SIAM J. CONTROL OPTIM.
Vol. 56, No. 5, pp. 3404–3431

ANALYSIS OF NORMALIZED LEAST MEAN SQUARES-BASED CONSENSUS ADAPTIVE FILTERS UNDER A GENERAL INFORMATION CONDITION*

SIYU XIE† AND LEI GUO†

Abstract. A distributed adaptive filter can estimate an unknown signal of interest by a set of sensors working cooperatively, when any individual sensor cannot fulfill the filtering task due to lack of necessary information condition. This paper considers these kinds of filtering problems and focuses on a class of consensus normalized least mean squares-based algorithms. A general and weakest possible cooperative information condition is introduced to guarantee the stability of these kinds of adaptive filters, without resorting to commonly used but stringent conditions such as independence and stationarity of the system signals, which makes our theory applicable to feedback systems. Moreover, this general information condition is shown to be not only sufficient but also necessary for stability of the adaptive filters for a large class of random signals with decaying dependence. We further show that the mean square tracking error matrix can be approximately calculated by a linear deterministic difference matrix equation that can be easily evaluated and analyzed.

Key words. distributed adaptive filter, consensus algorithm, least mean squares, information condition, stochastic stability, tracking performance

AMS subject classifications. 68W15, 62M20, 93E15

DOI. 10.1137/16M1106791

1. Introduction.

With the development of network technology, distributed filtering (or estimation) algorithms have been widely studied in many practical situations. Examples include collaborative spectral sensing in cognitive radio systems, target localization in biological networks, fish schooling, bee swarming, and bird flight in mobile adaptive networks; see, e.g., [34, 35]. It is well known that distributed filtering algorithms have some significant advantages over the centralized ones or the fusion methods (cf., e.g., [1, 47]). In fact, the centralized algorithms usually require the distributed sensor network to transmit the whole system information to a fusion center to estimate the unknown signals, which may lack robustness at the fusion center and need strong communication capability over the sensor networks. In many practical situations, sensors may only have the capability to exchange information locally with their neighbors, which is the main motivation for the development of distributed algorithms. Moreover, Sayed [33] showed that the performance of the distributed estimation algorithms could be improved to even reach the performance of the traditional unweighted centralized estimation algorithm by optimizing the combination weights of the network topology.

It goes without saying that different distributed strategies may give different distributed filtering algorithms, and three types of strategies have been investigated in the literature, i.e., the incremental [6, 32, 39], consensus [3, 10, 20, 26, 27, 29, 36, 37],

*Received by the editors December 6, 2016; accepted for publication (in revised form) August 3, 2018; published electronically September 25, 2018.
http://www.siam.org/journals/sicon/56-5/M110679.html
Funding: This work was supported by the National Natural Science Foundation of China under grants 11688101 and 61227902.
†Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, and School of Mathematical Science, University of Chinese Academy of Sciences, Beijing 100049, China (xiesiyu7@gmail.com, Lguo@amss.ac.cn).

and diffusion [4, 5, 24, 28, 30, 31] strategies. Incremental techniques require to determine a cyclic path that runs across all nodes, which is generally an NP-hard problem. Moreover, the consensus strategies were introduced early for distributed estimation problems in [3, 20, 27, 29, 36], and the diffusion ones were investigated in [4, 5, 24]. Note that these two strategies are fully distributed and attract more research attentions, since they are scalable and robust. For examples, Solo [37] gave a stability analysis of the consensus least mean squares (LMS) algorithm, Mateos, Schizas, and Giannakis [27] studied the consensus-based recursive least squares (RLS) algorithm, and Lopes and Sayed [24], Cattivelli and Sayed [4], Piggott and Solo [30, 31] introduced diffusion LMS schemes for distributed estimation and studied the stability and performance of diffusion algorithms. Moreover, Carli et al. [3], Olfati-Saber [29], and Cattivelli and Sayed [5] investigated distributed Kalman filtering algorithms. Chen et al. [10] investigated a class of consensus-type cooperative adaptive identification algorithms in continuous-time under a deterministic cooperative information condition. In addition, Nosrati et al. [28] studied the tracking behavior of a wide range of adaptive networks and analyzed the mean square error performance of centralized, incremental, and diffusion algorithms. Furthermore, Kar, Moura, and Poor [19] and Kar, Moura, and Ramanan [20] considered the problem of distributed parameter estimation and established a globally observable condition under which the distributed estimates are consistent and asymptotically normal. However, a diminishing adaptation gain was used in [19, 20], which may prevent the algorithms from tracking unknown time-varying dynamic processes. Thus, we will use a fixed gain in this work, which is widely used in the literature and enables the distributed algorithms to have the ability for tracking time-varying parameters.

Although the distributed LMS algorithms can be applied to a wide class of signals and requires no assumptions on the statistics of regression vectors and measurements, when it comes to stability and performance evaluation, some statistical conditions are necessarily required to carry out the theoretical analysis, as pointed out by Schizas, Mateos, and Giannakis in [36]. In fact, due to the mathematical difficulty in analyzing the product of random matrices (PRM), almost all the existing related literature require certain spatial and/or temporal independence conditions on the system signals in the stability and performance analysis.

That is not surprising, because even for the traditional single sensor case, the stability analysis of estimation algorithms also requires independence and stationarity in most of the literature (e.g., Widrow and Stearns [40] and Haykin [18]), with only a few exceptions. For example, Solo and Kong [38] and Macchi [25] relaxed the independence in a certain sense, Kushner and Yin [21] and Yin and Krishnamurthy [46] investigated the weak convergence property (when the adaptation gain tends to zero) for a large class of weakly dependent signals. In particular, in a series of works conducted by Guo and his coauthors ([14, 15, 16, 17]), a general class of tracking algorithms (including LMS, RLS, and Kalman filter) were investigated by introducing a general information (or excitation) condition, which require neither independence nor stationarity on the system signals, and hence do not exclude the applications of the theory to feedback systems.

As pointed out by Solo [37], the first step towards relaxing the spatial and temporal independence and stationarity conditions for the analysis of distributed adaptive filtering algorithms is made by Chen, Liu, and Guo [9], where a combine-then-adapt diffusion normalized LMS (NLMS) algorithm is considered. It has been shown that the whole sensor networks can fulfill the estimation task under a cooperative stochastic information condition [7, 8, 9]. However when the sensor network degenerates to a single sensor, the information condition in [7, 8, 9] cannot include the weakest-known information condition introduced and investigated in [14, 15, 16, 17], indicating that there is still much room for improvement. Other interesting work that does not require the temporal independence of the regressors can be found in, e.g., [30, 31, 37]. In particular, it appears that [30] is the first one to study the almost sure convergence of the standard LMS-based distributed algorithms. Recently, the exponential stability of the homogenous part of the consensus-type NLMS algorithm is established in [44] under a general cooperative stochastic information condition, which is considerably weaker than that previously used in [7, 8, 9]. However, neither performance results for the tracking error nor necessity results for the information condition used are given in [44]. This is the main motivation for the current paper.

In this paper, we consider the consensus NLMS-based filtering algorithm for estimating time-varying unknown signals, aiming at establishing a theory for the tracking performance. We will first show that both the stability and the tracking performance bounds can be established under a general cooperative information condition, which needs neither independence nor stationarity of the system signals. We will then show that the actual mean square tracking error matrix can be adequately approximated by a simple linear and deterministic difference matrix equation which can be easily evaluated, analyzed, and even optimized. We will further show that our information condition is actually a necessary one for a wide class of stochastic signals with decaying dependence. Also, the spatial and temporal independence and stationarity conditions are not required in our theoretical analyses, which indicates that our theory does not exclude applications to feedback systems. Moreover, the distributed algorithms in this paper can track unknown signals that may change constantly, while most of the literature concerns the estimation of time-invariant parameters. Furthermore, the weakest possible information condition used in this paper implies that the distributed adaptive filter can work well even if any individual filter is not stable due to lack of necessary information (i.e., the covariance matrix for each individual regressor is degenerate), which is a natural property for distributed algorithms but has not been justified rigorously in the existing literature.

The remainder of this paper is organized as follows. In section 2, we present the consensus NLMS-based adaptive filter with some notations. In section 3, we give some definitions and conditions to be used in later sections. The main results are stated in section 4. In section 5, we provide the proofs of the main theorems. Finally, we give some simulation results illustrating the advantages of the general information condition in section 6 and give some concluding remarks in section 7.

2. Problem formulation.

Let us consider the following time-varying stochastic linear regression model at sensor $i (i = 1, \dots, n)$

$$y_k^i = (\boldsymbol{\varphi}_k^i)^T \boldsymbol{\theta}_k + v_k^i, \quad k \geq 0, \tag{2.1}$$

where $(\cdot)^T$ denotes transpose operator; y_k^i and v_k^i are scalar observation and noise of sensor i at time k, respectively; $\boldsymbol{\varphi}_k^i$ is the $m \times 1$-dimensional stochastic regressor of sensor i at time k. It is well known that many problems from different application areas can be cast as (2.1) (see, e.g., [23, 33, 38, 40]).

In order to develop a strategy to update the estimation of the $m \times 1$-dimensional system signals or time-varying parameter vector $\boldsymbol{\theta}_k$ in real-time, it is usually convenient to denote the variation of $\boldsymbol{\theta}_k$ as follows:

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \gamma \boldsymbol{\omega}_k, \quad k \geq 1, \tag{2.2}$$

where γ is a nonnegative number reflecting the speed of parameter variate and $\boldsymbol{\omega}_k$ is an undefined $m \times 1$-dimensional vector. We remark that the widely studied time-invariant parameter scenario can be naturally included when $\gamma = 0$. Moreover, different assumptions on the sequences $\{\boldsymbol{\omega}_k\}$ and $\{v_k^i\}$ will result in different theoretical results. In fact, as will be shown later, for stability analysis and tracking error upper bounds, we only need some moment conditions on $\{\boldsymbol{\omega}_k\}$ and $\{v_k^i\}$, while for more accurate performance analysis, $\{\boldsymbol{\omega}_k\}$ and $\{v_k^i\}$ need to have white noise characters.

