constant $\mu^* \in (0,1)$, such that for any $0 < \mu \leq \mu^*$, we have for all $k \geq 1$,

$$\|\boldsymbol{\Pi}_{k+1}^{\mathrm{CTA}} - \widehat{\boldsymbol{\Pi}}_{k+1}^{\mathrm{CTA}}\| \leq c\bar{\delta}(\mu)\left[\mu + \frac{\gamma^2}{\mu} + (1-\alpha\mu)^{k+1}\right] \tag{21}$$

where $c > 0, \alpha \in (0,1)$ are constants and

$$\bar{\delta}(\mu) \stackrel{\triangle}{=} \min_{t\geq 1}\left\{t\sqrt{\mu}\log^3\frac{1}{\mu} + \bar{\phi}(t,\mu)\log\frac{1}{\mu}\right\},$$

which tends to zero as μ approaches to zero.

The proof of *Theorem 4.4* is given in Section V. *Theorem 4.4* provides a good approximation of the mean square tracking error matrix $\boldsymbol{\Pi}_k^{\mathrm{CTA}}$ by $\widehat{\boldsymbol{\Pi}}_k^{\mathrm{CTA}}$ for small parameter variation γ and small adaptation gain μ, since $\bar{\delta}(\mu)$ tends to zero as μ

tends to zero. $\widehat{\mathbf{\Pi}}_k^{\rm CTA}$ can be further simplified in the wide-sense stationary case, as demonstrated in the following theorem.

Theorem 4.5: Let

$$\begin{aligned} \boldsymbol{F} &= \mathbb{E}[\boldsymbol{F}_k] = \mathrm{diag}\{\boldsymbol{F}^1, \cdots, \boldsymbol{F}^n\}, \text{ with } \boldsymbol{F}^i = \boldsymbol{F}^j \\ \boldsymbol{T} &= \mathbb{E}[\boldsymbol{T}_k] = \mathbb{E}[\boldsymbol{L}_k \boldsymbol{V}_k \boldsymbol{V}_k^T \boldsymbol{L}_k^T] \\ \boldsymbol{Q}_\omega &\equiv \boldsymbol{Q}_\omega(k+1). \end{aligned}$$

Under the conditions of *Theorem 4.4*, we have $k \to \infty$,

$$\mathbf{\Pi}_k^{\rm CTA} = \mu \bar{\boldsymbol{R}}_v + \frac{\gamma^2}{\mu} \bar{\boldsymbol{R}}_\omega + O\left(\bar{\delta}(\mu)\left[\mu + \frac{\gamma^2}{\mu}\right]\right) + o(1)$$

where the term $o(1)$ tends to 0 as $k \to \infty$, and

$$\begin{aligned} \bar{\boldsymbol{R}}_v &= \int_0^\infty e^{-\boldsymbol{F}t} \boldsymbol{A}_{\rm ave} \boldsymbol{T} \boldsymbol{A}_{\rm ave} e^{-\boldsymbol{F}t} dt \\ \bar{\boldsymbol{R}}_\omega &= \int_0^\infty e^{-\boldsymbol{F}t} \boldsymbol{A}_{\rm ave} \boldsymbol{Q}_\omega \boldsymbol{A}_{\rm ave} e^{-\boldsymbol{F}t} dt \end{aligned}$$

and $\boldsymbol{A}_{\rm ave} = (\lim_{k\to\infty} \mathcal{A}^k) \otimes I_m$ with $(\lim_{k\to\infty}\mathcal{A}^k)_{ij} = \frac{1}{n}$ for all $i, j = 1, \ldots, n$.

Remark 4.7: Since the proof of *Theorem 4.5* is similar to [20, Th. 15], here we omit it. Note that $\lim_{\mu\to 0} \bar{\delta}(\mu) = 0$. As a result, we have for all small μ and large k

$$\mathbf{\Pi}_k^{\rm CTA} \sim \mu \bar{\boldsymbol{R}}_v + \frac{\gamma^2}{\mu} \bar{\boldsymbol{R}}_\omega.$$

Consequently, by taking "trace," i.e., Tr$(\cdot)$, on both sides and noticing the definition of $\mathbf{\Pi}_k^{\rm CTA}$, we have

$$\sum_{i=1}^n \mathbb{E}[\|\widetilde{\boldsymbol{\theta}}_k^{i,\rm CTA}\|^2] \sim \mu \mathrm{Tr}(\bar{\boldsymbol{R}}_v) + \frac{\gamma^2}{\mu} \mathrm{Tr}(\bar{\boldsymbol{R}}_\omega) \qquad (22)$$

which indicates that μ should be proportional to γ, and by minimizing the right-hand side, we get the "optimal" choice

$$\mu^O = \gamma \sqrt{\mathrm{Tr}(\bar{\boldsymbol{R}}_\omega)/\mathrm{Tr}(\bar{\boldsymbol{R}}_v)}$$

with the corresponding minimum value

$$\sum_{i=1}^n \mathbb{E}[\|\widetilde{\boldsymbol{\theta}}_k^{i,\rm CTA}\|^2] \sim 2\gamma \sqrt{\mathrm{Tr}(\bar{\boldsymbol{R}}_\omega) \cdot \mathrm{Tr}(\bar{\boldsymbol{R}}_v)}.$$

Remark 4.8: In a similar way, one can show that *Theorem 4.4* and *Theorem 4.5* hold for $\{\widetilde{\mathbf{\Theta}}_k^{\rm ATC}, k \geq 1\}$ under the same conditions. In particular, $\widehat{\Pi}_k^{\rm ATC}$ provides a good approximation of the mean square tracking error matrix $\Pi_k^{\rm ATC}$, where

$$\begin{aligned} \mathbf{\Pi}_k^{\rm ATC} &= \mathbb{E}[\widetilde{\mathbf{\Theta}}_k^{\rm ATC} (\widetilde{\mathbf{\Theta}}_k^{\rm ATC})^T] \\ \widehat{\mathbf{\Pi}}_{k+1}^{\rm ATC} &= (I_{mn} - \mathbb{E}[\boldsymbol{B}_k]) \widehat{\mathbf{\Pi}}_k^{\rm ATC} (I_{mn} - \mathbb{E}[\boldsymbol{B}_k])^T \\ &\quad + \mathbf{\Lambda}^2 (\mathcal{A} \otimes I_m) \mathbb{E}[\boldsymbol{T}_k] (\mathcal{A} \otimes I_m) + \gamma^2 \boldsymbol{Q}_\omega(k+1) \end{aligned}$$

and

$$\boldsymbol{B}_k = \mathbf{\Lambda} \boldsymbol{F}_k + \mathscr{L} - \mathbf{\Lambda} \boldsymbol{F}_k \mathscr{L}.$$

B. Necessity of the Information Condition

In this section, we will further show that *Condition 4.2* used in this paper is not only sufficient but also necessary for the stability of the distributed algorithm under some extra conditions on dependence, for example, the ϕ-mixing condition. A random process $\{\xi_k\}$ is called ϕ-mixing, if there exists a sequence $\phi(n) \to 0$ as $n \to \infty$, such that

$$\sup_{A \in \mathcal{F}_{t+s}^\infty, B \in \mathcal{F}_0^t} |P(A|B) - P(A)| \leq \phi(s) \quad \forall t, s$$

where $\mathcal{F}_t^s = \sigma\{\xi(u), t \leq u \leq s\}$. As is well known, the ϕ-mixing process includes a large class of important processes, for examples, deterministic processes, M-dependent processes, and processes generated from bounded white noise filtered through a stable finite-dimensional linear filter (see [21]).

Theorem 4.6: Consider the model (1) and the estimation error (6). Let $\{\boldsymbol{\varphi}_k^i\}$ be ϕ-mixing processes and suppose that *Condition 4.1* is satisfied. Then, there exists a constant $\mu^* \in (0,1)$, such that for any $0 < \mathbf{\Lambda} \leq \mu^* I_{mn}$, $\{I_{mn} - [\mathbf{\Lambda}\boldsymbol{F}_k + \mathscr{L} - \mathbf{\Lambda}\boldsymbol{F}_k\mathscr{L}], k \geq 1\}$ is L_p-exponentially stable $(p \geq 1)$ **if and only if** *Condition 4.2* holds.

Remark 4.9: The detailed proof will be given in the following section. Note that the ϕ-mixing property used in the above theorem is just for simplicity, which can be further relaxed, see [34] for related discussions. Similarly, for the ATC diffusion NLMS algorithm, *Theorem 4.6* is also true.

V. Proofs of the Main Theorems

A. Proof of Theorem 4.1

Before proving the theorem, we first list and prove some lemmas. The first one is about Kronecker product.

Lemma 5.1 ([35]): Let the eigenvalues of matrices $X \in \mathbb{R}^{n\times n}$ and $Y \in \mathbb{R}^{m\times m}$ are $\lambda_i^X (i = 1, \ldots, n), \lambda_j^Y (j=1, \ldots, m)$, respectively. Then, the eigenvalues of matrix $X \otimes Y$ are $\lambda_i^X \lambda_j^Y, i = 1, \ldots, n, j = 1, \ldots, m$. Furthermore, if $x_1, \ldots, x_p$ are linearly independent right eigenvectors of X corresponding to $\lambda_1^X, \ldots, \lambda_p^X$ $(p \leq n)$ and $y_1, \ldots, y_q$ are linearly independent right eigenvectors of Y corresponding to $\lambda_1^Y, \ldots, \lambda_q^Y$ $(q \leq m)$, then $x_i \otimes y_j$ is a right eigenvector of $X \otimes Y$ corresponding to $\lambda_i^X \lambda_j^Y$, and $\{x_i \otimes y_j, i = 1, \ldots, n, j = 1, \ldots, m\}$ are independent.

The following three lemmas are all about the properties of S^0 defined by (11), which can be found in [21].

Lemma 5.2 ([21]): If two sequences α_k and β_k satisfy $0 \leq \alpha_k \leq \beta_k \leq 1$ $\forall k \geq 0$, then $\{\alpha_k\} \in S^0(\lambda)$ implies $\{\beta_k\} \in S^0(\lambda)$.

Lemma 5.3 ([21]): Let $\{\alpha_k\} \in S^0(\lambda)$ and $\alpha_k \leq \alpha^* < 1$ $\forall k \geq 0$ where α^* is a constant. Then, for any $\epsilon \in (0,1)$, $\{\epsilon\alpha_k\} \in S^0(\lambda^{(1-\alpha^*)\epsilon})$.

Lemma 5.4 ([21]): Let $\alpha = \{\alpha_k, \mathcal{F}_k\}$ and $\beta = \{\beta_k, \mathcal{F}_k\}$ be adapted processes, such that

$$\alpha_k \in [0,1], \quad \mathbb{E}[\alpha_{k+1}|\mathcal{F}_k] \geq \beta_k, \quad k \geq 0.$$

Then, $\{\beta\} \in S^0(\lambda)$ implies that $\{\alpha\} \in S^0(\sqrt{\lambda})$.

The next three lemmas improve the results on symmetric random matrices established in [21], and provide some further results on asymmetric matrices, which will be shown to be satisfied by the random matrices in (6) and (8), see Lemmas 5.8 and 5.9.

Lemma 5.5: Let $\{A_k\}$ be a sequence of random matrices which is adapted to $\{\mathcal{F}_k\}$, and there exist a constant $\varepsilon \in (0,1)$, such that $A_k^T A_k \le (1-\varepsilon)(A_k + A_k^T), a.s.$ For any fixed integer $h > 0$, denote

$$\gamma_k = \lambda_{\min}\left\{\mathbb{E}\left[\frac{1}{1+4(1-\varepsilon)h}\sum_{i=kh+1}^{(k+1)h}(A_i + A_i^T)|\mathcal{F}_{kh}\right]\right\} \quad (23)$$

then for $k = sh+1, s \ge 1$, we have

$$\lambda_{\max}\{\mathbb{E}[\Psi^T(k+h,k)\Psi(k+h,k)|\mathcal{F}_{k-1}]\} \le 1 - \frac{\varepsilon\gamma_s}{1+4(1-\varepsilon)h} \quad (24)$$

where $\Psi(\cdot,\cdot)$ is defined as

$$\Psi(t+1,s) = (I - A_t)\Psi(t,s),\ \ \Psi(s,s) = I\ \forall t \ge s.$$

Proof: For simplicity, we omit a.s. for sample paths in the following proof. By $A_k^T A_k \le (1-\varepsilon)(A_k + A_k^T)$, we know that for all $k \ge 0$,

$$\begin{aligned}\|A_k\|^2 &= \lambda_{\max}(A_k^T A_k)\\ &\le (1-\varepsilon)\lambda_{\max}(A_k + A_k^T)\\ &\le (1-\varepsilon)\|A_k + A_k^T\|\\ &\le 2(1-\varepsilon)\|A_k\|\end{aligned}$$

then we have $\|A_k\| \le 2(1-\varepsilon)$. From this, we know that

$$\gamma_k \in \left[0, \frac{4(1-\varepsilon)h}{1+4(1-\varepsilon)h}\right].$$

We denote z_{k-1} as the unit eigenvector corresponding to the largest eigenvalue ρ_{k-1} of the matrix $\mathbb{E}[\Psi^T(k+h,k)\Psi(k+h,k)|\mathcal{F}_{k-1}]$ and recursively define $\{z_j, j \ge k\}$ by

$$z_j = (I - A_j)z_{j-1},\quad j \ge k. \quad (25)$$

Then, $z_{k+h-1} = \Psi(k+h,k)z_{k-1}$. Hence, we have

$$\begin{aligned}&\mathbb{E}[\|z_{k+h-1}\|^2|\mathcal{F}_{k-1}]\\ &= z_{k-1}^T\mathbb{E}[\Psi^T(k+h,k)\Psi(k+h,k)|\mathcal{F}_{k-1}]z_{k-1}\\ &= \rho_{k-1}\|z_{k-1}\|^2\\ &= \rho_{k-1}. \end{aligned}\quad (26)$$

By (25), we have

$$z_j = z_{k-1} - \sum_{i=k}^{j} A_i z_{i-1}\quad \forall j \in [k, k+h-1].$$

Hence, by $A_k^T A_k \le (1-\varepsilon)(A_k + A_k^T)$ and the Cr-inequality,

$$\begin{aligned}&\mathbb{E}[\|z_{j-1} - z_{k-1}\|^2|\mathcal{F}_{k-1}]\\ &= \mathbb{E}\left[\left\|\sum_{i=k}^{j-1}A_i z_{i-1}\right\|^2|\mathcal{F}_{k-1}\right]\\ &\le \mathbb{E}\left[\left(\sum_{i=k}^{j-1}\|A_i z_{i-1}\|\right)^2|\mathcal{F}_{k-1}\right]\\ &\le (j-k)\mathbb{E}\left[\sum_{i=k}^{j-1}\|A_i z_{i-1}\|^2|\mathcal{F}_{k-1}\right]\\ &\le h\mathbb{E}\left[\sum_{i=k}^{j-1} z_{i-1}^T A_i^T A_i z_{i-1}|\mathcal{F}_{k-1}\right]\\ &\le h(1-\varepsilon)\mathbb{E}\left[\sum_{i=k}^{j-1} z_{i-1}^T(A_i + A_i^T)z_{i-1}|\mathcal{F}_{k-1}\right]\\ &\le h(1-\varepsilon)\mathbb{E}\left[\sum_{i=k}^{j-1}\|(A_i + A_i^T)^{1/2}z_{i-1}\|^2|\mathcal{F}_{k-1}\right]. \end{aligned}\quad (27)$$

By the definition of λ_s, the Minkowski inequality, $\|A_i + A_i^T\| \le 4(1-\varepsilon)$ and (27), we can obtain

$$\begin{aligned}&\sqrt{[1+4(1-\varepsilon)h]\gamma_s}\\ &\le \left\{z_{k-1}^T\mathbb{E}\left[\sum_{i=sh+1}^{(s+1)h}(A_i + A_i^T)|\mathcal{F}_{sh}\right]z_{k-1}\right\}^{\frac{1}{2}}\\ &\le \left\{\mathbb{E}\left[\sum_{i=k}^{k+h-1}\|(A_i + A_i^T)^{1/2}z_{k-1}\|^2|\mathcal{F}_{k-1}\right]\right\}^{\frac{1}{2}}\\ &\le \left\{\mathbb{E}\left[\sum_{i=k}^{k+h-1}\|(A_i + A_i^T)^{1/2}z_{i-1}\|^2|\mathcal{F}_{k-1}\right]\right\}^{\frac{1}{2}}\\ &\quad + \left\{\mathbb{E}\left[\sum_{i=k}^{k+h-1}\|(A_i + A_i^T)^{1/2}(z_{i-1} - z_{k-1})\|^2|\mathcal{F}_{k-1}\right]\right\}^{\frac{1}{2}}\\ &\le [1 + \sqrt{4(1-\varepsilon)h}\cdot\sqrt{h(1-\varepsilon)}]\\ &\quad \times \left\{\mathbb{E}\left[\sum_{i=k}^{k+h-1}\|(A_i + A_i^T)^{1/2}z_{i-1}\|^2|\mathcal{F}_{k-1}\right]\right\}^{\frac{1}{2}}\\ &< [1+4(1-\varepsilon)h]\\ &\quad \times \left\{\mathbb{E}\left[\sum_{i=k}^{k+h-1}\|(A_i + A_i^T)^{1/2}z_{i-1}\|^2|\mathcal{F}_{k-1}\right]\right\}^{\frac{1}{2}}. \end{aligned}\quad (28)$$

Then, we have

$$\mathbb{E}\left[\sum_{i=k}^{k+h-1} z_{i-1}^T(A_i + A_i^T)z_{i-1}|\mathcal{F}_{k-1}\right] \ge \frac{\gamma_s}{1+4(1-\varepsilon)h}. \quad (29)$$

By (25) again, we have

$$\begin{aligned}&\|z_{k+h-1}\|^2\\ &= z_{k+h-2}^T(I - A_{k+h-1})^T(I - A_{k+h-1})z_{k+h-2}\\ &= z_{k+h-2}^T(I + A_{k+h-1}^T A_{k+h-1} - A_{k+h-1} - A_{k+h-1}^T)z_{k+h-2}\\ &\le \|z_{k+h-2}\|^2 - \varepsilon z_{k+h-2}^T(A_{k+h-1} + A_{k+h-1}^T)z_{k+h-2}\end{aligned}$$

$$\leq \|z_{k-1}\|^2 - \varepsilon \sum_{i=k}^{k+h-1} z_{i-1}^T (A_i + A_i^T) z_{i-1}$$
$$\leq 1 - \varepsilon \sum_{i=k}^{k+h-1} z_{i-1}^T (A_i + A_i^T) z_{i-1}.$$

Combining this with (26) and (28), we can obtain

$$\begin{aligned}\rho_{k-1} &= \mathbb{E}[\|z_{k+h-1}\|^2 | \mathcal{F}_{k-1}] \\ &\leq 1 - \varepsilon \mathbb{E}\left[\sum_{i=k}^{k+h-1} z_{i-1}^T (A_i + A_i^T) z_{i-1} | \mathcal{F}_{k-1}\right] \\ &\leq 1 - \frac{\varepsilon \gamma_s}{1 + 4(1-\varepsilon)h}.\end{aligned}$$

This completes the proof. ■

Lemma 5.6: Under the same conditions and notations of *Lemma 5.5*, consider the equation

$$x_k = \Psi(kh+1, (k-1)h+1) x_{k-1}, \quad k \geq k_0 + 1 \tag{30}$$

where x_{k_0} is deterministic and satisfies $\|x_{k_0}\| = 1$. Then, there exists a sequence $\{\alpha_k \in [0,1], k \geq k_0 + 1\}$ such that $\alpha_k \in \mathcal{F}_{kh}$, and

$$\|x_k\| \leq (1 - \alpha_k) \|x_{k-1}\|, \quad k \geq k_0 + 1 \tag{31}$$

and

$$\mathbb{E}[\alpha_{k+1} | \mathcal{F}_{kh}] \geq \frac{\varepsilon \gamma_k}{2[1 + 4(1-\varepsilon)h]}, \quad k \geq k_0 + 1. \tag{32}$$

Proof: For simplicity, we omit a.s. for sample paths in the following proof. Let us set for any $k \geq k_0 + 1$,

$$\alpha_k = \begin{cases} 1 - \frac{\|\Psi(kh+1,(k-1)h+1)x_{k-1}\|}{\|x_{k-1}\|}, & \text{if } \|x_{k-1}\| \neq 0 \\ 1, & \text{otherwise.} \end{cases} \tag{33}$$

By the conditions in *Lemma 5.5* and $\|A_k\| \leq 2(1-\varepsilon)$, we know that for any $k \geq 0$,

$$\begin{aligned}\|I - A_k\|^2 &= \lambda_{\max}[(I - A_k)^T (I - A_k)] \\ &= \lambda_{\max}(I + A_k^T A_k - A_k - A_k^T) \\ &\leq \lambda_{\max}[I - \varepsilon(A_k + A_k^T)] \\ &= 1 - \varepsilon \lambda_{\min}(A_k + A_k^T) \leq 1\end{aligned}$$

then we have $\|\Psi(t,s)\| \leq 1 \; \forall t \geq s \geq 0$. It is clear that $\alpha_k \in [0,1], \alpha_k \in \mathcal{F}_{kh}$ and (31) is true.

Now, consider the set $\Omega_k = \{\omega : \|x_k\| = 0\}$. Then, $\Omega_k \in \mathcal{F}_{kh}$ and

$$I_{\Omega_k} \mathbb{E}[\alpha_{k+1} | \mathcal{F}_{kh}] = \mathbb{E}[I_{\Omega_k} \alpha_{k+1} | \mathcal{F}_{kh}] = I_{\Omega_k}.$$

Hence, by noting $\gamma_k < 1$ and $\varepsilon \in (0,1)$, we know that (32) is true on the set Ω_k. Then, consider the set Ω_k^c, we have by *Lemma 5.5*

$$\begin{aligned}&\mathbb{E}[\|\Psi((k+1)h+1, kh+1)x_k\| | \mathcal{F}_{kh}] \\ &\leq \{\mathbb{E}[\|\Psi((k+1)h+1, kh+1)x_k\|^2 | \mathcal{F}_{kh}]\}^{1/2} \\ &\leq \{x_k^T \mathbb{E}[\Psi^T((k+1)h+1, kh+1). \\ &\quad \Psi((k+1)h+1, kh+1) | \mathcal{F}_{kh}] x_k\}^{1/2} \\ &\leq \left\{x_k^T \left[1 - \frac{\varepsilon \gamma_k}{1 + 4(1-\varepsilon)h}\right] x_k\right\}^{1/2} \\ &\leq \left\{1 - \frac{\varepsilon \gamma_k}{2[1 + 4(1-\varepsilon)h]}\right\} \|x_k\|.\end{aligned}$$

Consequently, by (33) we have

$$\begin{aligned}I_{\Omega_k^c} \mathbb{E}[\alpha_{k+1} | \mathcal{F}_{kh}] &\geq I_{\Omega_k^c} \left\{1 - \left(1 - \frac{\varepsilon \gamma_k}{2[1 + 4(1-\varepsilon)h]}\right)\right\} \\ &= \frac{\varepsilon \gamma_k}{2[1 + 4(1-\varepsilon)h]} I_{\Omega_k^c}.\end{aligned}$$

Hence, (32) is also true on the set Ω_k^c. This completes the proof. ■

Lemma 5.7: Let $\{A_k\}$ be a sequence of random matrices which is adapted to $\{\mathcal{F}_k\}$, and there exist a constant $\varepsilon \in (0,1)$, such that $A_k^T A_k \leq (1-\varepsilon)(A_k + A_k^T), a.s.$ If there exists an integer $h > 0$ such that $\{\gamma_k\} \in S^0(\gamma), \gamma \in (0,1)$ where γ_k is defined by (23), then $\{A_k\} \in S_p(\gamma^\alpha)$, where

$$\alpha = \begin{cases} \frac{\varepsilon}{8h[1+4(1-\varepsilon)h]^2}, & 1 \leq p \leq 2; \\ \frac{\varepsilon}{4h[1+4(1-\varepsilon)h]^2 p}, & p > 2. \end{cases} \tag{34}$$

Proof: Now, for any $t > s + h$, let us define

$$k_0 = \min\{k : s \leq kh + 1 \leq t\}$$
$$k_1 = \max\{k : s \leq kh + 1 \leq t\}.$$

Then, it is clear that

$$(k_1 + 1)h + 1 > t, \; (k_0 - 1)h + 1 < s$$

and

$$\mathbb{E}[\|\Psi(t,s)\|^2] \leq \mathbb{E}[\|\Psi(k_1 h + 1, k_0 h + 1)\|^2]$$

where $\Psi(t,s)$ is defined in *Lemma 5.5*. Hence, for $\{A_i\} \in S_2(\gamma^\beta)$ and $\beta = \frac{\varepsilon}{8h[1+4(1-\varepsilon)h]^2}$, it suffices to find a constant c which is free of k_1 and k_0 such that, for all $k_1 \geq k_0$,

$$\mathbb{E}[\|\Psi(k_1 h + 1, k_0 h + 1)\|^2] \leq c \gamma^{2\beta h (k_1 - k_0 + 1)}. \tag{35}$$

Now, consider (30), we have

$$x_{k_1} = \Psi(k_1 h + 1, k_0 h + 1) x_{k_0}.$$

To prove this, we need only to prove that for any deterministic x_{k_0} with $\|x_{k_0}\| = 1$,

$$\mathbb{E}[\|x_{k_1}\|^2] \leq c \gamma^{2h\beta(k_1 - k_0)} \tag{36}$$

where c is independent of k_0, k_1, and x_{k_0}.