To start with, we first consider the following class of NLMS algorithm for an individual sensor $i (i = 1, \dots, n)$ [18, 25, 38, 40]:

$$\widehat{\boldsymbol{\theta}}_{k+1}^i = \widehat{\boldsymbol{\theta}}_k^i + \mu \frac{\boldsymbol{\varphi}_k^i}{1 + \| \boldsymbol{\varphi}_k^i \|^2} [y_k^i - (\boldsymbol{\varphi}_k^i)^T \widehat{\boldsymbol{\theta}}_k^i], \qquad k \geq 0, \tag{2.3}$$

where $\mu \in (0, 1]$ is called step size or adaption rate and the increment of the algorithm is opposite to the stochastic gradient of the mean square prediction error

$$e_k^i(\boldsymbol{\theta}) = \mathbb{E}[y_k^i - (\boldsymbol{\varphi}_k^i)^T \boldsymbol{\theta}]^2, \qquad k \geq 0, i = 1, \dots, n,$$

where $\mathbb{E}[\cdot]$ denotes the mathematical expectation operator. Thus it is a type of steepest descent algorithm that aims at minimizing $e_k^i(\boldsymbol{\theta})$ recursively. As will be seen later, the introduced normalization factor $\frac{1}{1+\|\boldsymbol{\varphi}_k^i\|^2}$ will make it possible to deal with unbounded correlated random signals that may have heavy tail and to avoid assuming sufficiently small step size in the theoretical analysis. Here we assume that time instance k of the measurement equation and the iteration instance k of the algorithm are the same, which means that the algorithm is online and updated from time to time using new measurement data. Because of simplicity and robustness, the LMS-like algorithm is one of the most basic algorithms in many areas, such as signal processing, system identification, adaptive filtering, adaptive control, and so on.

In sensor networks, we consider a set of n nodes distributed over the geographic region. The network connections are usually modeled as a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ with the set of nodes $\mathcal{V} = \{1, 2, \dots, n\}$, the set of edges $\mathcal{E}$ where $(i, j) \in \mathcal{E}$ if and only if node j is a neighbor of node i, and the weighted adjacency matrix $\mathcal{A} = \{a_{ij}\}_{n \times n}$ with $0 \leq a_{ij} \leq 1, \sum_{j=1}^n a_{ij} = 1, \forall i, j = 1, \dots, n$, where $a_{ij} > 0 \Leftrightarrow (i, j) \in \mathcal{E}$. The matrix weights a_{ij} can reflect both the topology of the network and the interaction strength among neighboring nodes. Here we assume that matrix $\mathcal{A}$ is symmetrical. Node i denotes the ith sensor, and the set of neighbors of sensor i is denoted as

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

The Laplacian matrix $\mathcal{L}$ of the graph $\mathcal{G}$ is defined as $\mathcal{L} = \mathcal{D} - \mathcal{A}$ [13], where $\mathcal{D} = \text{diag}\{d_1, d_2, \dots, d_n\}$ and $d_i = \sum_{j=1}^n a_{ij} = 1$, i.e., $\mathcal{L} = I_{n \times n} - \mathcal{A}$.

In the following, $\mathbb{R}^{m \times n}$ denotes the set of $m \times n$ matrices. Since we are dealing with vector parameters, most of the matrix manipulations will involve Kronecker product [22], which is defined as follows for any matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{p \times q}$:

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \cdots & a_{mn}B \end{pmatrix} \in \mathbb{R}^{mp \times nq}.$$

In network systems, if sensor i has access only to the information from its neighbors $\{l \in \mathcal{N}_i\}$, we can adopt the following consensus NLMS-based algorithm:

$$\widehat{\boldsymbol{\theta}}_{k+1}^i = \widehat{\boldsymbol{\theta}}_k^i + \mu\left\{\frac{\boldsymbol{\varphi}_k^i}{1+\parallel \boldsymbol{\varphi}_k^i \parallel^2}[y_k^i - (\boldsymbol{\varphi}_k^i)^T\widehat{\boldsymbol{\theta}}_k^i] - \nu \sum_{l\in\mathcal{N}_i} a_{li}(\widehat{\boldsymbol{\theta}}_k^i - \widehat{\boldsymbol{\theta}}_k^l)\right\}, \tag{2.4}$$

where μ and ν are suitably chosen step size parameters in $(0, 1)$.

Remark 2.1. Early work for consensus techniques can be found in [2, 12]. The right-hand side of the above algorithm can be viewed as consisting of two parts: the first part is the usual (normalized) LMS update which tries to minimize the prediction error as explained above, while the second part tries to minimize the weighted distance between estimates of the agent i and its neighbors as explained in [36]. Also, as will be shown in Lemma 5.3 of section 5, the range of the step size parameters μ and ν will make $0 \le \mu\boldsymbol{G}_k \le I_{mn}$, a property to be used in the stability analysis, (see [14] for more information), where $\boldsymbol{G}_k$ is defined below, and this inequallity means that $\mu\boldsymbol{G}_k$ and $I_{mn} - \mu\boldsymbol{G}_k$ are both positive semidefinite matrices.

The goal of this paper is to establish the stability as well as to conduct performance analysis of the above consensus LMS-based adaptive filter under a weakest possible information condition without independence and stationarity assumptions. For convenience of analysis, we introduce the following set of notations:

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

$$\boldsymbol{V}_k \triangleq \text{col}\{v_k^1, \ldots, v_k^n\} \quad (n\times 1), \qquad \boldsymbol{\Omega}_k \triangleq \text{col}\{\underbrace{\boldsymbol{\omega}_k, \ldots, \boldsymbol{\omega}_k}_{n}\} \quad (mn\times 1),$$

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

$$\widetilde{\boldsymbol{\Theta}}_k \triangleq \text{col}\{\widetilde{\boldsymbol{\theta}}_k^1, \ldots, \widetilde{\boldsymbol{\theta}}_k^n\}, \text{ where } \widetilde{\boldsymbol{\theta}}_k^i = \widehat{\boldsymbol{\theta}}_k^i - \boldsymbol{\theta}_k,$$

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

$$\boldsymbol{F}_k \triangleq \boldsymbol{L}_k\boldsymbol{\Phi}_k^T \quad (mn\times mn), \qquad \boldsymbol{G}_k \triangleq \boldsymbol{F}_k + \nu(\mathcal{L}\otimes I_m) \quad (mn\times mn),$$

where col$\{\cdots\}$ denotes a vector by stacking the specified vectors, and 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. Note also that $\boldsymbol{\Omega}_k$ and $\boldsymbol{\Theta}_k$ mean just the n-times replication of identical vectors $\boldsymbol{\omega}_k$ and $\boldsymbol{\theta}_k$, respectively. By (2.1) and (2.2), we have

$$\boldsymbol{Y}_k = \boldsymbol{\Phi}_k^T\boldsymbol{\Theta}_k + \boldsymbol{V}_k \tag{2.5}$$

and

$$\boldsymbol{\Theta}_{k+1} = \boldsymbol{\Theta}_k + \gamma\boldsymbol{\Omega}_{k+1}. \tag{2.6}$$

Then, we have from (2.4)

$$\widehat{\boldsymbol{\Theta}}_{k+1} = \widehat{\boldsymbol{\Theta}}_k + \mu\boldsymbol{L}_k(\boldsymbol{Y}_k - \boldsymbol{\Phi}_k^T\widehat{\boldsymbol{\Theta}}_k) - \mu\nu(\mathcal{L}\otimes I_m)\widehat{\boldsymbol{\Theta}}_k, \tag{2.7}$$

where $\mathcal{L}$ is the Laplacian matrix of graph $\mathcal{G}$. Now, denoting $\widetilde{\boldsymbol{\Theta}}_k = \widehat{\boldsymbol{\Theta}}_k - \boldsymbol{\Theta}_k$, substituting (2.5) into (2.7), and noticing (2.6), we can get

$$\widetilde{\boldsymbol{\Theta}}_{k+1} = \widetilde{\boldsymbol{\Theta}}_k - \mu\boldsymbol{L}_k\boldsymbol{\Phi}_k^T\widetilde{\boldsymbol{\Theta}}_k - \mu\nu(\mathcal{L}\otimes I_m)\widehat{\boldsymbol{\Theta}}_k + \mu\boldsymbol{L}_k\boldsymbol{V}_k - \gamma\boldsymbol{\Omega}_{k+1},$$

and because $(\mathcal{L} \otimes I_m)\boldsymbol{\Theta}_k = 0$, we can obtain

$$\widetilde{\boldsymbol{\Theta}}_{k+1} = \widetilde{\boldsymbol{\Theta}}_k - \mu \boldsymbol{L}_k \boldsymbol{\Phi}_k^T \widetilde{\boldsymbol{\Theta}}_k - \mu\nu(\mathcal{L} \otimes I_m)\widetilde{\boldsymbol{\Theta}}_k + \mu \boldsymbol{L}_k \boldsymbol{V}_k - \gamma \boldsymbol{\Omega}_{k+1}.$$

By the notation $\boldsymbol{G}_k$, we have the following error equation for the consensus NLMS algorithm:

$$\widetilde{\boldsymbol{\Theta}}_{k+1} = (I_{mn} - \mu \boldsymbol{G}_k)\widetilde{\boldsymbol{\Theta}}_k + \mu \boldsymbol{L}_k \boldsymbol{V}_k - \gamma \boldsymbol{\Omega}_{k+1}. \tag{2.8}$$

As is well known, a prime problem in any filtering theory is to assess and quantify the estimation accuracy, expressed normally by the mean square tracking error. It is easy to see from the above distributed filtering error equation (2.8) that such an error hinges on the stability of the homogeneous part of this equation, which depends essentially on the properties of PRM. It is well-known that the analysis of PRM is a difficult mathematical problem in general, unless the random matrices involved are independent and stationary. One of the main purposes of this paper is to overcome this difficulty by using both the specific structure of the consensus LMS and some of our earlier results on single LMS (see [14, 15, 16, 17]).

Now, we give an outline of the main results of the paper. For analyzing the above error equation (2.8) without invoking statistical independence on the system signals, we need to introduce some definitions and notations on the stability of random matrices, as well as some statistical conditions on the system signals and noises, which will be collected and explained in some detail in section 3. Our main theoretical results will be presented in section 4, where both the exponential stability of the algorithm and the tracking error upper bounds will be established under a general cooperative information condition (see Theorems 4.1 and 4.2), followed by some refined theorems on the approximate expression for the true mean square tracking error matrix (see Theorem 4.3 and Corollary 4.4). Moreover, our cooperative information condition is not only general enough to include possible feedback signals but also is a necessary one in some sense as will be shown in Theorem 4.5. The detailed proofs will be given in section 5 with some more technical derivations given in the appendices. Another remarkable feature of our theory is that the "cooperative property" of the consensus adaptive filter can indeed be justified; see section 6.

3. Definitions and conditions.

3.1. Some definitions. Let $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ be two symmetric matrices; then $A \geq B$ means $A - B$ is a positive semidefinite matrix. Let also $\lambda_{max}(\cdot)$ and $\lambda_{min}(\cdot)$ denote the largest and the smallest eigenvalues of matrix $(\cdot)$, respectively. For any matrix $X \in \mathbb{R}^{m\times n}$, the Euclidean norm is defined as $\| X \| = \{\lambda_{max}(XX^T)\}^{\frac{1}{2}}$, and for any random matrix A, its L_p-norm is defined as $\| A \|_{L_p} = \{\mathbb{E}[\| A \|^p]\}^{\frac{1}{p}}$. Throughout the paper, we use $\mathcal{F}_k = \sigma\{\boldsymbol{\varphi}_j^i, \boldsymbol{\omega}_j, v_{j-1}^i, i = 1, \ldots, n, j \leq k\}$ to denote the σ-algebra generated by $\{\boldsymbol{\varphi}_j^i, \boldsymbol{\omega}_j, v_{j-1}^i, i = 1, \ldots, n, j \leq k\}$, where the definition of σ-algebra together with that of conditional mathematical expectation to be used later can be found in [11]. To proceed with further discussions, we need the following definitions introduced in [14].

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-stable.

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

$$S_p(\lambda) = \left\{ A : \left\| \prod_{j=i+1}^{k} (I - A_j) \right\|_{L_p} \leq M\lambda^{k-i} \quad \forall k \geq i+1, \forall i \geq 0, \text{for some } M > 0 \right\}, \tag{3.1}$$

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

For convenience of discussions, we introduce the subclass of $S_1(\lambda)$ for a scalar sequence $a = \{a_k, k \geq 0\}$

$$S^0(\lambda) = \left\{ a : a_k \in [0,1], \mathbb{E}\left[\prod_{j=i+1}^{k} (1 - a_j) \right] \leq M\lambda^{k-i} \quad \forall k \geq i+1, \forall i \geq 0, \text{for some } M > 0 \right\}, \tag{3.2}$$

where $\lambda \in [0, 1)$. Clearly, 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 $\{\beta_k\} \in S^0(\lambda)$.

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 $h \geq 1$,*

$$\left\| \sum_{i=k+1}^{k+h} x_i \right\|_{L_p} \leq C_p^x h^{\frac{1}{2}}. \tag{3.3}$$

Remark 3.1. It is known that the martingale difference, zero mean ϕ- and α-mixing sequences, and the linear process driven by white noises are all in $\mathcal{M}_p$ [16].

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

$$\|A_k\| \leq M b_k \quad \forall k \geq 0. \tag{3.4}$$

3.2. Conditions. To establish a general theory on the consensus adaptive filter, we need the following conditions.

Condition 3.1 (network topology). *The graph $\mathcal{G}$ is connected.*

Remark 3.2. In order to estimate the unknown signals cooperatively, it is natural to require the connectivity of the sensor network; otherwise the distributed adaptive filtering algorithms may not be able to achieve consensus due to the existence of isolated nodes in the network and the lack of necessary information at individual sensors. Moreover, by the connectivity, we know that the Laplacian matrix $\mathcal{L}$ of graph $\mathcal{G}$ has one zero eigenvalue only, with other eigenvalues being positive and not more than 2 [13]. We remark that time-varying network topology may also be considered in a similar way as in [43].

CONDITION 3.2 (information condition). *Let $\{\boldsymbol{\varphi}_k^i, \mathcal{F}_k, k \geq 0\}, i = 1, \dots, n$, be n adapted sequences, and 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*

$$\lambda_k \triangleq \lambda_{min}\left\{\mathbb{E}\left[\frac{1}{n(h+1)}\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\Big|\mathcal{F}_k\right]\right\}, \tag{3.5}$$

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

Remark 3.3. Let us give some intuitive explanations for the above condition used in the paper. To start with, let us first consider the extreme case where the regressor process $\boldsymbol{\varphi}_j^i$ is identically zero. It is clear that Condition 3.2 is not satisfied since the process λ_k is zero, which is indeed a trivial case where the system is not identifiable since the observations contain no information about the unknown parameters. Hence, to estimate the unknown parameters, some nonzero information (or excitation) conditions should be imposed on the regressors $\boldsymbol{\varphi}_j^i$, and such excitation should be persistent for estimating unknown persistently time-varying parameters (see, e.g., [14]), which is usually called the persistence of excitation condition in the single sensor case in the literature. In the theoretical investigations of stochastic systems, such a persistence of excitation condition is usually replaced by both the full rank condition of the covariance matrices of the regressors and some statistical independence conditions. However, as mentioned in the introduction, statistical independence assumptions are too stringent and cannot be satisfied for signals generated from feedback systems. To make the adaptive filtering theory applicable to more general dynamical systems, many efforts have been made in the past and various improved information conditions have been introduced and investigated. For the traditional single sensor case, the conditional mathematical expectation-based information condition as introduced in [14], appears to be quite general and even necessary. Our Condition 3.2 is a suitable generalization of the information condition used in [14] from single sensor to sensor networks and is thus called cooperative information condition. We remark that the assumption on the conditional mathematical expectation in Condition 3.2 implies that the system signals will have some kind of "persistence of excitation" since the prediction of the "future" is nondegenerate given the "past", which is required to track unknown signals that may change constantly.

We remark also that the information condition used in [9] is only a special case of Condition 3.2 with $h = 1$. Moreover, under Condition 3.2 the distributed adaptive filtering network can be shown to fulfill the estimation task cooperatively even when any individual filter cannot, which is a natural property for distributed adaptive filters but has not been justified theoretically before. Furthermore, Condition 3.2 is arguably the weakest possible information condition for the stability of the consensus NLMS algorithm, since it can be shown that it is not only sufficient but also necessary under some extra assumptions; see Theorems 4.1 and 4.5 below. We further remark that the denominator of (3.5) makes the upper bound of λ_k be $\frac{h}{h+1}$, which is needed for the proof of our main results.

CONDITION 3.3. *For some $p \geq 1$, the initial estimation error is bounded, i.e., $\|\widetilde{\boldsymbol{\Theta}}_0\|_{L_{2p}} < \infty$. Furthermore, $\{\boldsymbol{L}_k\boldsymbol{V}_k\} \in \mathcal{M}_{2p}$ and $\{\boldsymbol{\Omega}_k\} \in \mathcal{M}_{2p}$.*

Remark 3.4. This condition simply implies that both the noises and parameter variations are weakly dependent with certain bounded moments.

CONDITION 3.4. *Let $\mathcal{F}_k = \sigma\{\boldsymbol{\varphi}_j^i, \boldsymbol{\omega}_j, v_{j-1}^i, i = 1, \ldots, n, j \le k\}$, and assume that $\forall k \ge 1$,*

$$\begin{aligned} &\mathbb{E}[\boldsymbol{V}_k|\mathcal{F}_k] = 0, \mathbb{E}[\boldsymbol{\Omega}_{k+1}|\mathcal{F}_k] = 0, \mathbb{E}[\boldsymbol{\Omega}_{k+1}\boldsymbol{V}_k^T|\mathcal{F}_k] = 0,\\ &\mathbb{E}[\boldsymbol{V}_k\boldsymbol{V}_k^T|\mathcal{F}_k] = \boldsymbol{P}_v(k) \ge 0,\\ &\mathbb{E}[\boldsymbol{\Omega}_{k+1}\boldsymbol{\Omega}_{k+1}^T] = \boldsymbol{Q}_\omega(k+1) \ge 0,\\ &\sup_k(\|\boldsymbol{V}_k\|_{L_8} + \|\boldsymbol{\Omega}_k\|_{L_8}) < \infty, \end{aligned} \tag{3.6}$$

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

$$\|\mathbb{E}[\boldsymbol{F}_k|\mathcal{F}_{k-t}] - \mathbb{E}[\boldsymbol{F}_k]\|_{L_4} \le \bar{\phi}(t,\mu). \tag{3.7}$$

Remark 3.5. If we are only interested in getting the upper bound of $\widetilde{\boldsymbol{\Theta}}_k$, then some moment conditions on $\boldsymbol{V}_k$ and $\boldsymbol{\Omega}_k$ are sufficient; see Theorem 4.1 below. A refined upper bound can also be obtained under the additional Conditions 3.3; see Theorem 4.2. Moreover, the stronger Condition 3.4 will only be used to obtain the approximate tracking performances as will be shown in Theorem 4.3. Condition 3.4 means that the measurement noise $\boldsymbol{V}_k$ and the parameter variation $\boldsymbol{\Omega}_{k+1}$ have white noise characters, which are commonly used in many works [4, 24, 28] and is a worst case analysis since the future behavior of the model is unpredictable, as mentioned in [16]. 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. We may also note that time-varying covariances $\boldsymbol{P}_v(k)$ and $\boldsymbol{Q}_\omega(k+1)$ are allowed, which may cover some special model drifts of interests. Also, (3.7) describes the decaying correlation between $\boldsymbol{F}_k$ and $\mathcal{F}_{k-t}$, which can be guaranteed by imposing a certain weak dependence condition on the regressor $\{\boldsymbol{\varphi}_k^i\}$, e.g., ϕ-mixing property [16].