Since $\gamma_k \in [0, \frac{4(1-\varepsilon)h}{1+4(1-\varepsilon)h}]$ and $\{\gamma_k\} \in S^0(\gamma), \gamma \in (0,1)$, then by *Lemma 5.3* we know that

$$\left\{\frac{\varepsilon \gamma_k}{2[1 + 4(1-\varepsilon)h]}\right\} \in S^0(\gamma^{4h\beta}).$$

From this, (32), *Lemma 5.4* and its proof in [21], we know that

$$\mathbb{E}\left[\prod_{k=k_0+1}^{k_1} (1 - \alpha_k)\right] \leq c \gamma^{2h\beta(k_1 - k_0)}$$

for some constant c independent of k_0, k_1 and x_{k_0}. Then, by (31) we know that (36) is true. Hence, $\{A_k\} \in S_2(\gamma^\beta)$.

For $1 \le p \le 2$, we use the monotonicity of the norm $\|\cdot\|_{L_p}$, while for $p > 2$, we apply the inequality $\|I - A_j\| \le 1$, then we have

$$\left\| \prod_{j=i+1}^{k} (I - A_j) \right\|_{L_p} \le \begin{cases} \left\| \prod_{j=i+1}^{k} (I - A_j) \right\|_{L_2}, & 1 \le p \le 2 \\ \left\| \prod_{j=i+1}^{k} (I - A_j) \right\|_{L_2}^{2/p}, & p > 2. \end{cases}$$

Consequently, $\{A_k\} \in S_p(\gamma^\alpha)$ where α is defined in (34). This completes the proof. ∎

Remark 5.1: Note that *Lemma 5.7* will be used to connect the *Condition 4.2* with the L_p-exponential stability of the homogeneous part of the error (6).

In the following two lemmas, we will prove that the random matrices $\{\boldsymbol{\Lambda F}_k + \mathscr{L} - \boldsymbol{\Lambda F}_k\mathscr{L}, k \ge 0\}$ and $\{\boldsymbol{\Lambda F}_k + \mathscr{L} - \boldsymbol{\Lambda}\mathscr{L}\boldsymbol{F}_k, k \ge 0\}$ satisfy the inequality in *Lemma 5.5*.

Lemma 5.8: Consider the distributed filtering error (6), and denote $\boldsymbol{A}_k = \boldsymbol{\Lambda F}_k + \mathscr{L} - \boldsymbol{\Lambda F}_k\mathscr{L}$. Then under *Condition 4.1*, there exist constants $\mu^* \in (0,1)$ and $\varepsilon \in (0,1)$, such that for any $0 < \boldsymbol{\Lambda} \le \mu^* I_{mn}$,

$$\boldsymbol{A}_k^T \boldsymbol{A}_k \le (1-\varepsilon)(\boldsymbol{A}_k + \boldsymbol{A}_k^T), \quad \text{a.s.} \tag{37}$$

Proof: By the symmetrical and doubly stochastic property of the matrix $\mathcal{A}$ and *Lemma 5.1*, we know that there exists a constant $\delta \in (0,1)$ such that $0 \le \mathscr{L} \le 2(1-\delta)I_{mn}$. In addition, we know that $0 \le \boldsymbol{F}_k \le I_{mn}$.

For simplicity, we omit the subscript k, the dimension mn, and $a.s.$ for sample paths hereafter. We first have

$$\begin{aligned} \boldsymbol{A}^T\boldsymbol{A} &= (\boldsymbol{\Lambda F} + \mathscr{L} - \boldsymbol{\Lambda F}\mathscr{L})^T(\boldsymbol{\Lambda F} + \mathscr{L} - \boldsymbol{\Lambda F}\mathscr{L}) \\ &= \boldsymbol{\Lambda}^2\boldsymbol{F}^2 + \boldsymbol{\Lambda F}\mathscr{L} - \boldsymbol{\Lambda}^2\boldsymbol{F}^2\mathscr{L} + \mathscr{L}\boldsymbol{F\Lambda} + \mathscr{L}^2 \\ &\quad - 2\mathscr{L}\boldsymbol{\Lambda F}\mathscr{L} - \mathscr{L}\boldsymbol{F}^2\boldsymbol{\Lambda}^2 + \mathscr{L}\boldsymbol{\Lambda}^2\boldsymbol{F}^2\mathscr{L}. \end{aligned}$$

Since $\mathscr{L}\boldsymbol{\Lambda}^2\boldsymbol{F}^2\mathscr{L} - 2\mathscr{L}\boldsymbol{\Lambda F}\mathscr{L} \le 0$, we can obtain

$$\begin{aligned} \boldsymbol{A}^T\boldsymbol{A} &\le \boldsymbol{\Lambda}^2\boldsymbol{F}^2 + \mathscr{L}^2 + \boldsymbol{\Lambda F}\mathscr{L} + \mathscr{L}\boldsymbol{F\Lambda} \\ &\quad - \boldsymbol{\Lambda}^2\boldsymbol{F}^2\mathscr{L} - \mathscr{L}\boldsymbol{F}^2\boldsymbol{\Lambda}^2. \end{aligned}$$

Also, we have

$$\boldsymbol{A} + \boldsymbol{A}^T = 2(\boldsymbol{\Lambda F} + \mathscr{L}) - (\boldsymbol{\Lambda F}\mathscr{L} + \mathscr{L}\boldsymbol{F\Lambda}).$$

From this, to prove (37), we need only to prove that there exist constants $\mu^* \in (0,1)$ and $\varepsilon \in (0,1)$, such that for any $0 < \boldsymbol{\Lambda} \le \mu^* I_{mn}$,

$$\begin{aligned} \boldsymbol{\Lambda}^2\boldsymbol{F}^2 + \mathscr{L}^2 &\le 2(1-\varepsilon)(\boldsymbol{\Lambda F} + \mathscr{L}) + (\boldsymbol{\Lambda}^2\boldsymbol{F}^2\mathscr{L} + \mathscr{L}\boldsymbol{F}^2\boldsymbol{\Lambda}^2) \\ &\quad - (2-\varepsilon)(\boldsymbol{\Lambda F}\mathscr{L} + \mathscr{L}\boldsymbol{F\Lambda}). \end{aligned} \tag{38}$$

Denote $\mu_1 = 2(1-\delta)$, since $0 \le \boldsymbol{F}_k \le I$, then for any $0 < \boldsymbol{\Lambda} \le \mu_1 I_{mn}$, we have

$$\begin{aligned} \boldsymbol{\Lambda}^2\boldsymbol{F}^2 + \mathscr{L}^2 &\le \boldsymbol{\Lambda}^2\boldsymbol{F} + 2(1-\delta)\mathscr{L} \le 2(1-\delta)(\boldsymbol{\Lambda F} + \mathscr{L}) \\ &= 2\left(1 - \frac{\delta}{2}\right)(\boldsymbol{\Lambda F} + \mathscr{L}) - \delta(\boldsymbol{\Lambda F} + \mathscr{L}). \end{aligned}$$

Now, we choose $\varepsilon = \frac{\delta}{2}$. To prove (38), we need only to prove that there exists a constant $\mu^* \in (0,1)$, such that for any $0 < \boldsymbol{\Lambda} \le \mu^* I_{mn}$,

$$\begin{aligned} &(2-\varepsilon)(\boldsymbol{\Lambda F}\mathscr{L} + \mathscr{L}\boldsymbol{F\Lambda}) - (\boldsymbol{\Lambda}^2\boldsymbol{F}^2\mathscr{L} + \mathscr{L}\boldsymbol{F}^2\boldsymbol{\Lambda}^2) \\ &\le \delta(\boldsymbol{\Lambda F} + \mathscr{L}). \end{aligned} \tag{39}$$

In fact, for any mn-dimensional unit column vector x with $\|x\| = 1$, we have by noting $\|\boldsymbol{F}\| \le 1$ and $\|\mathscr{L}\| \le 2$,

$$\begin{aligned} &x^T[(2-\varepsilon)(\boldsymbol{\Lambda F}\mathscr{L} + \mathscr{L}\boldsymbol{F\Lambda}) - (\boldsymbol{\Lambda}^2\boldsymbol{F}^2\mathscr{L} + \mathscr{L}\boldsymbol{F}^2\boldsymbol{\Lambda}^2)]x \\ &= 2(2-\varepsilon)x^T\boldsymbol{\Lambda F}\mathscr{L}x - 2x^T\boldsymbol{\Lambda}^2\boldsymbol{F}^2\mathscr{L}x \\ &\le 2(2-\varepsilon)\|x^T\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{F}\boldsymbol{\Lambda}^{\frac{1}{2}}\| \cdot \|\mathscr{L}x\| \\ &\quad + 2\|x^T\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{F}^2\boldsymbol{\Lambda}^{\frac{3}{2}}\| \cdot \|\mathscr{L}x\| \\ &\le 2(2-\varepsilon)\|x^T\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{F}^{\frac{1}{2}}\| \cdot \|\boldsymbol{F}^{\frac{1}{2}}\| \cdot \|\boldsymbol{\Lambda}^{\frac{1}{2}}\| \cdot \|\mathscr{L}^{\frac{1}{2}}\| \cdot \|\mathscr{L}^{\frac{1}{2}}x\| \\ &\quad + 2\|x^T\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{F}^{\frac{1}{2}}\| \cdot \|\boldsymbol{F}^{\frac{3}{2}}\| \cdot \|\boldsymbol{\Lambda}^{\frac{3}{2}}\| \cdot \|\mathscr{L}^{\frac{1}{2}}\| \cdot \|\mathscr{L}^{\frac{1}{2}}x\| \\ &\le \sqrt{2}(2-\varepsilon)\|\boldsymbol{\Lambda}^{\frac{1}{2}}\|(2\|x^T\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{F}^{\frac{1}{2}}\| \cdot \|\mathscr{L}^{\frac{1}{2}}x\|) \\ &\quad + \sqrt{2}\|\boldsymbol{\Lambda}^{\frac{3}{2}}\|(2\|x^T\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{F}^{\frac{1}{2}}\| \cdot \|\mathscr{L}^{\frac{1}{2}}x\|) \\ &\le \sqrt{2}\Big[(2-\varepsilon)\|\boldsymbol{\Lambda}^{\frac{1}{2}}\| + \|\boldsymbol{\Lambda}^{\frac{3}{2}}\|\Big](x^T\boldsymbol{\Lambda F}x + x^T\mathscr{L}x). \end{aligned} \tag{40}$$

Here, if we choose μ to satisfy $\sqrt{2}(2-\varepsilon)\|\boldsymbol{\Lambda}^{\frac{1}{2}}\| \le \frac{\delta}{2}$ and $\sqrt{2}\|\boldsymbol{\Lambda}^{\frac{3}{2}}\| \le \frac{\delta}{2}$, then (39) holds. Hence, we can choose

$$\mu^* = \min\left\{2(1-\delta), \frac{\delta^2}{2(4-\delta)^2}, \frac{\delta^{2/3}}{2}\right\} \tag{41}$$

where $\delta \in (0,1)$ is a constant which is related to the Laplacian matrix $\mathcal{L}$ of the network graph. Consequently, there exists a constant $\mu^* \in (0,1)$ such that for any $0 < \boldsymbol{\Lambda} \le \mu^* I_{mn}$, (37) holds. This completes the proof. ∎

In a similar way, one can prove the following lemma.