4. The main results.

4.1. Stability and performance results. In this section, we first present the L_p-exponential stability of the homogeneous part of (2.8) and a preliminary tracking error bound in the following theorem.

THEOREM 4.1. *Consider the model* (2.5) *and the estimation error equation* (2.8). *Suppose that Conditions* 3.1 *and* 3.2 *are satisfied. For any $\mu \in (0,1)$ and $\nu \in (0,1)$ satisfying $\mu(1+2\nu) \le 1$, then $\{I_{mn} - \mu\boldsymbol{G}_k, k \ge 1\}$ is L_p-exponentially stable ($p \ge 1$). Furthermore, if for some $p \ge 1$ and $\beta > 1$, $\sigma_p \triangleq \sup_k \|\xi_k log^\beta(e+\xi_k)\|_{L_p} < \infty, \|\widetilde{\boldsymbol{\Theta}}_0\|_{L_p} < \infty$ hold where $\xi_k = \|\boldsymbol{V}_k\| + \|\boldsymbol{\Omega}_{k+1}\|$, then $\{\widetilde{\boldsymbol{\Theta}}_k, k \ge 1\}$ is L_p-stable and*

$$\limsup_{k\to\infty} \|\widetilde{\boldsymbol{\Theta}}_k\|_{L_p} \le c[\sigma_p log(e + \sigma_p^{-1})],$$

where c is a positive constant.

Remark 4.1. Intuitively, by Theorem 4.1 we know that when both the noise and the parameter variation are small, the processes ξ_k and σ_p will also be small, and hence the parameter tracking error $\widetilde{\boldsymbol{\Theta}}_k$ will be small too. Here we only require that the observation noise and the parameter variation satisfy a moment condition, and no Gaussian property is required. Note also that the condition $\mu(1+2\nu) \le 1$ reveals how the innovation update gain μ and the consensus update gain ν are related in the distributed adaptive filtering algorithm (2.4). With Condition 3.3, we can further establish a better upper bound on the error $\widetilde{\mathbf{\Theta}}_{k+1}$.

THEOREM 4.2. *Assume that Conditions* 3.1–3.3 *are satisfied; then for any* $\mu \in (0,1)$ *and* $\nu \in (0,1)$ *satisfying* $\mu(1+2\nu) \leq 1$, *we have* $\forall k \geq 0$,

$$\|\widetilde{\mathbf{\Theta}}_{k+1}\|_{L_p} = O\bigg(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}} + (1-\alpha_{2p}\mu)^{k+1}\bigg), \tag{4.1}$$

where $\alpha_{2p} \in (0,1)$ *is a constant which is defined as in Lemma* 5.7 *and* O *is a constant depends only on* α_{2p}.

Remark 4.2. The detailed proofs of Theorems 4.1 and 4.2 are given in section 5. We remark that the upper bound in Theorem 4.2 significantly improves the "crude" bound given in Theorem 4.1, and it roughly indicates the trade-off between noise sensitivity and tracking ability.

Rather than the upper bounds, we may further get the approximate value of the mean square tracking error matrix by strengthening the conditions used in Theorem 4.2. Now, following the ideas in the single sensor case (see, e.g., [16, 23]), to approximate the true mean square tracking error matrix

$$\mathbf{\Pi}_k = \mathbb{E}[\widetilde{\mathbf{\Theta}}_k \widetilde{\mathbf{\Theta}}_k^T],$$

we define the following linear deterministic difference equation for $\widehat{\mathbf{\Pi}}_k$:

$$\widehat{\mathbf{\Pi}}_{k+1} = (I_{mn} - \mu\mathbb{E}[\boldsymbol{G}_k])\widehat{\mathbf{\Pi}}_k(I_{mn} - \mu\mathbb{E}[\boldsymbol{G}_k])^T + \mu^2\mathbb{E}[\boldsymbol{T}_k] + \gamma^2\boldsymbol{Q}_\omega(k+1), \tag{4.2}$$

where

$$\begin{aligned}
&\mathbb{E}[\boldsymbol{T}_k] = \mathbb{E}[\boldsymbol{L}_k\boldsymbol{V}_k\boldsymbol{V}_k^T\boldsymbol{L}_k^T],\\
&\boldsymbol{Q}_\omega(k+1) = \mathbb{E}[\mathbf{\Omega}_{k+1}\mathbf{\Omega}_{k+1}^T],\\
&\widehat{\mathbf{\Pi}}_0 = \mathbb{E}[\widetilde{\mathbf{\Theta}}_0\widetilde{\mathbf{\Theta}}_0^T].
\end{aligned}$$

Note that by Lemma 5.9, the above linear deterministic equation for $\widehat{\mathbf{\Pi}}_k$ is stable, which will be used to approximate the true mean square tracking error matrix $\mathbf{\Pi}_k$. This is the main content of the following theorem.

THEOREM 4.3. *Let Conditions* 3.1–3.4 *hold. Then for any* $\mu \in (0,1)$ *and* $\nu \in (0,1)$ *satisfying* $\mu(1+2\nu) \leq 1$ *we have* $\forall k \geq 1$,

$$\|\mathbf{\Pi}_{k+1} - \widehat{\mathbf{\Pi}}_{k+1}\| \leq c\bar{\delta}(\mu)\bigg[\mu + \frac{\gamma^2}{\mu} + (1-\alpha\mu)^{k+1}\bigg], \tag{4.3}$$

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

$$\bar{\delta}(\mu) \triangleq \min_{t\geq 1}\{\sqrt{\mu}t + \bar{\phi}(t,\mu)\},$$

which tends to zero as μ *approaches zero.*

Remark 4.3. The detailed proof of Theorem 4.3 is given in section 5. By Theorem 4.3, $\widehat{\mathbf{\Pi}}_k$ provides a good approximation to the mean square tracking error matrix $\mathbf{\Pi}_k$ for small parameter variation γ and small adaptation gain μ, since $\bar{\delta}(\mu)$ tends to zero as μ tends to zero. Note that the explicit solution of $\widehat{\mathbf{\Pi}}_k$ can be obtained in terms of $\mathbb{E}[\boldsymbol{G}_k], \mathbb{E}[\boldsymbol{T}_k], \boldsymbol{Q}_\omega(k+1)$ and the initial condition $\widehat{\mathbf{\Pi}}_0$, based on which the value of $\widehat{\mathbf{\Pi}}_k$ can be calculated and examined conveniently in terms of both μ and γ. This point will become clearer in the special case of (wide-sense) stationary signals, and the above theorem can be further simplified as illustrated by the following corollary.

COROLLARY 4.4. *Let*

$$\boldsymbol{F} = \mathbb{E}[\boldsymbol{F}_k] = diag\{\boldsymbol{F}^1, \cdots, \boldsymbol{F}^n\}, \boldsymbol{G} = \boldsymbol{F} + \nu(\mathcal{L} \otimes I_m),$$
$$\boldsymbol{T} = \mathbb{E}[\boldsymbol{T}_k] = \mathbb{E}[\boldsymbol{L}_k \boldsymbol{V}_k \boldsymbol{V}_k^T \boldsymbol{L}_k^T], \boldsymbol{Q}_\omega = \boldsymbol{Q}_\omega(k+1).$$

Then, under the conditions of Theorem 4.3*, we have*

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

where $o(1)$ *is a term which tends to zero as* $k \to \infty$*, and*

$$\bar{\boldsymbol{R}}_v = \int_0^\infty e^{-\boldsymbol{G}t}\boldsymbol{T}e^{-\boldsymbol{G}t}dt, \quad \bar{\boldsymbol{R}}_\omega = \int_0^\infty e^{-\boldsymbol{G}t}\boldsymbol{Q}_\omega e^{-\boldsymbol{G}t}dt.$$

Remark 4.4. Since the proof of Corollary 4.4 is similar to [16, Corollary 4.2], here we omit it. Note that $\lim_{\mu\to 0}\bar{\delta}(\mu) = 0$. As a result, we have $\forall$ small μ and large k:

$$\mathbf{\Pi}_k \sim \mu \bar{\boldsymbol{R}}_v + \frac{\gamma^2}{\mu}\bar{\boldsymbol{R}}_\omega. \tag{4.4}$$

Consequently, by taking "trace" ($Tr(\cdot)$) on both sides and by the definition of $\mathbf{\Pi}_k$, we have

$$\sum_{i=1}^n \mathbb{E}[\|\widetilde{\boldsymbol{\theta}}_k^i\|^2] \sim \mu Tr(\bar{\boldsymbol{R}}_v) + \frac{\gamma^2}{\mu} Tr(\bar{\boldsymbol{R}}_\omega), \tag{4.5}$$

which indicates that μ should be proportional to γ, and by minimizing the right-hand side, we get the "optimal" choice $\mu^* = \gamma\sqrt{Tr(\bar{\boldsymbol{R}}_\omega)/Tr(\bar{\boldsymbol{R}}_v)}$ with the corresponding minimum value:

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

Remark 4.5. When the regressors are independent and the parameter process is a random walk, the steady-state mean square tracking error was derived for different distributed adaptive filtering algorithms (e.g., [28]), but most of the literature in the analysis of the distributed adaptive filtering algorithms is concerned with the special case of constant parameter (i.e., $\gamma = 0$), under either independence conditions (e.g., [37]), or stationarity and ergodicity conditions ([36]), where the mean square errors were also approximated recursively for different distributed strategies. We remark that the existing theoretical results in distributed adaptive filters do not seem to be applicable to the natural case where any individual sensor does not have the capability of tracking unknown signals due to lack of necessary information (e.g., when the information matrix $\mathbb{E}[\boldsymbol{\varphi}_k^i(\boldsymbol{\varphi}_k^i)^T]$ degenerates at a single sensor i), whereas our results can be applied to this important case, thanks to the cooperative information condition used in the paper. This fact will be further illustrated in the simulation examples to be given in section 6.

We remark that the trade-off structure of (4.4) for the tracking error depends on both the adaptation gain μ and the speed of parameter variation γ, which is similar to the formula in the traditional single sensor case for both independent signals [18, 23, 25, 40] and correlated signals [16, 17]. Clearly, (4.4) shows that the smaller the adaptation gain, the slower the parameter variation can be effectively dealt with. As will be shown in the next section, in the current network case, the connectivity of the network and the cooperative information condition (Conditions 3.1 and 3.2) guarantee the positive definite property of the matrix $\boldsymbol{G}$, and hence the finite value of the matrices $\bar{\boldsymbol{R}}_v$ and $\bar{\boldsymbol{R}}_\omega$ in (4.4).

4.2. Necessity of the information condition. In this subsection, we will further show that Condition 3.2 used in this paper is not only sufficient but also necessary for the stability of the consensus LMS-based adaptive filter under some extra conditions on dependence, for example, the ϕ-mixing properties. A random process $\{\xi_k\}$ is called ϕ-mixing if there exists a sequence $\phi(n) \to 0$ as $n \to \infty$ such that

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

where $\mathcal{F}_t^s = \sigma\{\xi(u), t \le u \le s\}$. It is known that for any $\mathcal{F}_t^\infty$-measurable f_t, with $|f_t| \le 1$, the following inequality holds [21]:

$$|\mathbb{E}[f_{t+h}|\mathcal{F}_0^t] - \mathbb{E}[f_{t+h}]| \le 2\phi(h) \quad \forall t, h.$$

As is well known, the ϕ-mixing process includes a large class of important processes, for example, deterministic processes, M-dependent processes, and processes generated from bounded white noise filtered through a stable finite-dimensional linear filter [14]. The following theorem gives necessary and sufficient conditions for the L_p-exponential stability of such processes.

THEOREM 4.5. *Let $\{\boldsymbol{\varphi}_k^i\}$ be ϕ-mixing processes, and suppose that Condition* 3.1 *is satisfied. Then for any $\mu \in (0,1)$ and $\nu \in (0,1)$ satisfying $\mu(1+2\nu) \le 1$, $\{I_{mn} - \mu\boldsymbol{G}_k, k \ge 1\}$ is L_p-exponentially stable $(p \ge 1)$ if and only if Condition* 3.2 *holds.*

Remark 4.6. The ϕ-mixing property used in the above theorem is just for simplicity, which can be further relaxed for Theorem 4.5; see [17] for related discussions.

5. Proofs of the main results.

5.1. Proof of Theorem 4.1. Before proving Theorem 4.1, we first introduce some preliminary lemmas. Firstly, for the Kronecker product, we have the following result.

LEMMA 5.1 (see [22]). *Let $\lambda_i (i = 1, \dots, n), \mu_j (j = 1, \dots, m)$ be the eigenvalues of matrix $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times m}$, respectively. Then the eigenvalues of matrix $A \otimes B$ are $\lambda_1\mu_1, \dots, \lambda_1\mu_m, \lambda_2\mu_1, \dots, \lambda_2\mu_m, \dots, \lambda_n\mu_1, \dots, \lambda_n\mu_m$. Moreover, if $x_1, \dots, x_p$ are linearly independent right eigenvectors of A corresponding to eigenvalues $\lambda_1, \dots, \lambda_p (p \le n)$ and $y_1, \dots, y_q$ are linearly independent right eigenvectors of B corresponding to eigenvalues $\mu_1, \dots, \mu_q (q \le m)$, then $x_i \otimes y_j$ are linearly independent right eigenvectors of $A \otimes B$ corresponding to $\lambda_i\mu_j$.*

In addition, we also need the following lemmas in the proof of Corollary 5.5.

LEMMA 5.2 (see [14]). *Let $\{\alpha_k\} \in S^0(\lambda)$ and $\alpha_k \le \alpha^* < 1$ $\forall k \ge 0$, where α^* is a constant; then for any $\epsilon \in (0,1)$, $\{\epsilon\alpha_k\} \in S^0(\lambda^{(1-\alpha^*)\epsilon})$.*

LEMMA 5.3. *For any* $\mu \in (0,1)$ *and* $\nu \in (0,1)$ *satisfying* $\mu(1+2\nu) \le 1$, *we have* $0 \le \mu \boldsymbol{G}_k \le I_{mn}$.

Proof. By Condition 3.1, matrix $\mathcal{L}$ has n real eigenvalues in an ascending order

$$0 = t_1 < t_2 \le t_3 \le \cdots \le t_n \le 2.$$

Correspondingly, by Lemma 5.1 we know that $\mathcal{L} \otimes I_m$ has mn real eigenvalues in an ascending order

$$0 = l_1 = l_2 = \cdots = l_m < l_{m+1} \le l_{m+2} \le \cdots \le l_{mn} \le 2. \tag{5.1}$$

Then the matrix $\mathcal{L} \otimes I_m$ is positive semidefinite and $\mathcal{L} \otimes I_m \le 2I_{mn}$. Since $\mu \in (0,1)$ and $\nu \in (0,1)$ satisfying $\mu(1+2\nu) \le 1$, the conclusion holds by the definition of $\boldsymbol{G}_k$. ☐

LEMMA 5.4. *Suppose that Condition* 3.1 *is satisfied. For any matrices* $A_j^i \in \mathbb{R}^{m\times m}, j \ge 0, i = 1, \ldots, n$, *with* $0 \le A_j^i \le cI_m$, *where* $c > 0$ *is a constant, denote* $\boldsymbol{A}_j = diag\{A_j^1, \ldots, A_j^n\}$. *Then for any* $k \ge 0, h \ge 1$, *we have*

$$\lambda_{min}\bigg(\sum_{j=k+1}^{k+h}[\boldsymbol{A}_j + \nu(\mathcal{L}\otimes I_m)]\bigg) \ge \sigma\lambda_{min}\bigg(\sum_{i=1}^{n}\sum_{j=k+1}^{k+h} A_j^i\bigg), \tag{5.2}$$

where $\nu > 0, \sigma = \frac{l_{m+1}\nu}{2nc + l_{m+1}\nu n}$ *and* l_{m+1} *is the* $(m+1)$*th smallest eigenvalue of* $\mathcal{L}\otimes I_m$, *which equals to the second smallest eigenvalue of Laplacian matrix* $\mathcal{L}$ *by Lemma* 5.1.

Proof. See Appendix A. ☐

The following result plays a key role in the proof of Theorem 4.1.

COROLLARY 5.5. *For any* $\mu \in (0,1)$ *and* $\nu \in (0,1)$ *satisfying* $\mu(1+2\nu) \le 1$, *suppose that Conditions* 3.1 *and* 3.2 *are satisfied; then* $\rho_k \in S^0(\rho)$, *where*

$$\rho_k \triangleq \lambda_{min}\bigg\{\mathbb{E}\bigg[\frac{1}{1+h}\sum_{j=k+1}^{k+h}\mu\boldsymbol{G}_j\Big|\mathcal{F}_k\bigg]\bigg\}, \tag{5.3}$$

and $\rho = \lambda^{\epsilon}, \epsilon = \frac{l_{m+1}\nu\mu}{(1+h)(2+\nu l_{m+1})}$ *and* l_{m+1} *is the* $(m+1)$*th smallest eigenvalue of* $\mathcal{L}\otimes I_m$, *which equals to the second smallest eigenvalue of matrix* $\mathcal{L}$.

Proof. For any $k \ge 0$, let $A_{k,j}^i = \mathbb{E}\Big[\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\big|\mathcal{F}_k\Big], i = 1, \ldots, n, j = k+1, \ldots, k+h$, where h is defined in Condition 3.2. Then for any $k \ge 0$, by Lemma 5.4 and since $0 \le \boldsymbol{F}_k \le I_{mn}$, we have

$$\frac{\rho_k}{\mu} \ge \frac{l_{m+1}\nu}{2n + l_{m+1}\nu n} n\lambda_k = \frac{l_{m+1}\nu}{2 + l_{m+1}\nu}\lambda_k = \sigma^{'}\lambda_k,$$

where λ_k is defined in (3.5) and $0 < \sigma^{'} = \frac{l_{m+1}\nu}{2+l_{m+1}\nu} < 1$.

From Lemma 5.3, we know that $\rho_k \in [0,1]$. By Lemma 5.2 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 $\rho = \lambda^{\epsilon}$ and

$$\epsilon = (1-\alpha^*)\sigma^{'} = \frac{l_{m+1}\nu\mu}{(h+1)(2+l_{m+1}\nu)} > 0.$$

This completes the proof. ☐

As will be seen shortly, the constant ρ will determine the rate of exponential convergence of the homogeneous part of (2.8). Note that λ of Condition 3.2 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 $\rho = \lambda^{\epsilon}$ shows explicitly how the stability of the distributed adaptive filtering algorithms is connected with the cooperativity of the system information, the connectivity of the network topology, as well as the adaptation gains μ and ν.