Lemma 5.9: Consider the distributed filtering error (8), denote $\boldsymbol{B}_k = \boldsymbol{\Lambda F}_k + \mathscr{L} - \boldsymbol{\Lambda}\mathscr{L}\boldsymbol{F}_k$. Then, there exist constants $\mu^* \in (0,1)$ and $\varepsilon \in (0,1)$, such that for any $0 < \boldsymbol{\Lambda} \le \mu^* I_{mn}$,

$$\boldsymbol{B}_k^T\boldsymbol{B}_k \le (1-\varepsilon)(\boldsymbol{B}_k + \boldsymbol{B}_k^T), \quad \text{a.s.} \tag{42}$$

To accomplish the proof of *Theorem 4.1*, we need the following lemma.

Lemma 5.10: Denote $\boldsymbol{A}_k = \boldsymbol{\Lambda F}_k + \mathscr{L} - \boldsymbol{\Lambda F}_k\mathscr{L}$ and $\sigma_{\min} = \min\{\sigma_1, \ldots, \sigma_n\}$, and suppose that *Conditions 4.1 and 4.2* are satisfied. Then, there exist constants $\mu^* \in (0,1)$ and $\varepsilon \in (0,1)$, such that for any $0 < \boldsymbol{\Lambda} \le \mu^* I_{mn}$,, we have $\gamma_k \in S^0(\gamma)$, where

$$\gamma_k \triangleq \lambda_{\min}\left\{\mathbb{E}\left[\frac{1}{1+4(1-\varepsilon)h}\sum_{j=k+1}^{k+h}(\boldsymbol{A}_j + \boldsymbol{A}_j^T)\Big|\mathcal{F}_k\right]\right\} \tag{43}$$

and $\gamma = \lambda^\nu, \nu = \frac{0.5hl_{m+1}\sigma_{\min}\mu^*}{(2+l_{m+1})(1+h)[1+4(1-\varepsilon)h]}$, l_{m+1} is the $(m+1)$th smallest eigenvalue of matrix $\mathscr{L}$, which equals to the second smallest eigenvalue of matrix $\mathcal{L}$.

Remark 5.2: The detailed proof is given in Appendix A. Note that the constant ν determines the rate of exponential convergence of the homogeneous part of (6). Note also that λ of *Conditions 4.2* can be regarded as a measure of the cooperativity of the system information, and that l_{m+1} can be regarded as a measure of the connectivity of the graph $\mathcal{G}$; hence, the formula $\gamma = \lambda^\nu$ shows explicitly how the stability of the distributed algorithms is connected with the cooperativity of the system information, the connectivity of the network topology, as well as the step-size.

Proof of Theorem 4.1: By *Lemmas 5.7* and *5.10*, we know that there exists a constant $\mu^* \in (0,1)$, such that for any $0 < \mathbf{\Lambda} \leq \mu^* I_{mn}$,

$$\{\mathbf{\Lambda F}_k + \mathscr{L} - \mathbf{\Lambda F}_k \mathscr{L}\} \in S_p(\gamma^\alpha)$$

where α is defined by (34). Then by *Definition 3.2*, it is obvious that $\{I_{mn} - (\mathbf{\Lambda F}_k + \mathscr{L} - \mathbf{\Lambda F}_k \mathscr{L}), k \geq 1\}$ is L_p-exponentially stable $(p \geq 1)$.

Remark 5.3: Since *Lemma 5.10* also holds for random matrices $\mathbf{\Lambda F}_k + \mathscr{L} - \mathbf{\Lambda} \mathscr{L} \mathbf{F}_k$ in the ATC diffusion NLMS algorithm, we know that $\{I_{mn} - (\mathbf{\Lambda F}_k + \mathscr{L} - \mathbf{\Lambda} \mathscr{L} \mathbf{F}_k), k \geq 1\}$ is also L_p-exponentially stable $(p \geq 1)$.

B. Proof of Theorem 4.3

Before establishing further performance results, we first have the following lemma.

Lemma 5.11: Suppose that *Conditions 4.1* and 4.2 are satisfied. Then for any $p \geq 2$, there exist constants $\mu^* \in (0,1), \varepsilon \in (0,1)$ and $M > 0$, such that for all $0 < \mathbf{\Lambda} \leq \mu^* I_{mn}$ and $\forall k \geq i+1 \geq 0$

$$\left\| \prod_{j=i+1}^{k} (I_{mn} - [\mathbf{\Lambda F}_j + \mathscr{L} - \mathbf{\Lambda F}_j \mathscr{L}]) \right\|_{L_p} \leq M_p (1 - \mu^* \alpha_p)^{k-i} \tag{44}$$

where M_p and α_p are positive constants depending on $\{\mathbf{F}_j, j > 0\}, \mathscr{L}$ and p.

Proof: Denote

$$b = \frac{0.5 h l_{m+1} \sigma_{\min}}{(2 + l_{m+1})(1+h)[1 + 4(1-\varepsilon)h]}.$$

By the proof of *Theorem 4.1*, we have

$$\{\mathbf{\Lambda F}_k + \mathscr{L} - \mathbf{\Lambda F}_k \mathscr{L}\} \in S_p(\lambda^{b\alpha\mu^*})$$

where α is defined in *Lemma 5.7*. Then by the definition of S_p, we know that

$$\left\| \prod_{j=i+1}^{k} (I_{mn} - [\mathbf{\Lambda F}_j + \mathscr{L} - \mathbf{\Lambda F}_j \mathscr{L}]) \right\|_{L_p} \leq M \{\beta_p\}^{\mu^*(k-i)}$$

where

$$\beta_p = \lambda^{\frac{b\varepsilon}{4h[1+4(1-\varepsilon)h]^2 p}}$$

and l_{m+1} is the second smallest eigenvalue of matrix $\mathcal{L}$.

Note that $\beta_p \in (0,1)$. Then, there exists a constant $\alpha_p \in (0,1)$ such that $\alpha_p = 1 - \beta_p = 1 - \lambda^{\frac{b\varepsilon}{4h[1+4(1-\varepsilon)h]^2 p}}$. According to the Bernoulli inequality and since $\mu^* \in (0,1)$, we have $\beta_p^{\mu^*} = (1-\alpha_p)^{\mu^*} < 1 - \mu^* \alpha_p$. This completes the proof. ■

The following proof of *Theorem 4.3* is similar to [19, Th. 2.2], and we omit it here.

C. Proof of Theorem 4.4

Before proving the theorem, we first give the following lemma.

Lemma 5.12: Let $\{\mathbf{A}_k\}$ be a sequence of random matrices which is adapted to $\{\mathcal{F}_k\}$, and there exists a constant $\varepsilon \in (0,1)$, such that $\mathbf{A}_k^T \mathbf{A}_k \leq (1-\varepsilon)(\mathbf{A}_k + \mathbf{A}_k^T)$, a.s. If $\{\mathbf{A}_k\} \in S_1(\lambda)$ for some $\lambda \in [0,1)$, then there exists an integer $h > 0$ such that

$$\inf_k \lambda_{\min} \left\{ \sum_{j=kh+1}^{(k+1)h} \mathbb{E}[\mathbf{A}_j + \mathbf{A}_j^T] \right\} > 0. \tag{45}$$

Proof: For simplicity, we omit a.s. for sample paths in the following proof. By the conditions, we know that there exists a suitably large integer $h \geq 2$ such that

$$\mathbb{E}\left\| \prod_{j=k+1}^{k+h} (I_{mn} - \mathbf{A}_j) \right\| \leq M\lambda^h < \frac{1}{2} \quad \forall k \geq 0. \tag{46}$$

By *Lemma 5.8* we know that $\|\mathbf{A}_k\| \leq 2(1-\varepsilon) \quad \forall k \geq 0$. Let us denote

$$\gamma_k = \lambda_{\min}\left\{ \mathbb{E}\left[\sum_{j=k+1}^{k+h} (\mathbf{A}_j + \mathbf{A}_j^T) \right] \right\}$$

and let $\boldsymbol{\alpha}_k$ be the unit eigenvector corresponding to γ_k. Then, we have $\gamma_k = \mathbb{E}\left[\sum_{j=k+1}^{k+h} \boldsymbol{\alpha}_k^T (\mathbf{A}_j + \mathbf{A}_j^T) \boldsymbol{\alpha}_k\right]$.

Hence, if $t \geq 2$, we have for any integers $j_s \in [k+1, k+h], s = 1, \ldots, t, t \leq h$,

$$\begin{aligned}
&\mathbb{E}[\boldsymbol{\alpha}_k^T \mathbf{A}_{j_1} \cdot \cdots \cdot \mathbf{A}_{j_t} \boldsymbol{\alpha}_k] \\
&\leq \mathbb{E}[\|\boldsymbol{\alpha}_k^T \mathbf{A}_{j_1}\| \cdot \|\mathbf{A}_{j_2} \cdots \mathbf{A}_{j_{t-1}}\| \cdot \|\mathbf{A}_{j_t} \boldsymbol{\alpha}_k\|] \\
&\leq [2(1-\varepsilon)]^{t-2} \mathbb{E}[\|\boldsymbol{\alpha}_k^T \mathbf{A}_{j_1}\| \cdot \|\mathbf{A}_{j_t} \boldsymbol{\alpha}_k\|] \\
&\leq [2(1-\varepsilon)]^{t-2} \{\mathbb{E}[\|\boldsymbol{\alpha}_k^T \mathbf{A}_{j_1}\|^2] \cdot \mathbb{E}[\|\mathbf{A}_{j_t} \boldsymbol{\alpha}_k\|^2]\}^{\frac{1}{2}} \\
&= [2(1-\varepsilon)]^{t-2} \{\mathbb{E}[\boldsymbol{\alpha}_k^T \mathbf{A}_{j_1}^T \mathbf{A}_{j_1} \boldsymbol{\alpha}_k] \cdot \mathbb{E}[\boldsymbol{\alpha}_k^T \mathbf{A}_{j_t}^T \mathbf{A}_{j_t} \boldsymbol{\alpha}_k]\}^{\frac{1}{2}} \\
&\leq [2(1-\varepsilon)]^{t-2} \max_{k+1 \leq j \leq k+h} \mathbb{E}[\boldsymbol{\alpha}_k^T \mathbf{A}_j^T \mathbf{A}_j \boldsymbol{\alpha}_k] \\
&\leq [2(1-\varepsilon)]^{t-2} (1-\varepsilon) \max_{k+1 \leq j \leq k+h} \mathbb{E}[\boldsymbol{\alpha}_k^T (\mathbf{A}_j^T + \mathbf{A}_j) \boldsymbol{\alpha}_k] \\
&\leq [2(1-\varepsilon)]^{t-2} (1-\varepsilon) \gamma_k
\end{aligned}$$

where we have used *Lemma 5.8*. If $t = 1$, we have for any integer $j_1 \in [k+1, k+h]$,

$$\begin{aligned}
&\mathbb{E}[\boldsymbol{\alpha}_k^T \mathbf{A}_{j_1} \boldsymbol{\alpha}_k] \leq \mathbb{E}[\|\boldsymbol{\alpha}_k^T\| \cdot \|\mathbf{A}_{j_1} \boldsymbol{\alpha}_k\|] \leq \mathbb{E}\|\mathbf{A}_{j_1} \boldsymbol{\alpha}_k\| \\
&\leq \{\mathbb{E}\|\mathbf{A}_{j_1} \boldsymbol{\alpha}_k\|^2\}^{\frac{1}{2}} \leq \max_{k+1 \leq j \leq k+h} \{\mathbb{E}[\boldsymbol{\alpha}_k^T \mathbf{A}_j^T \mathbf{A}_j \boldsymbol{\alpha}_k]\}^{\frac{1}{2}} \\
&\leq (1-\varepsilon)^{\frac{1}{2}} \max_{k+1 \leq j \leq k+h} \{\mathbb{E}[\boldsymbol{\alpha}_k^T (\mathbf{A}_j^T + \mathbf{A}_j) \boldsymbol{\alpha}_k]\}^{\frac{1}{2}} \\
&\leq (1-\varepsilon)^{\frac{1}{2}} \gamma_k^{\frac{1}{2}}.
\end{aligned}$$