LEMMA 5.6 (see [14]). *Let $\{A_k^i, \mathcal{F}_k\}, i = 1,\dots,n$, be n adapted sequences of random matrices and $0 \le A_k^i \le I$ $\forall i = 1,\dots,n$, and denote $\boldsymbol{A}_k = diag\{A_k^1,\dots,A_k^n\}$. If there exists an integer $h > 0$ such that $\{\zeta_k\} \in S^0(\zeta), \zeta \in (0,1)$, where*

$$\zeta_k \triangleq \lambda_{min}\left\{\mathbb{E}\left[\frac{1}{1+h}\sum_{j=k+1}^{k+h}\boldsymbol{A}_j\Big|\mathcal{F}_k\right]\right\}, \tag{5.4}$$

then $\{\boldsymbol{A}_k\} \in S_p(\zeta^{\alpha})$, where

$$\alpha = \begin{cases} \frac{1}{8h(1+h)^2}, & 1 \le p \le 2,\\ \frac{1}{4h(1+h)^2 p}, & p > 2. \end{cases} \tag{5.5}$$

Notice that Lemma 5.6 will be used to connect the information condition with the L_p-exponential stability of the homogeneous part of the algorithm, which in conjunction with Corollary 5.5, makes it possible to accomplish the proof of Theorem 4.1 as follows.

By Corollary 5.5 and Lemma 5.6, we know that

$$\{\mu\boldsymbol{G}_k\} \in S_p(\rho^{\alpha}),$$

where α is defined by (5.5). Then by Definition 3.2, it is obvious that $\{I_{mn} - \mu\boldsymbol{G}_k, k \ge 1\}$ is L_p-exponentially stable ($p \ge 1$). Moreover, the upper bound for the error can be derived in a similar way as that of [14, Theorem 4.2], details will be omitted here.

5.2. Proof of Theorem 4.2.

LEMMA 5.7. *Suppose that Conditions* 3.1 *and* 3.2 *are satisfied. Then for any $p \ge 2$ such that for any $\mu \in (0,1)$ and $\nu \in (0,1)$ satisfying $\mu(1+2\nu) \le 1$ and $\forall k \ge i+1 > 0$*

$$\left\|\prod_{j=i+1}^{k}(I_{mn} - \mu\boldsymbol{G}_j)\right\|_{L_p} \le M_p(1-\mu\alpha_p)^{k-i}, \tag{5.6}$$

where M_p and α_p are positive constants depending on $\{\boldsymbol{G}_j, j > 0\}$ and p.

Proof. Denote

$$b = \frac{l_{m+1}\nu}{(1+h)(2+\nu l_{m+1})}.$$

By the proof of Theorem 4.1, we have

$$\left\|\prod_{j=i+1}^{k}(I_{mn} - \mu\boldsymbol{G}_j)\right\|_{L_p} \le M\{\beta_p\}^{\mu(k-i)},$$

where

$$\beta_p = \lambda^{\frac{b}{4h(1+h)^2 p}},$$

and l_{m+1} is the $(m+1)$th smallest eigenvalue of $\mathcal{L} \otimes I_m$, which equals to the second smallest eigenvalue of matrix $\mathcal{L}$.

Note that $\beta_p \in (0,1)$. Then we define a constant $\alpha_p \in (0,1)$ as $\alpha_p = 1 - \beta_p = 1 - \lambda^{\frac{b}{4h(1+h)^2 p}}$. 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 of this lemma. ∎

By Lemma 5.7, and since the proof of Theorem 4.2 is similar to Lemma A.3 in [16], here we omit that proof.

5.3. Proof of Theorem 4.3. Before proving the theorem, we first give the following lemma in [14].

LEMMA 5.8 (see [14]). *Let $\{\boldsymbol{A}_i, \mathcal{F}_i\}$ be an adapted sequence of random matrices and $0 \le \boldsymbol{A}_i \le I$. If $\{A_i\} \in S_1(\lambda)$ for some $\lambda \in [0,1)$, then there exists an integer $h > 0$ such that*

$$\inf_k \lambda_{min}\left\{\sum_{j=kh+1}^{(k+1)h} \mathbb{E}[\boldsymbol{A}_j]\right\} > 0. \tag{5.7}$$

LEMMA 5.9. *Suppose that Conditions 3.1 and 3.2 are satisfied. Then there exist constants $M > 0$ and $\beta \in (0,1)$ which depend only on sequence $\{\boldsymbol{G}_k\}$ such that for any $\mu \in (0,1)$ and $\nu \in (0,1)$ satisfying $\mu(1+2\nu) \le 1$ and $\forall k \ge i+1 > 0$*

$$\left\| \prod_{j=i+1}^{k} (I_{mn} - \mu\mathbb{E}[\boldsymbol{G}_j]) \right\| \le M(1-\mu\beta)^{k-i}. \tag{5.8}$$

Proof. Take any $\mu_0 \in (0,1)$ satisfying $\mu_0(1+2\nu) \le 1$. Since $0 \le \mu_0\boldsymbol{G}_k \le I_{mn}$, by Lemma 5.8 and Theorem 4.1, we know that there exist constants $h_0 > 0$ and $\delta_0 > 0$ such that

$$\sum_{j=kh_0+1}^{(k+1)h_0} \mu_0\mathbb{E}[\boldsymbol{G}_j] \ge \delta_0 I_{mn} \quad \forall k \ge 0.$$

Consequently, we have for any $\mu \in (0,1)$ satisfying $\mu(1+2\nu) \le 1$,

$$\sum_{j=kh_0+1}^{(k+1)h_0} \mu\mathbb{E}[\boldsymbol{G}_j] \ge \mu\delta I_{mn} \quad \forall k \ge 0,$$

where $\delta = \frac{\delta_0}{\mu_0}$. Hence for deterministic sequence $\{\mathbb{E}[\mu\boldsymbol{G}_k], k \ge 0\}$ by [14, Theorem 2.1], we have

$$\left\| \prod_{j=kh_0+1}^{(k+1)h_0} (I_{mn} - \mu\mathbb{E}[\boldsymbol{G}_j]) \right\| \le \left\{1 - \frac{\mu\delta}{1+h_0}\right\}^{1/2}.$$

It is easy to know that there exist constants $\beta \in (0,1)$ and $M > 0$ which depend on sequence $\{\boldsymbol{G}_k\}$,

$$\left\| \prod_{j=i+1}^{k} (I_{mn} - \mu\mathbb{E}[\boldsymbol{G}_j]) \right\| \le M(1-\mu\beta)^{k-i}.$$

This completes the proof. ∎

The proof of Theorem 4.3 is as follows. Let us define the following new sequence:

$$\bar{\boldsymbol{\Theta}}_{k+1} = (I_{mn} - \mu\mathbb{E}[\boldsymbol{G}_k])\bar{\boldsymbol{\Theta}}_k + \mu\boldsymbol{L}_k\boldsymbol{V}_k - \gamma\boldsymbol{\Omega}_{k+1}, \tag{5.9}$$

where $\bar{\boldsymbol{\Theta}}_0 = \widetilde{\boldsymbol{\Theta}}_0$. By (5.9) it is evident that

$$\widehat{\boldsymbol{\Pi}}_k = \mathbb{E}[\bar{\boldsymbol{\Theta}}_k\bar{\boldsymbol{\Theta}}_k^T], \quad k \geq 0.$$

Hence by the Schwarz inequality

$$\begin{aligned} \|\boldsymbol{\Pi}_{k+1} - \widehat{\boldsymbol{\Pi}}_{k+1}\| =& \|\mathbb{E}[\widetilde{\boldsymbol{\Theta}}_{k+1}\widetilde{\boldsymbol{\Theta}}_{k+1}^T - \bar{\boldsymbol{\Theta}}_{k+1}\bar{\boldsymbol{\Theta}}_{k+1}^T]\| \\ \leq& \|\widetilde{\boldsymbol{\Theta}}_{k+1} - \bar{\boldsymbol{\Theta}}_{k+1}\|_{L_2}(\|\widetilde{\boldsymbol{\Theta}}_{k+1}\|_{L_2} + \|\bar{\boldsymbol{\Theta}}_{k+1}\|_{L_2}). \end{aligned} \tag{5.10}$$

Similar to the proof of Theorem 4.2, and using Lemma 5.9, it is easy to obtain

$$\|\bar{\boldsymbol{\Theta}}_{k+1}\|_{L_2} = O\left(\sqrt{\mu} + \frac{\gamma}{\sqrt{\mu}} + (1-\alpha\mu)^{k+1}\right), \tag{5.11}$$

where $\alpha \in (0,1)$ is a constant (without loss of generality, it may be taken as the same as that in Theorem 4.2).

The following proof is similar to [15, Theorem 4.1]; here we omit it, and we know that by Conditions 3.1–3.4, (4.3) holds.