Consequently, by (46) we have

$$\begin{aligned}
\frac{1}{2} &> \mathbb{E}\left[\boldsymbol{\alpha}_k^T \left\{ \prod_{j=k+1}^{k+h} (I_{mn} - \mathbf{A}_j) \right\} \boldsymbol{\alpha}_k \right] \\
&= 1 - \sum_{k+1 \leq j_1 \leq k+h} \mathbb{E}[\boldsymbol{\alpha}_k^T \mathbf{A}_{j_1} \boldsymbol{\alpha}_k] \\
&\quad - \sum_{t=2}^{h} \sum_{k+1 \leq j_1 < \ldots < j_t \leq k+h} \mathbb{E}[\boldsymbol{\alpha}_k^T \mathbf{A}_{j_1} \cdots \mathbf{A}_{j_t} \boldsymbol{\alpha}_k] \\
&\geq 1 - h(1-\varepsilon)^{\frac{1}{2}} \gamma_k^{\frac{1}{2}}
\end{aligned}$$

$$-\sum_{2=1}^{h}\sum_{k+1\le j_1<\ldots<j_t\le k+h}[2(1-\varepsilon)]^{t-2}(1-\varepsilon)\gamma_k$$

$$\geq 1-h(1-\varepsilon)^{\frac{1}{2}}\gamma_k^{\frac{1}{2}}-\sum_{t=2}^{h}\binom{h}{t}[2(1-\varepsilon)]^{t-2}(1-\varepsilon)\gamma_k.$$

Let us denote

$$a_1=\sum_{t=2}^{h}\binom{h}{t}[2(1-\varepsilon)]^{t-2}(1-\varepsilon)>0,\ a_2=h(1-\varepsilon)^{\frac{1}{2}}>0$$

then we have $a_1\gamma_k+a_2\sqrt{\gamma_k}\geq\frac{1}{2}$, which implies that

$$\gamma_k^{\frac{1}{2}}\geq\frac{\sqrt{a_2^2+2a_1}-a_2}{2a_1}\ \forall k$$

or

$$\gamma_k\geq\left(\frac{\sqrt{a_2^2+2a_1}-a_2}{2a_1}\right)^2\ \forall k. \tag{47}$$

Hence, *Lemma 5.12* is true. ■

Lemma 5.13: Suppose that *Conditions 4.1* and 4.2 are satisfied. Then, there exist constants $M>0$, $\beta\in(0,1)$ and $\mu^*\in(0,1)$, such that for any $0<\mathbf{\Lambda}\leq\mu^* I_{mn}$ and $\forall k\geq i\geq 0$

$$\left\|\prod_{j=i+1}^{k}(I_{mn}-\mathbb{E}[\mathbf{\Lambda}\boldsymbol{F}_j+\mathscr{L}-\mathbf{\Lambda}\boldsymbol{F}_j\mathscr{L}])\right\|\leq M(1-\mu^*\beta)^{k-i}. \tag{48}$$

Proof: By *Lemma 5.12* and *Theorem 4.1*, there exist constants $h_0>0$ and $\xi>0$ which depend on sequence $\{\boldsymbol{A}_k\}$, such that

$$\sum_{j=kh_0+1}^{(k+1)h_0}\mathbb{E}[\boldsymbol{A}_j+\boldsymbol{A}_j^T]\geq\mu^*\xi I_{mn}\quad\forall k\geq 0$$

where $\boldsymbol{A}_j=\mathbf{\Lambda}\boldsymbol{F}_j+\mathscr{L}-\mathbf{\Lambda}\boldsymbol{F}_j\mathscr{L}$.

Hence, by *Lemma 5.5*, for the deterministic sequence $\{\mathbb{E}[\mathbf{\Lambda}\boldsymbol{F}_k+\mathscr{L}-\mathbf{\Lambda}\boldsymbol{F}_k\mathscr{L}],k\geq 0\}$, we have

$$\left\|\prod_{j=kh_0+1}^{(k+1)h_0}(I_{mn}-\mathbb{E}[\mathbf{\Lambda}\boldsymbol{F}_j+\mathscr{L}-\mathbf{\Lambda}\boldsymbol{F}_j\mathscr{L}])\right\|$$
$$\leq\left\{1-\frac{\mu^*\varepsilon\xi}{1+4(1-\varepsilon)h_0}\right\}^{1/2}.$$

It is easy to know that there exist constants $\beta\in(0,1)$ and $M>0$ such that

$$\left\|\prod_{j=i+1}^{k}(I_{mn}-\mathbb{E}[\mathbf{\Lambda}\boldsymbol{F}_j+\mathscr{L}-\mathbf{\Lambda}\boldsymbol{F}_j\mathscr{L}])\right\|\leq M(1-\mu^*\beta)^{k-i}.$$

This completes the proof. ■

Proof of Theorem 4.4: Let us define the following new sequence

$$\bar{\mathbf{\Theta}}_{k+1}^{\mathrm{CTA}}=(I_{mn}-\mathbb{E}[\mathbf{\Lambda}\boldsymbol{F}_k+\mathscr{L}-\mathbf{\Lambda}\boldsymbol{F}_k\mathscr{L}])\bar{\mathbf{\Theta}}_k^{\mathrm{CTA}}$$
$$+\mathbf{\Lambda}\boldsymbol{L}_k\boldsymbol{V}_k-\gamma\mathbf{\Omega}_{k+1} \tag{49}$$

with $\bar{\mathbf{\Theta}}_0^{\mathrm{CTA}}=\widetilde{\mathbf{\Theta}}_0^{\mathrm{CTA}}$. By (49), it is evident that

$$\widehat{\mathbf{\Pi}}_k^{\mathrm{CTA}}=\mathbb{E}[\bar{\mathbf{\Theta}}_k^{\mathrm{CTA}}(\bar{\mathbf{\Theta}}_k^{\mathrm{CTA}})^T]\quad k\geq 0.$$

Hence, by the Schwarz inequality

$$\|\mathbf{\Pi}_{k+1}^{\mathrm{CTA}}-\widehat{\mathbf{\Pi}}_{k+1}^{\mathrm{CTA}}\|$$
$$=\|\mathbb{E}[\widetilde{\mathbf{\Theta}}_{k+1}^{\mathrm{CTA}}(\widetilde{\mathbf{\Theta}}_{k+1}^{\mathrm{CTA}})^T-\bar{\mathbf{\Theta}}_{k+1}^{\mathrm{CTA}}(\bar{\mathbf{\Theta}}_{k+1}^{\mathrm{CTA}})^T]\|$$
$$\leq\|\widetilde{\mathbf{\Theta}}_{k+1}^{\mathrm{CTA}}-\bar{\mathbf{\Theta}}_{k+1}^{\mathrm{CTA}}\|_{L_2}(\|\widetilde{\mathbf{\Theta}}_{k+1}^{\mathrm{CTA}}\|_{L_2}+\|\bar{\mathbf{\Theta}}_{k+1}^{\mathrm{CTA}}\|_{L_2}). \tag{50}$$

Similar to the proof of *Theorem 4.3*, and using *Lemma 5.13*, it is easy to obtain

$$\|\bar{\mathbf{\Theta}}_{k+1}^{\mathrm{CTA}}\|_{L_2}=O\left(\left[\sqrt{\mu}+\frac{\gamma}{\sqrt{\mu}}\right]\log\frac{1}{\mu}+(1-\alpha\mu)^{k+1}\right) \tag{51}$$

where $\alpha\in(0,1)$ is a constant (without loss of generality, it may be taken as the same as that in *Theorem 4.3*).

The remaining proof is similar to [20, Th. 13], and is omitted here. Thus, we know that by *Conditions 4.1–4.4*, (21) holds. We remark that for random matrices $\{\mathbf{\Lambda}\boldsymbol{F}_k+\mathscr{L}-\mathbf{\Lambda}\mathscr{L}\boldsymbol{F}_k\}$ in the ATC diffusion NLMS algorithm, *Lemma 5.13* also holds. Therefore, *Theorems 4.4* and *4.5* hold for $\{\widetilde{\mathbf{\Theta}}_k^{\mathrm{ATC}},k\geq 1\}$ under the same conditions.

D. Proof of Theorem 4.6

Lemma 5.14: Let *Condition 4.1* be satisfied. If there exists a constant $\mu^*\in(0,1)$, such that for any $0<\mathbf{\Lambda}\leq\mu^* I_{mn}$, $\{\mathbf{\Lambda}\boldsymbol{F}_k+\mathscr{L}-\mathbf{\Lambda}\boldsymbol{F}_k\mathscr{L}\}\in S_1$, then there exists $h>0$ such that

$$\inf_k\lambda_{\min}\left\{\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\mathbb{E}\left[\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\right]\right\}>0. \tag{52}$$

The detailed proof of *Lemma 5.14* is given in Appendix B. We remark that the converse assertion of *Lemma 5.14* may not be true in general and this can be seen from [21, Example 2.1]. However, it will be true if we impose additional assumptions on $\{\boldsymbol{\varphi}_k^i\}$, for example, the ϕ-mixing condition.

Next, we prove *Theorem 4.6*.

Sufficiency: By *Theorem 4.1*, we know that $\{I_{mn}-[\mathbf{\Lambda}\boldsymbol{F}_k+\mathscr{L}-\mathbf{\Lambda}\boldsymbol{F}_k\mathscr{L}],k\geq 1\}$ is L_p-exponentially stable ($p\geq 1$).

Necessity: By *Lemma 5.14*, we can obtain (52). Since $\{\boldsymbol{\varphi}_k^i\}$ is ϕ-mixing, we know that $\{\sum_{i=1}^n\frac{\boldsymbol{\varphi}_k^i(\boldsymbol{\varphi}_k^i)^T}{1+\|\boldsymbol{\varphi}_k^i\|^2}\}$ is also ϕ-mixing. Then by (52), the ϕ-mixing property and [21, Th. 2.3], we know that *Condition 4.2* holds. This completes the proof.

VI. Simulation Results

In this section, we will construct a simulation example to illustrate that for regression vectors that are generated by linear stochastic state space models (where the regressors are strongly correlated and satisfy our cooperative information condition), even none of the sensors can estimate the parameters individually, the whole sensor network can still fulfill the filtering task cooperatively and effectively. Let us take $n=3$ with the following adjacency matrix:

$$\mathcal{A}=\begin{pmatrix}2/3 & 1/3 & 0\\ 1/3 & 1/2 & 1/6\\ 0 & 1/6 & 5/6\end{pmatrix}$$

then the corresponding graph is connected. We will estimate or track an unknown 3-D signal $\boldsymbol{\theta}_k$. Let us consider two cases: $\gamma=0$ ($\boldsymbol{\theta}_k$ is time-invariant) and $\gamma=0.1$ ($\boldsymbol{\theta}_k$ is time-varying)

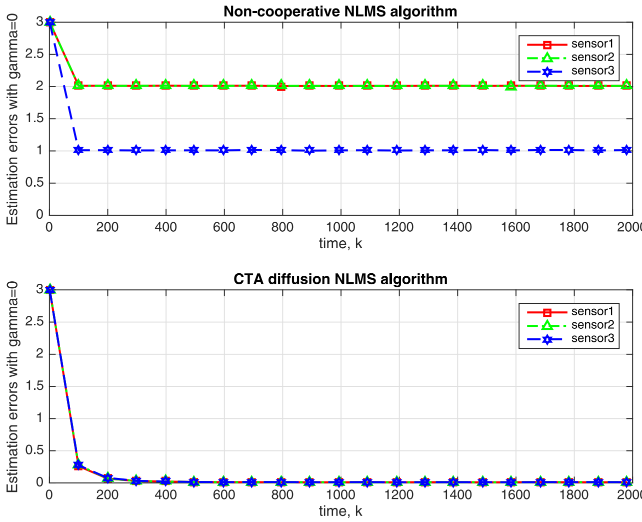

Fig. 1. Estimation errors of the three sensors with $\gamma = 0$.

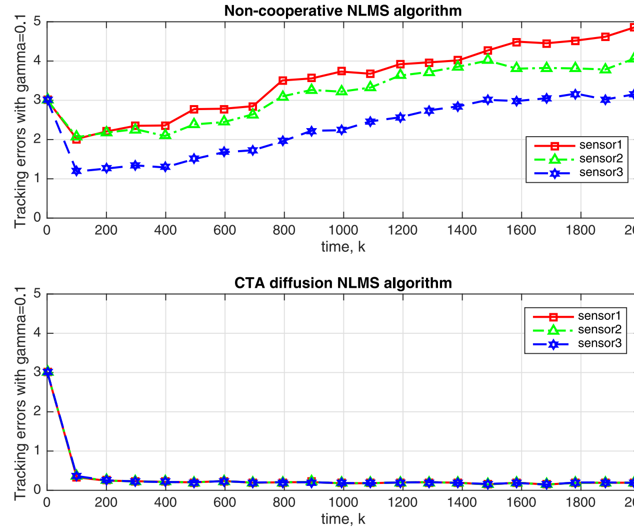

Fig. 2. Tracking errors of the three sensors with $\gamma = 0.1$.

with the parameter variation $\boldsymbol{\omega}_k \sim N(0, 0.3, 3, 1)$ (Gaussian distribution) in (2). In both cases, the observation noises $\{v_k^i, k \geq 1, i = 1, 2, 3\}$ are i.i.d. with $v_k^i \sim N(0, 0.3, 1, 1)$ in (1), where $\boldsymbol{\varphi}_k^i (i = 1, 2, 3)$ are generated by a state space model

$$\begin{cases} \boldsymbol{x}_k^i = A_i \boldsymbol{x}_{k-1}^i + B_i \xi_k^i \\ \boldsymbol{\varphi}_k^i = C_i \boldsymbol{x}_k^i \end{cases}$$

where $\{\xi_k^i, k \geq 1, i = 1, 2, 3\}$ are i.i.d. with $\xi_k^i \sim N(0, 0.3, 1, 1)$, and

$$A_1 = A_2 = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/5 \end{pmatrix}, A_3 = \begin{pmatrix} 4/5 & 0 & 0 \\ 4/5 & 0 & 0 \\ 4/5 & 0 & 0 \end{pmatrix}$$

$$B_1 = (1, 0, 0)^T, B_2 = (1, 0, 0)^T, B_3 = (1, 0, 0)^T$$

$$C_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, C_2 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, C_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It can be verified that *Condition 4.2* is satisfied with $h = 2$. Moreover, it is not difficult to verify that the necessary information condition in [21] is not satisfied for any individual sensor, since the three subsystems are not observable.

For numerical simulations, let $\boldsymbol{x}_0^1 = \boldsymbol{x}_0^2 = \boldsymbol{x}_0^3 = (1, 1, 1)^T$, $\boldsymbol{\theta}_0 = (1, 1, 1)^T, \hat{\boldsymbol{\theta}}_0^i = (0, 0, 0)^T (i = 1, 2, 3), \mu = 0.3$. Here, we repeat the simulation for $m = 500$ times with the same initial states. Then, for sensor $i (i = 1, 2, 3)$, we can get m sequences $\{\|\hat{\boldsymbol{\theta}}_k^{i,j} - \boldsymbol{\theta}_k^j\|^2, k = 1, 100, 200, \ldots, 2000\} (j = 1, \ldots, m)$, where the superscript j denotes the jth simulation result. We use $\frac{1}{m} \sum_{j=1}^m \|\hat{\boldsymbol{\theta}}_k^{i,j} - \boldsymbol{\theta}_k^j\|^2 (i = 1, 2, 3, k = 1, 100, 200, \ldots, 2000)$ to approximate the estimation or tracking errors with $\gamma = 0$ in Fig. 1 and with $\gamma = 0.1$ in Fig. 2.

When θ_k is time-invariant, the upper one in Fig. 1 is the noncooperative NLMS algorithm in which the estimation errors of the three sensors are all quite large because all the sensors do not satisfy the information condition in [21]. The lower one in Fig. 1 is the CTA diffusion NLMS algorithms in which all the estimation errors converge to a small neighborhood of zero as k increases, since the whole system satisfies *Condition 4.2*. In Fig. 2, θ_k is time-varying. The upper one is the individual situation in which the tracking errors of the three sensors keep large, and the lower one is the distributed situation in which all the mean square tracking errors converge nicely as k increases. Moreover, we can obtain some similar results for the ATC diffusion NLMS algorithm, which are omitted here.

VII. Concluding Remarks

This paper has established the stability and performance of a basic class of distributed adaptive filtering algorithms based on diffusion strategies, under a general stochastic cooperative information condition on the system regressor processes. This condition is not only sufficient for stability, but also necessary in a certain sense, which is a natural generation of the weakest known conditional information condition for single sensor case introduced by Guo [21], [32]. In fact, due to the difficulty in analyzing product of random matrices, almost all of the existing theory and analyses on distributed adaptive filtering algorithms require such stringent signal conditions as independence and stationarity, and thus exclude applications to feedback control systems. Moreover, our main results also demonstrate a desired but rarely rigorously established fact: the distributed adaptive filters can track a dynamic process of interest from noisy measurements by a set of sensors working cooperatively, in the natural scenario where none of the sensors can fulfill the estimation task individually. Of course, there are still a number of interesting problems for further investigation, for examples, how to establish similar theoretical results for other filtering algorithms, and how to combine distributed adaptive filters with distributed control problems, etc.

Appendix A
Proof of Lemma 5.10

For simplicity, we omit a.s. for sample paths in the following proof. By *Lemma 5.8*, we know that there exist constants $\mu^* \in (0, 1)$ and $\varepsilon = \delta/2$, such that for any $0 < \boldsymbol{\Lambda} < \mu^* I_{mn}$, (37) holds. Then by the proof of *Lemma 5.5*, we know that for any $\mu \in (0, \mu^*)$, $\gamma_k \in [0,1]$ holds.

By the notations in *Lemma 5.8*, we then have

$$\boldsymbol{A}_j + \boldsymbol{A}_j^T = 2(\boldsymbol{\Lambda F}_j + \mathscr{L}) - (\boldsymbol{\Lambda F}_j \mathscr{L} + \mathscr{L}\boldsymbol{F}_j\boldsymbol{\Lambda}).$$

Similar to the proof of (40), we have

$$\boldsymbol{\Lambda F}_j \mathscr{L} + \mathscr{L}\boldsymbol{F}_j\boldsymbol{\Lambda} \le \sqrt{2}\|\boldsymbol{\Lambda}^{\frac{1}{2}}\|(\boldsymbol{\Lambda F}_j + \mathscr{L}).$$

Then, we can obtain

$$\begin{aligned}\boldsymbol{A}_j + \boldsymbol{A}_j^T &\ge (2-\sqrt{2}\|\boldsymbol{\Lambda}^{\frac{1}{2}}\|)(\boldsymbol{\Lambda F}_j + \mathscr{L})\\ &\ge 0.5(\boldsymbol{\Lambda F}_j + \mathscr{L})\\ &\ge 0.5(\sigma_{\min}\mu^*\boldsymbol{F}_j + \mathscr{L})\\ &\ge 0.5\sigma_{\min}\mu^*(\boldsymbol{F}_j + \mathscr{L})\end{aligned}$$

where $\sigma_{\min} = \min\{\sigma_1, \ldots, \sigma_n\}$. Denote

$$\rho_k \triangleq \lambda_{\min}\left\{\mathbb{E}\left[\frac{0.5\sigma_{\min}\mu^*}{1+4(1-\varepsilon)h}\sum_{j=k+1}^{k+h}(\boldsymbol{F}_j + \mathscr{L})\Big|\mathcal{F}_k\right]\right\}. \quad (53)$$

Since $0 \le \rho_k \le \gamma_k \le 1$, by *Lemma 5.2* we know that to prove $\gamma_k \in S^0(\lambda^\nu)$, we need only to prove $\rho_k \in S^0(\lambda^\nu)$.

According to *Condition 4.1*, $\mathcal{L}$ has only one zero eigenvalue whose unit eigenvector is $\frac{1}{\sqrt{n}}(1,\ldots,1)^T$, i.e., $\frac{1}{\sqrt{n}}\mathbf{1}$ where $\mathbf{1} = (1,\ldots,1)^T_{n\times 1}$. Correspondingly, by *Lemma 5.1*, $\mathscr{L}$ has m zero eigenvalues whose orthogonal unit eigenvectors are

$$\boldsymbol{\xi}_1 = \frac{1}{\sqrt{n}}\mathbf{1}\otimes\boldsymbol{e}_1, \ldots, \boldsymbol{\xi}_m = \frac{1}{\sqrt{n}}\mathbf{1}\otimes\boldsymbol{e}_m$$

where $\boldsymbol{e}_i$ is a unit column vector with the ith element is 1 and the dimension is m. The other eigenvalues of $\mathscr{L}$ are $l_{m+1} \le \cdots \le l_{mn}$ arranged in a nondecreasing order whose orthogonal unit eigenvectors are denoted as $\boldsymbol{\xi}_{m+1}, \ldots, \boldsymbol{\xi}_{mn}$ correspondingly. Here, for an arbitrary unit vector $\boldsymbol{\eta} \in \mathbb{R}^{mn}$, it can be expressed as

$$\boldsymbol{\eta} = \sum_{j=1}^m x_j\boldsymbol{\xi}_j + \sum_{j=m+1}^{mn} x_j\boldsymbol{\xi}_j \triangleq \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2$$

where $\sum_{j=1}^m x_j^2 + \sum_{j=m+1}^{mn} x_j^2 = 1$. Now, let

$$\boldsymbol{H}_k^i \triangleq \sum_{j=k+1}^{k+h}\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}$$

$$\boldsymbol{H}_k \triangleq \text{diag}\{\boldsymbol{H}_k^1, \ldots, \boldsymbol{H}_k^n\}.$$

By the definition of $\boldsymbol{F}_j$ and denote

$$a \triangleq \frac{0.5\sigma_{\min}\mu^*}{1+4(1-\varepsilon)h}$$

we have $ah \le 1$ and

$$\begin{aligned}\boldsymbol{\Delta}_k &\triangleq \mathbb{E}\left[\frac{0.5\sigma_{\min}\mu^*}{1+4(1-\varepsilon)h}\sum_{j=k+1}^{k+h}(\boldsymbol{F}_j+\mathscr{L})\Big|\mathcal{F}_k\right]\\ &= \mathbb{E}[a\boldsymbol{H}_k + ah\mathscr{L}|\mathcal{F}_k].\end{aligned} \quad (54)$$

Note that

$$\begin{aligned}\boldsymbol{\Gamma}_k &\triangleq \mathbb{E}\left[\frac{1}{n(h+1)}\sum_{i=1}^n\sum_{j=k+1}^{k+h}\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\Big|\mathcal{F}_k\right]\\ &= \mathbb{E}\left[\frac{1}{n(h+1)}\sum_{i=1}^n\boldsymbol{H}_k^i\Big|\mathcal{F}_k\right].\end{aligned} \quad (55)$$

Let us now consider the following quadratic form of (54):

$$\begin{aligned}\boldsymbol{\eta}^T\boldsymbol{\Delta}_k\boldsymbol{\eta} &= (\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)^T\boldsymbol{\Delta}_k(\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)\\ &= a\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_k|\mathcal{F}_k]\boldsymbol{\eta}_1 + a\boldsymbol{\eta}_2^T\mathbb{E}[\boldsymbol{H}_k|\mathcal{F}_k]\boldsymbol{\eta}_2\\ &\quad + 2a\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_k|\mathcal{F}_k]\boldsymbol{\eta}_2\\ &\quad + ah\boldsymbol{\eta}_1^T\mathscr{L}\boldsymbol{\eta}_1 + ah\boldsymbol{\eta}_2^T\mathscr{L}\boldsymbol{\eta}_2\\ &\quad + 2ah\boldsymbol{\eta}_1^T\mathscr{L}\boldsymbol{\eta}_2\\ &\triangleq s_1+s_2+s_3+s_4+s_5+s_6.\end{aligned} \quad (56)$$

By [14, Proof of Th. 1], we have

$$\boldsymbol{\eta}^T\boldsymbol{\Delta}_k\boldsymbol{\eta} \ge (1-\delta)s_1 + \left(1-\frac{1}{\delta}\right)s_2 + s_4 + s_5 + s_6 \quad (57)$$

where $\delta > 0$ can be any constant. Now, we proceed to estimate s_i one by one.