5.4. Proof of Theorem 4.5. Before proving the theorem, we first give some critical lemmas.

LEMMA 5.10. *Suppose that Condition* 3.1 *is satisfied. For any matrices* $A_j^i \in \mathbb{R}^{m\times m}, j \geq 0, i = 1, \dots, n$, *with* $0 \leq A_j^i \leq cI_m$, *where* $c > 0$ *is a constant, we denote* $\boldsymbol{A}_j = diag\{A_j^1, \dots, A_j^n\}$. *If there exists a constant* $h \geq 1$ *such that*

$$\inf_k \lambda_{min}\left(\sum_{j=k+1}^{k+h}[\boldsymbol{A}_j + \nu(\mathcal{L}\otimes I_m)]\right) > 0, \tag{5.12}$$

then we have

$$\inf_k \lambda_{min}\left(\sum_{i=1}^{n}\sum_{j=k+1}^{k+h} A_j^i\right) > 0. \tag{5.13}$$

Proof. See Appendix B. ∎

COROLLARY 5.11. *For any* $\mu \in (0,1)$ *and* $\nu \in (0,1)$ *satisfying* $\mu(1+2\nu) \leq 1$, *if Condition* 3.1 *is satisfied and* $\{\mu\boldsymbol{G}_k\} \in S_1(\lambda)$, *then there exists* $h > 0$ *such that*

$$\inf_k \lambda_{min}\left\{\sum_{i=1}^{n}\sum_{j=k+1}^{k+h}\mathbb{E}\left[\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\right]\right\} > 0. \tag{5.14}$$

Proof. By the conditions in Corollary 5.11 and using Lemma 5.8, we know that there exists an integer $h > 0$ such that

$$\inf_k \lambda_{min}\left\{\sum_{j=k+1}^{k+h}\mathbb{E}[\boldsymbol{G}_j]\right\} > 0.$$

Then denote $A_j^i = \mathbb{E}\left[\frac{\boldsymbol{\varphi}_j^i(\boldsymbol{\varphi}_j^i)^T}{1+\|\boldsymbol{\varphi}_j^i\|^2}\right], i = 1, \dots, n, j \geq 0$, and by Lemma 5.10, we know (5.14) holds. ∎

We remark that the converse assertion of Corollary 5.11 may not true in general, and this can be seen from [14, Example 2.1]. However, it will be true if we impose additional assumptions on $\{\boldsymbol{\varphi}_k^i\}$, for example, the ϕ-mixing properties. Next, we prove Theorem 4.5.

Sufficiency. By Theorem 4.1, $\{I_{mn} - \mu \boldsymbol{G}_k, k \geq 1\}$ is L_p-exponentially stable ($p \geq 1$).

Necessity. By Corollary 5.11, we can obtain (5.14). Since $\{\boldsymbol{\varphi}_k^i\}$ is ϕ-mixing, we know that $\left\{ \sum_{i=1}^n \frac{\boldsymbol{\varphi}_k^i (\boldsymbol{\varphi}_k^i)^T}{1+\|\boldsymbol{\varphi}_k^i\|^2} \right\}$ is also ϕ-mixing. Then by (5.14), the ϕ-mixing property, and [14, Theorem 2.3], we know that Condition 3.2 holds. This completes the proof.

6. Simulation results. We now provide two examples to illustrate and demonstrate the theoretical results obtained in this paper.

Example 1. We will provide an example of nonindependent signals to illustrate that the approximate formula (4.5) for the filtering error is quite accurate and that the whole sensor network can track the unknown parameters effectively even if none of the sensors can track the unknown parameters individually.

Let us consider a network with 20 sensors, which is shown in Figure 6.1, and the corresponding graph is connected. Here we utilize a conventional way to construct the combination weights, which is called the Metropolis rule [41], i.e.,

$$a_{li} = \begin{cases} 1 - \sum_{j \neq i} a_{ij} & \text{if } \; l = i, \\ 1/\max\{n_i, n_l\} & \text{if } \; l \in \mathcal{N}_i \setminus \{k\}, \end{cases}$$

where n_i denotes the degree of the node i, i.e., the size of its neighborhood (without the node i itself). By the definition of the Metropolis rule, we know that the adjacency matrix $\mathcal{A}$ is symmetrical.

We will estimate or track an unknown signal $\boldsymbol{\theta}_k \in \mathbb{R}^{10}$, and let every entry of $\boldsymbol{\theta}_0$ be 1. Here we consider two cases: $\gamma = 0$ ($\boldsymbol{\theta}_k$ is time-invariant) and $\gamma = 0.1$ ($\boldsymbol{\theta}_k$ is

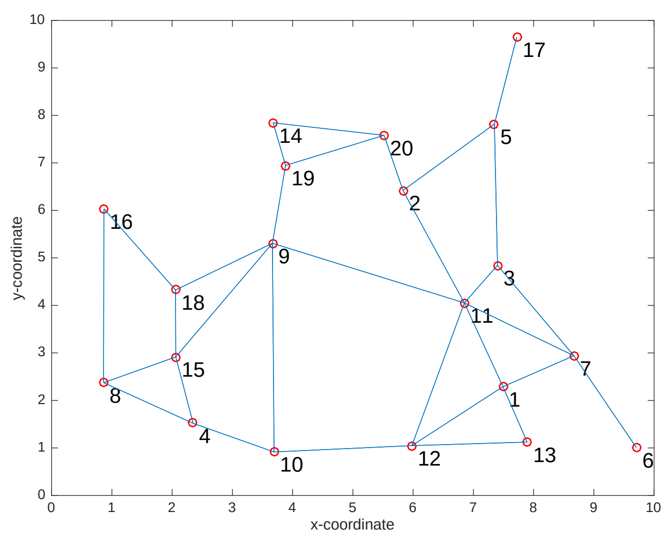

FIG. 6.1. *Network topology.*

time-varying) with the parameter variation $\boldsymbol{\omega}_k \sim N(0, 0.2^2, 10, 1)$ (Gaussian distribution with 0 mean and 0.2^2 variance) in (2.2). In both cases, the observation noises $\{v_k^i, k \geq 1, i = 1, \ldots, 20\}$ in (2.1) are independent and identically distributed with $v_k^i \sim N(0, \sigma_i^2, 1, 1)$, where $\sigma_i^2 (i = 1, \ldots, 20)$ are randomly generated over $[0.03, 0.3]$. Let $\boldsymbol{\varphi}_k^i \in \mathbb{R}^{10} (i = 1, \ldots, 20)$ be 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 $\boldsymbol{x}_k^i \in \mathbb{R}^{10}, A_i \in \mathbb{R}^{10\times 10}, B_i \in \mathbb{R}^{10}, C_i \in \mathbb{R}^{10\times 10}$, and $\xi_k^i \in \mathbb{R}$. Let $\{\xi_k^i, k \geq 1, i = 1, \ldots, 20\}$ be independent and identically distributed with $\xi_k^i \sim N(0, 0.3^2, 1, 1)$, and

$$\begin{aligned}
&A_{1+3j} = A_{2+3j} = \text{diag}\{\underbrace{1/2, \ldots, 1/2}_{10}\}, (j = 0, \ldots, 6),\\
&A_3 = A_{15} = [\text{col}\{\underbrace{4/5, \ldots, 4/5}_{10}\}, \mathbf{0}_{10\times 9}], A_6 = A_{18} = [\mathbf{0}_{10\times 3}, \text{col}\{\underbrace{4/5, \ldots, 4/5}_{10}\}, \mathbf{0}_{10\times 6}],\\
&A_9 = [\mathbf{0}_{10\times 6}, \text{col}\{\underbrace{4/5, \ldots, 4/5}_{10}\}, \mathbf{0}_{10\times 3}], A_{12} = [\mathbf{0}_{10\times 9}, \text{col}\{\underbrace{4/5, \ldots, 4/5}_{10}\}],\\
&B_1 = B_3 = [1; \mathbf{0}_{9\times 1}], B_2 = [0; 1; \mathbf{0}_{8\times 1}],\\
&B_4 = B_6 = B_{16} = B_{18} = [\mathbf{0}_{3\times 1}; 1; \mathbf{0}_{6\times 1}], B_5 = B_{17} = [\mathbf{0}_{4\times 1}; 1; \mathbf{0}_{5\times 1}],\\
&B_7 = B_9 = B_{19} = [\mathbf{0}_{6\times 1}; 1; \mathbf{0}_{3\times 1}], B_8 = B_{20} = [\mathbf{0}_{7\times 1}; 1; \mathbf{0}_{2\times 1}],\\
&B_{10} = B_{12} = [\mathbf{0}_{7\times 1}; 1; \mathbf{0}_{2\times 1}], B_{11} = [\mathbf{0}_{8\times 1}; 1; \mathbf{0}_{1\times 1}],\\
&B_{13} = B_{15} = [\mathbf{0}_{8\times 1}; 1; 0], B_{14} = [\mathbf{0}_{9\times 1}; 1],\\
&C_1 = \text{diag}\{\underbrace{1, 0, \ldots, 0}_{10}\}, C_2 = \text{diag}\{\underbrace{0, 1, \ldots, 0}_{10}\}, \ldots, C_{10} = \text{diag}\{\underbrace{0, 0, \ldots, 1}_{10}\},\\
&C_{11} = C_{12} = C_{13} = C_{14} = C_{15} = C_{10},\\
&C_{16} = C_4, C_{17} = C_5, C_{18} = C_6, C_{19} = C_7, C_{20} = C_8,
\end{aligned}$$

where $\mathbf{0}_{s\times t}$ denotes an $s \times t$-dimensional matrix with all entries being zero. It can be verified that Condition 3.2 is satisfied with $h = 2$, but is not for $h = 1$ [9]. Moreover, it is not difficult to verify that the necessary information condition in [14] is not satisfied for any individual sensor, since all the 20 subsystems are not observable. Let

$$\boldsymbol{x}_0^i = \text{col}\{\underbrace{1, \ldots, 1}_{10}\}, \quad \hat{\boldsymbol{\theta}}_0^i = \text{col}\{\underbrace{0, \ldots, 0}_{10}\} (i = 1, \ldots, 20),$$

and $\mu = 0.5, \nu = 0.5$. Here we repeat the simulation for $m = 500$ times with the same initial states. Then for sensor $i (i = 1, \ldots, 20)$, we can get m sequences

$$\{\|\hat{\boldsymbol{\theta}}_k^{i,j} - \boldsymbol{\theta}_k^j\|^2, k = 1, 100, 200, \ldots, 2000\}, \qquad j = 1, \ldots, m,$$

where the superscript j denotes the jth simulation result. We use

$$\frac{1}{20m} \sum_{i=1}^{20} \sum_{j=1}^{m} \|\hat{\boldsymbol{\theta}}_k^{i,j} - \boldsymbol{\theta}_k^j\|^2, \qquad k = 1, 100, 200, \ldots, 2000,$$

to approximate the estimation or tracking errors with $\gamma = 0$ and $\gamma = 0.1$ in Figure 6.2.