Note that

$$\begin{aligned}s_1 &= a\mathbb{E}[\boldsymbol{\eta}_1^T\boldsymbol{H}_k\boldsymbol{\eta}_1|\mathcal{F}_k]\\ &= a\mathbb{E}\left[\left(\sum_{j=1}^m x_j\boldsymbol{\xi}_j\right)^T\boldsymbol{H}_k\left(\sum_{j=1}^m x_j\boldsymbol{\xi}_j\right)\Big|\mathcal{F}_k\right]\\ &= a\mathbb{E}[\boldsymbol{X}^T\boldsymbol{\Xi}^T\boldsymbol{H}_k\boldsymbol{\Xi X}|\mathcal{F}_k]\end{aligned} \quad (58)$$

where $\boldsymbol{X} = [x_1, \ldots, x_m]^T, \boldsymbol{\Xi} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m]$. By [14, Proof of Th. 1], we have

$$\mathbb{E}[\boldsymbol{\Xi}^T\boldsymbol{H}_k\boldsymbol{\Xi}|\mathcal{F}_k] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n\boldsymbol{H}_k^i|\mathcal{F}_k\right] = (h+1)\boldsymbol{\Gamma}_k. \quad (59)$$

Substitute (59) into (58), it can be deduced that

$$s_1 \ge a(h+1)\boldsymbol{X}^T\boldsymbol{\Gamma}_k\boldsymbol{X} \ge a(h+1)\lambda_k\sum_{j=1}^m x_j^2 \ge ah\lambda_k\sum_{j=1}^m x_j^2. \quad (60)$$

Notice that

$$|s_2| \le ah\|\boldsymbol{\eta}_2\|^2 = ah\sum_{j=m+1}^{mn}x_j^2. \quad (61)$$

Since $\boldsymbol{\eta}_1 = \sum_{j=1}^m x_j\boldsymbol{\xi}_j$ and $\boldsymbol{\xi}_j (1 \le j \le m)$ is the eigenvector corresponding to the zero eigenvalue, we have

$$s_4 = s_6 = 0. \quad (62)$$

For s_5, we know that

$$s_5 = ah\sum_{j=m+1}^{mn}l_jx_j^2 \ge ahl_{m+1}\sum_{j=m+1}^{mn}x_j^2. \quad (63)$$

Denote $y \triangleq \sum_{j=1}^{m} x_j^2$. Since $\rho_k = \lambda_{\min}(\mathbf{\Delta}_k)$ and here we choose $\delta \in (0,1)$, we have by (56)

$$\begin{aligned}\rho_k &\geq (1-\delta)ah\lambda_k y + \left(1-\frac{1}{\delta}\right)ah(1-y) + ahl_{m+1}(1-y)\\ &= \left[(1-\delta)ah\lambda_k - \left(l_{m+1}+1-\frac{1}{\delta}\right)ah\right]y\\ &\quad + \left(l_{m+1}+1-\frac{1}{\delta}\right)ah, \quad y\in[0,1]. \qquad (64)\end{aligned}$$

Here we choose $\delta = 1/(1+0.5l_{m+1})$ and since $\lambda_k \in [0,1]$, then we have for $y \in [0,1]$

$$\rho_k \geq \left[\frac{0.5l_{m+1}ah}{1+0.5l_{m+1}}\lambda_k - 0.5l_{m+1}ah\right]y + 0.5l_{m+1}ah. \qquad (65)$$

It is easy to obtain

$$\rho_k \geq \frac{0.5l_{m+1}ah}{1+0.5l_{m+1}}\lambda_k = \frac{l_{m+1}ah}{2+l_{m+1}}\lambda_k = c\lambda_k$$

where $0 < c = \frac{l_{m+1}ah}{2+l_{m+1}} < 1$.

By *Lemma 5.3* and since $\lambda_k \in [0,\alpha^*]$, $\alpha^* = \frac{h}{h+1}$ and $\{\lambda_k\} \in S_0(\lambda)$, we have $\{\rho_k\} \in S^0(\rho)$, where $\rho = \lambda^\nu$ and

$$\nu = (1-\alpha^*)c = \frac{0.5hl_{m+1}\sigma_{\min}\mu^*}{(2+l_{m+1})(1+h)[1+4(1-\varepsilon)h]} > 0$$

where $\varepsilon = \delta/2$. This completes the proof.

Appendix B
Proof of Lemma 5.14

For simplicity, we omit a.s. for sample paths in the following proof. Let $\mathbf{A}_k = \mathbf{\Lambda F}_k + \mathscr{L} - \mathbf{\Lambda F}_k\mathscr{L}$ and denote

$$\gamma_k = \lambda_{\min}\left\{\mathbb{E}\left[\sum_{j=k+1}^{k+h}(\mathbf{A}_j + \mathbf{A}_j^T)\right]\right\}.$$

Then since $0 < \mathbf{\Lambda} \leq \mu^* I_{mn}$ and by *Lemma 5.12*, we know that

$$\gamma_k \geq \left(\frac{\sqrt{a_2^2+2a_1}-a_2}{2a_1}\right)^2 > 0 \ \forall k. \qquad (66)$$

Similar to the proof of (40), for any mn-dimensional unit column vector x, we can obtain

$$\begin{aligned}&x^T(\mathbf{A}_j+\mathbf{A}_j^T)x\\ &= x^T[2(\mathbf{\Lambda F}_j+\mathscr{L}) - (\mathbf{\Lambda F}_j\mathscr{L}+\mathscr{L}\mathbf{F}_j\mathbf{\Lambda})]x\\ &= 2x^T(\mathbf{\Lambda F}_j+\mathscr{L})x - 2x^T\mathbf{\Lambda F}_j\mathscr{L}x\\ &\leq 2x^T(\mathbf{\Lambda F}_j+\mathscr{L})x + 2\|x^T\mathbf{\Lambda}^{\frac12}\mathbf{F}_j\mathbf{\Lambda}^{\frac12}\|\cdot\|\mathscr{L}x\|\\ &\leq 2x^T(\mathbf{\Lambda F}_j+\mathscr{L})x + \sqrt{2}\|\mathbf{\Lambda}^{\frac12}\|\cdot x^T(\mathbf{\Lambda F}_j+\mathscr{L})x\\ &\leq (2+\sqrt{2}\|\mathbf{\Lambda}^{\frac12}\|)\cdot x^T(\mathbf{\Lambda F}_j+\mathscr{L})x\\ &\leq (2+\sqrt{2})x^T(\mathbf{F}_j+\mathscr{L})x\end{aligned}$$

then we have

$$\mathbf{A}_j + \mathbf{A}_j^T \leq (2+\sqrt{2})(\mathbf{F}_j+\mathscr{L}).$$

Denote

$$\rho_k = \lambda_{\min}\left\{\mathbb{E}\left[\sum_{j=k+1}^{k+h}(\mathbf{F}_j+\mathscr{L})\right]\right\}.$$

Then by the definition of γ_k, we have

$$\rho_k \geq \frac{\gamma_k}{2+\sqrt{2}} > 0 \qquad (67)$$

which means that $\inf_k \rho_k > 0$ by (66). We next show that there exists a positive constant δ such that

$$\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\right] \geq \delta I_m$$

for all $k \geq 0$. We first prove that for any $k \geq 0$, the eigenvalues of matrices $\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\right]$ are all positive. This can be done through contradiction by assuming that there exists a time instant k^* such that the smallest eigenvalue of matrix $\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k^*+1}^{k^*+h}\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\right]$ is 0. Denote the corresponding unit eigenvector as β_{k^*}, then we have

$$\boldsymbol{\beta}_{k^*}^T\left(\sum_{i=1}^{n}\sum_{j=k^*+1}^{k^*+h}\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\right]\right)\boldsymbol{\beta}_{k^*} = 0. \qquad (68)$$

Similar to the proof of *Lemma 5.9*, we know that $\mathcal{L}$ has only one zero eigenvalue whose unit eigenvector is $\frac{1}{\sqrt{n}}\mathbf{1}$ where $\mathbf{1} = (1,\ldots,1)_{n\times1}^T$. Correspondingly, $\mathcal{L}\otimes I_m$ has m zero eigenvalues whose orthogonal unit eigenvectors are

$$\boldsymbol{\xi}_1 = \frac{1}{\sqrt{n}}\mathbf{1}\otimes\boldsymbol{e}_1,\ldots,\boldsymbol{\xi}_m = \frac{1}{\sqrt{n}}\mathbf{1}\otimes\boldsymbol{e}_m$$

where $\boldsymbol{e}_i$ is a unit column vector with the ith element is 1 and the dimension is m. The other eigenvalues of $\mathcal{L}\otimes I_m$ are $l_{m+1},\ldots,l_{mn}$ whose orthogonal unit eigenvectors are denoted as $\boldsymbol{\xi}_{m+1},\ldots,\boldsymbol{\xi}_{mn}$, respectively. Note that for an arbitrary unit vector $\boldsymbol{\eta}\in\mathbb{R}^{mn}$, it can be expressed as

$$\boldsymbol{\eta} = \sum_{j=1}^{m}x_j\boldsymbol{\xi}_j + \sum_{j=m+1}^{mn}x_j\boldsymbol{\xi}_j \triangleq \boldsymbol{\eta}_1+\boldsymbol{\eta}_2$$

where $\sum_{j=1}^{m}x_j^2 + \sum_{j=m+1}^{mn}x_j^2 = 1$. Now, let

$$\begin{aligned}\mathbf{H}_{k^*}^i &= \sum_{j=k^*+1}^{k^*+h}\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\\ \mathbf{H}_{k^*} &= \mathrm{diag}\{\mathbf{H}_{k^*}^1,\ldots,\mathbf{H}_{k^*}^n\}.\end{aligned}$$

By the definition of $\mathbf{F}_j$, we have

$$\begin{aligned}\mathbf{\Delta}_{k^*} &\triangleq \mathbb{E}\left[\sum_{j=k^*+1}^{k^*+h}(\mathbf{F}_j+\mathscr{L})\right]\\ &= \mathbb{E}[\mathbf{H}_{k^*}+h\mathscr{L}]. \qquad (69)\end{aligned}$$

Note that

$$\Gamma_{k^*} \triangleq \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k^*+1}^{k^*+h}\frac{\varphi_j^i(\varphi_j^i)^T}{1+\|\varphi_j^i\|^2}\right] = \mathbb{E}\left[\sum_{i=1}^{n}\boldsymbol{H}_{k^*}^i\right]. \tag{70}$$

Let us consider the following quadratic form of (69):

$$\begin{aligned}&\boldsymbol{\eta}^T\boldsymbol{\Delta}_{k^*}\boldsymbol{\eta}\\ &=(\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)^T\boldsymbol{\Delta}_{k^*}(\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)\\ &=\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_1+\boldsymbol{\eta}_2^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_2+2\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_2\\ &\quad+h\boldsymbol{\eta}_1^T\mathscr{L}\boldsymbol{\eta}_1+h\boldsymbol{\eta}_2^T\mathscr{L}\boldsymbol{\eta}_2+2h\boldsymbol{\eta}_1^T\mathscr{L}\boldsymbol{\eta}_2\\ &\triangleq s_1^{k^*}+s_2^{k^*}+s_3^{k^*}+s_4+s_5+s_6.\end{aligned} \tag{71}$$

For matrices $\boldsymbol{\zeta}_1$ and $\boldsymbol{\zeta}_2$, we have following inequality:

$$2\boldsymbol{\zeta}_1^T\boldsymbol{\zeta}_2 \le \delta\boldsymbol{\zeta}_1^T\boldsymbol{\zeta}_1+\frac{1}{\delta}\boldsymbol{\zeta}_2^T\boldsymbol{\zeta}_2 \tag{72}$$

where $\delta > 0$ can be any constant. Let

$$\boldsymbol{\zeta}_1 \triangleq (\mathbb{E}[\boldsymbol{H}_{k^*}])^{1/2}\boldsymbol{\eta}_1,\ \boldsymbol{\zeta}_2 \triangleq (\mathbb{E}[\boldsymbol{H}_{k^*}])^{1/2}\boldsymbol{\eta}_2$$

and substitute this into (72), it is easy to have

$$\begin{aligned}2\boldsymbol{\zeta}_1^T\boldsymbol{\zeta}_2 &= 2\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_2 \le \delta\boldsymbol{\zeta}_1^T\boldsymbol{\zeta}_1+\frac{1}{\delta}\boldsymbol{\zeta}_2^T\boldsymbol{\zeta}_2\\ &=\delta\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_1+\frac{1}{\delta}\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_1.\end{aligned}$$

Then, we can obtain

$$\begin{aligned}s_3^{k^*} &= 2\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_2 \le \delta\boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_1+\frac{1}{\delta}\boldsymbol{\eta}_2^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_2\\ &=\delta s_1^{k^*}+\frac{1}{\delta}s_2^{k^*}.\end{aligned} \tag{73}$$

From (71) and (73), it is obvious that

$$\boldsymbol{\eta}^T\boldsymbol{\Delta}_{k^*}\boldsymbol{\eta} \le (1+\delta)s_1^{k^*}+\left(1+\frac{1}{\delta}\right)s_2^{k^*}+s_4+s_5+s_6. \tag{74}$$

Now, we will estimate $s_1^{k^*}, s_2^{k^*}, s_4, s_5$ and s_6. By *Lemma 5.4*, we know that

$$\begin{aligned}s_1^{k^*} &= \boldsymbol{\eta}_1^T\mathbb{E}[\boldsymbol{H}_{k^*}]\boldsymbol{\eta}_1\\ &=\mathbb{E}\left[\left(\sum_{j=1}^{m}x_j\boldsymbol{\xi}_j\right)^T\boldsymbol{H}_{k^*}\left(\sum_{j=1}^{m}x_j\boldsymbol{\xi}_j\right)\right]\\ &=\mathbb{E}[\boldsymbol{X}^T\boldsymbol{\Xi}^T\boldsymbol{H}_{k^*}\boldsymbol{\Xi}\boldsymbol{X}]\\ &=\mathbb{E}\left[\frac{1}{n}\boldsymbol{X}^T\left(\sum_{i=1}^{n}\boldsymbol{H}_{k^*}^i\right)\boldsymbol{X}\right]\\ &=\frac{1}{n}\cdot\boldsymbol{X}^T\boldsymbol{\Gamma}_{k^*}\boldsymbol{X}\end{aligned} \tag{75}$$

where $\boldsymbol{X}=[x_1,\ldots,x_m]^T\in\mathbb{R}^m$, $\boldsymbol{\Xi}=[\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_m]$.

Notice that

$$|s_2^{k^*}| \le h\|\boldsymbol{\eta}_2\|^2 = h\sum_{j=m+1}^{mn}x_j^2. \tag{76}$$

Since $\boldsymbol{\eta}_1=\sum_{j=1}^{m}x_j\boldsymbol{\xi}_j$ and $\boldsymbol{\xi}_j(1\le j\le m)$ is the eigenvector corresponding to the zero eigenvalue, we have

$$s_4 = s_6 = 0. \tag{77}$$

For s_5, we know that

$$s_5 = h\sum_{j=m+1}^{mn}l_jx_j^2 \le hl_{mn}\sum_{j=m+1}^{mn}x_j^2. \tag{78}$$

Denote

$$y \triangleq \sum_{j=1}^{m}x_j^2 \in [0,1].$$

By (74)–(78) and since $\rho_{k^*}=\lambda_{\min}(\Delta_{k^*})$, we know that for any $y\in[0,1]$,

$$\begin{aligned}\rho_{k^*} &\le \frac{1+\delta}{n}\cdot\boldsymbol{X}^T\boldsymbol{\Gamma}_{k^*}\boldsymbol{X}+\left(1+\frac{1}{\delta}\right)h(1-y)\\ &\quad+hl_{mn}(1-y).\end{aligned} \tag{79}$$

We can take $\boldsymbol{X}=\boldsymbol{\beta}_{k^*}$, then we have $y=1$ and

$$\rho_{k^*} \le \frac{1+\delta}{n}\cdot\boldsymbol{X}^T\boldsymbol{\Gamma}_{k^*}\boldsymbol{X}=0$$

which contradicts with $\rho_k > 0\ \forall k\ge 0$.

In a similar way, we can prove that all of the eigenvalues of the above matrix must have a uniform lower bound $\delta > 0$ with respect to $k\ge 0$. This is done through contradiction by assuming that there exist unit eigenvectors $\boldsymbol{\beta}_k$ and a sequence $\{k_s\}_{s=1}^{\infty}$ such that

$$\begin{aligned}&\lim_{s\to\infty}\boldsymbol{\beta}_{k_s}^T\left\{\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k_s+1}^{k^*+h}\frac{\varphi_j^i(\varphi_j^i)^T}{1+\parallel\varphi_j^i\parallel^2}\right]\right\}\boldsymbol{\beta}_{k_s}\\ &=\lim_{s\to\infty}\boldsymbol{\beta}_{k_s}^T\sigma_{k_s}\boldsymbol{\beta}_{k_s}=0\end{aligned} \tag{80}$$

where σ_{k_s} is the eigenvalue corresponding to the eigenvector $\boldsymbol{\beta}_{k_s}$. Similar to the above proof, we can take $\boldsymbol{X}=\boldsymbol{\beta}_{k_s}$, then it is obvious that

$$\lim_{s\to\infty}\rho_{k_s} \le \frac{1+\delta}{n}\cdot\lim_{s\to\infty}\boldsymbol{\beta}_{k_s}^T\boldsymbol{\Gamma}_{k_s}\boldsymbol{\beta}_{k_s}=0$$

which contradicts with $\inf_k\rho_k>0$. Therefore, we conclude that there exists a positive constant δ such that

$$\mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\frac{\varphi_j^i(\varphi_j^i)^T}{1+\parallel\varphi_j^i\parallel^2}\right]\ge\delta I_m$$

for all $k\ge 0$. This completes the proof.

REFERENCES

[1] S. Kar, J. M. F. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3575–3605, Jun. 2012.

[2] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2365–2382, Jun. 2009.

[3] A. Rastegarnia, M. A. Tinati, and A. Khalili, "A distributed incremental LMS algorithm with reliability of observation consideration," in *Proc. IEEE Int. Conf. Commun. Syst.*, Singapore, Nov. 2010, pp. 67–70.

[4] R. Carli *et al.*, "Distributed Kalman filtering based on consensus strategies," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 4, pp. 622–633, May 2008.

[5] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 674–690, Aug. 2011.

[6] 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. Autom. Control*, vol. 59, no. 1, pp. 91–106, Jan. 2014.

[7] N. Takahashi, I. Yamada, and A. H. Sayed, "Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4795–4810, Sep. 2010.

[8] R. Abdolee and B. Champagne, "Diffusion LMS strategies in sensor networks with noisy input data," *IEEE/ACM Trans. Netw.*, vol. 24, no. 1, pp. 3–14, Feb. 2016.

[9] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chanbers "Steady-state analysis of diffusion LMS adaptive networks with noisy links," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 974–979, Feb. 2012.

[10] M. J. Piggott and V. Solo, "Stability of distributed adaptive algorithms II: Diffusion algorithms," in *Proc. 54th IEEE Conf. Decis. Control*, 2015, pp. 7428–7433.

[11] H. Nosrati, M. Shamsi, S. M. Taheri, and M. H. Sedaaghi, "Adaptive networks under non-stationary conditions: Formulation, performance analysis, and application," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4300–4314, Aug. 2015.

[12] J. Chen and A. H. Sayed, "On the learning behavior of adaptive networks Part I: Transient analysis," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3487–3517, Jun. 2015.

[13] O. N. Gharehshiran, V. Krishnamurthy, and G. Yin, "Distributed energy-aware diffusion least mean squares: Game-theoretic learning," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 5, pp. 821–836, Oct. 2013.

[14] S. Y. Xie and L. Guo, "Stability of distributed LMS under cooperative stochastic excitation," in *Proc. 34th Chin. Control Conf.*, Hangzhou, China, pp. 7445–7450, Jul. 2015.

[15] S. Y. Tu and A. H. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6217–6234, Dec. 2012.

[16] H. J. Kushner, *Approximatation and Weak Convergence Methods for Random Processes With Applications to Stochastic Systems Theory*. Cambridge, MA, USA: MIT Press, 1984.

[17] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Applications*. New York, NY, USA: Springer-Verlag, 1997.

[18] C. Chen, Z. X. Liu, and L. Guo, "Stability of diffusion adaptive filters," in *Proc. IFAC World Congress*, Cape Town, South Africa, Aug. 2014, pp. 10409–10414.

[19] C. Chen, Z. X. Liu, and L. Guo, "Performance bounds of distributed adaptive filters with cooperative correlated signals," *Sci. China Inf. Sci.*, vol. 59, no. 11, 2016, Art. no. 112202.

[20] C. Chen, Z. X. Liu, and L. Guo, "Performance analysis of distributed adaptive filters," *Commun. Inf. Syst.*, vol. 15, no. 4, pp. 453–476, 2015.

[21] L. Guo, "Stability of recursive stochastic tracking algorithms," *SIAM J. Control Optim.*, vol. 32, no. 5, pp. 1195–1225, Sep. 1994.

[22] L. Guo and L. Ljung, "Exponential stability of general tracking algorithms," *IEEE Trans. Autom. Control*, vol. 40, no. 8, pp. 1376–1387, Aug. 1995.

[23] S. Y. Xie and L. Guo, "Exponential stability of LMS-based distributed adaptive filters," in *Proc. 20th World Congr. Int. Fed. Autom. Control*, Toulouse, France, Jul. 9–14, 2017, pp. 13764–13769.

[24] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification*. Cambridge, MA, USA: MIT Press, 1983.

[25] A. H. Sayed, *Adaptation, Learning, and Optimization Over Networks*, (Foundations and Trends in Machine Learning). Breda, Netherlands: NOW Publishers, vol. 7, no. 4/5, pp. 311–801, Jul. 2014.

[26] B. Widrow and S. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1985.

[27] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1995.

[28] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[29] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach With Applications in Transmission*. New York, NY, USA: Wiley, 1995.

[30] Y. S. Chow and H. Teicher, *Probability Theory*. New York, NY, USA: Springer, Mar. 1978.

[31] C. Godsil and G. Royle, *Algebraic Graph Theory*. New York, NY, USA: Springer, 2001.

[32] L. Guo, "Estimating time-varying parameters by Kalman filter based algorithm: Stability and convergence," *IEEE Trans. Autom. Control*, vol. 35, no. 2, pp. 141–147, Feb. 1990.

[33] L. Guo and L. Ljung, "Performance analysis of general tracking algorithms," *IEEE Trans. Autom. Control*, vol. 40, no. 8, pp. 1388–1402, Aug. 1995.

[34] L. Guo, L. Ljung, and G. J. Wang, "Necessary and sufficient conditions for stability of LMS," *IEEE Trans. Autom. Control*, vol. 42, no. 6, pp. 761–770, Jun. 1997.

[35] A. J. Laub, *Matrix Analysis for Scientists and Engineers*. Philadelphia, PA, USA: SIAM, 2005.

Siyu Xie received the B.S. degree in information and computing science (systems control) from Beijing University of Aeronautics and Astronautics, Beijing, China. She is currently working toward the Ph.D. degree majoring in control theory at Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China.

Her research interests include networked systems, distributed adaptive filters, machine learning, compressive sensing, and distributed control.

Dr. Xie received the IEEE CSS Beijing Chapter Young Author Prize at Chinese Control Conference in 2015, and China National Scholarship for graduate students in 2017.

Lei Guo (M'88–SM'96–F'99) was born in 1961. He received the B.S. degree in mathematics from Shandong University, Jinan, China, in 1982, and the Ph.D. degree in control theory from the Chinese Academy of Sciences, Beijing, China, in 1987.

He was a Postdoctoral Fellow at the Australian National University (1987–1989). Since 1992, he has been a Professor with the Institute of Systems Science, CAS. He has been the President of the Academy of Mathematics and Systems Science, CAS. He is currently the Director of the National Center for Mathematics and Interdisciplinary Sciences, CAS. He has worked on problems in stochastic adaptive control, system identification, adaptive signal processing, nonstationary time series analysis, and the maximum capability of feedback. His current research interests include distributed adaptive filtering, nonlinear uncertain systems control, multiagent systems, game-based control systems, and PID control theory, among others.

Prof. Guo became a member of Chinese Academy of Sciences in 2001, a Fellow of Academy of Sciences for the Developing World (TWAS) in 2002, a Foreign Member of Royal Swedish Academy of Engineering Sciences in 2007, a Fellow of International Federation of Automatic Control (IFAC) in 2007 "for fundamental contributions to the theory of adaptive control and estimation of stochastic systems, and to the understanding of the maximum capability of feedback." He was the recipient of the 12th IFAC World Congress Young Author Prize, and the coauthor of a SIGEST paper published by SIAM Review in 2014. He was also awarded an honorary doctorate by Royal Institute of Technology (KTH), Sweden, in 2014, Outstanding Service Award of IFAC in 2011, National Natural Science Prizes of China in 1987 and 1997. He was twice invited to deliver plenary lectures at the IFAC World Congress in 1999 and 2014, respectively, was an invited speaker at the International Congress of Mathematicians in 2002, and was selected as a Distinguished Lecturer of the IEEE Control Systems Society in 2012. He has served as a Council Member of IFAC, a member of IEEE Control Systems Award Committee, a member of the IFAC Award Committee, the General Co-Chair of 48th IEEE-CDC, the Congress Director of the 8th International Congress on Industrial and Applied Mathematics, the President of China Society for Industrial and Applied Mathematics, the Vice-President of Chinese Mathematical Society, and the Vice-President of Chinese Association of Automation, etc.