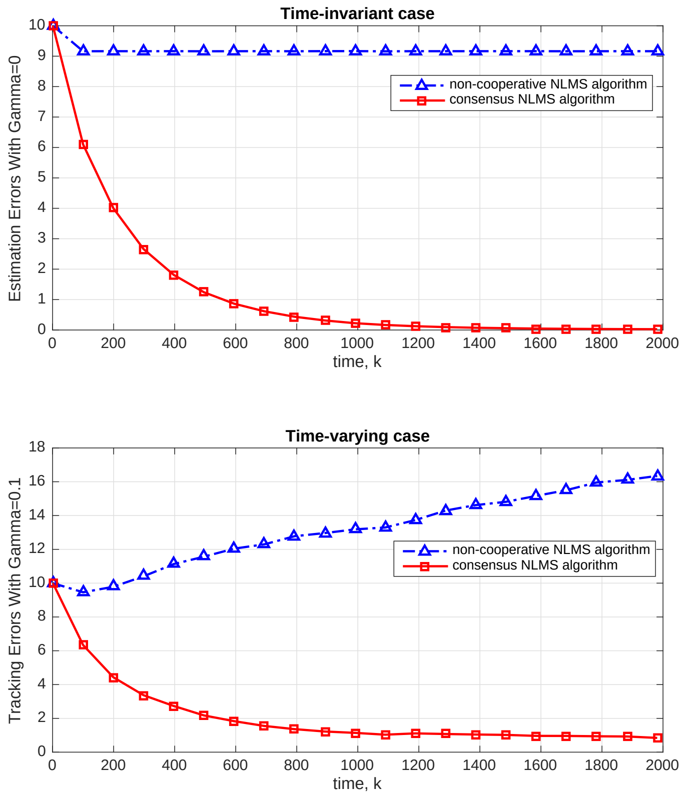

FIG. 6.2. *Estimation and tracking errors in Example* 1.

The upper graph of Figure 6.2 is the time-invariant case. If all the sensors use the noncooperative NLMS algorithm to estimate $\boldsymbol{\theta}_k$, the estimation errors are quite large, because all the sensors do not satisfy the information condition in [14]. However, if they utilize the consensus NLMS algorithm, the estimation errors converge to a small neighborhood of zero as k increases, since the whole system satisfies Condition 3.2. In the lower graph of Figure 6.2, $\boldsymbol{\theta}_k$ is time-varying. The tracking errors keep growing as k increases when all the sensors use the noncooperative NLMS algorithm, and the mean square tracking errors converge nicely as k increases when they all apply the consensus NLMS algorithm. In fact, in the time-varying case, we have the empirical estimation of the mean square error $\frac{1}{m}\sum_{i=1}^{20}\sum_{j=1}^{m}\|\hat{\boldsymbol{\theta}}_{2000}^{i,j}-\boldsymbol{\theta}_{2000}^{j}\|^2 = 17.964$.

On the other hand, using the formula (4.5) in Corollary 4.4, we can obtain $tr\bar{\boldsymbol{R}}_v \approx 0.8014$ and $tr\bar{\boldsymbol{R}}_\omega \approx 917.2435$; then we have $\sum_{i=1}^{20}\mathbb{E}\|\widetilde{\boldsymbol{\theta}}_{2000}^{i}\|^2 \sim \mu tr\bar{\boldsymbol{R}}_v + \frac{\gamma^2}{\mu}tr\bar{\boldsymbol{R}}_\omega \approx 18.7456$, which in comparison with the above empirical value shows that the approximate formula (4.5) is quite accurate.

Example 2. We next construct another example to show that if our cooperative information condition (Condition 3.2) is not satisfied, then the sensor network cannot fulfill the estimation or tracking task. Here we assume that $A_i, B_i (i = 1,\dots,20), C_j (j = 1,\dots,5,13,\dots,20)$ remain the same and

$$C_6 = C_{11} = C_1, C_7 = C_{12} = C_2, C_8 = C_3, C_9 = C_4, C_{10} = C_5.$$

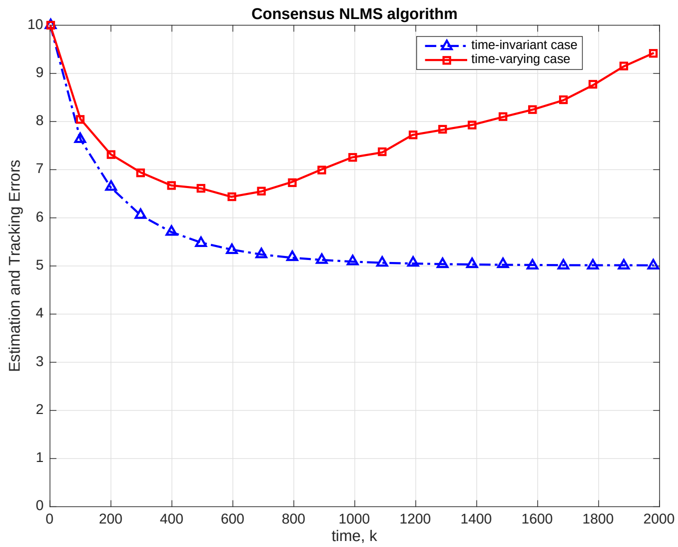

FIG. 6.3. *Estimation and tracking errors in Example* 2.

It is easy to verify that Condition 3.2 is not satisfied. We select the same initial states as in the first example and apply the consensus NLMS-based adaptive filtering algorithm for all sensors; then plot the estimation and tracking errors of the sensor network with $\gamma = 0$ and $\gamma = 0.1$ in Figure 6.3.

The lower line in Figure 6.3 is the time-invariant case and the other line is the time-varying case. The estimsation and tracking errors by using the consensus NLMS algorithm all stay large because the sensor network does not satisfy Condition 3.2.

7. Concluding remarks. In this paper, we have investigated a basic class of NLMS-based distributed adaptive filtering algorithms and proven that such algorithms can fulfill the estimation task even if any individual sensor or subsystem cannot, under a general cooperative information condition that does not exclude applications to feedback systems. This made it possible for further investigation on related problems concerning the combination of control and communication. Moreover, for the commonly used ϕ-mixing processes, we have proven that our general information condition is also necessary for the stability of the consensus NLMS-based filters. In addition, the mean square tracking error matrix has also been assessed under different noise conditions, providing either upper bounds or approximated expressions which can be easily evaluated, analyzed, and even optimized. We remark that the standard LMS-based distributed adaptive filter and the diffusion adaptive filter may also be investigated but different analysis methods need to be used (see, e.g., [42, 43, 45]). Of course, there are still a number of interesting problems for further investigation, for example, to consider other kinds of distributed filtering algorithms, to incorporate with more complicated network topologies, and to combine distributed filtering with distributed control problems, etc.

Appendix A. Proof of Lemma 5.4. According to Condition 3.1, $\mathcal{L}$ 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, $\mathcal{L} \otimes I_m$ 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 $l_{m+1} \leq \cdots \leq l_{mn}$ 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 [10]

$$\boldsymbol{H}_k^i \triangleq \sum_{j=k+1}^{k+h} A_j^i, \quad \boldsymbol{H}_k \triangleq \mathrm{diag}\{\boldsymbol{H}_k^1, \ldots, \boldsymbol{H}_k^n\};$$

then we denote

$$\boldsymbol{\Delta}_k \triangleq \sum_{j=k+1}^{k+h} [\boldsymbol{A}_j + \nu(\mathcal{L} \otimes I_m)] = \boldsymbol{H}_k + h\nu(\mathcal{L} \otimes I_m) \tag{A.1}$$

and

$$\boldsymbol{\Gamma}_k \triangleq \sum_{i=1}^{n} \sum_{j=k+1}^{k+h} A_j^i = \sum_{i=1}^{n} \boldsymbol{H}_k^i. \tag{A.2}$$

By (A.1), let us consider the following form:

$$\begin{aligned}\boldsymbol{\eta}^T \boldsymbol{\Delta}_k \boldsymbol{\eta} &= (\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2)^T \boldsymbol{\Delta}_k (\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2) \\ &= \boldsymbol{\eta}_1^T \boldsymbol{H}_k \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2^T \boldsymbol{H}_k \boldsymbol{\eta}_2 + 2\boldsymbol{\eta}_1^T \boldsymbol{H}_k \boldsymbol{\eta}_2 \\ &\quad + \boldsymbol{\eta}_1^T [h\nu(\mathcal{L} \otimes I_m)]\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2^T [h\nu(\mathcal{L} \otimes I_m)]\boldsymbol{\eta}_2 + 2\boldsymbol{\eta}_1^T [h\nu(\mathcal{L} \otimes I_m)]\boldsymbol{\eta}_2 \\ &\triangleq s_1 + s_2 + s_3 + s_4 + s_5 + s_6.\end{aligned} \tag{A.3}$$

Since $\boldsymbol{H}_k \geq 0$, we can decompose it as $\boldsymbol{H}_k = \boldsymbol{H}_k^{\frac{1}{2}} \boldsymbol{H}_k^{\frac{1}{2}}$. Note that for any two column vectors $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ with the same dimensions, we have following inequality:

$$2\boldsymbol{M}_1^T \boldsymbol{M}_2 \leq \delta \boldsymbol{M}_1^T \boldsymbol{M}_1 + \frac{1}{\delta} \boldsymbol{M}_2^T \boldsymbol{M}_2, \tag{A.4}$$

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

$$\boldsymbol{M}_1 \triangleq -\boldsymbol{H}_k^{\frac{1}{2}} \boldsymbol{\eta}_1, \quad \boldsymbol{M}_2 \triangleq \boldsymbol{H}_k^{\frac{1}{2}} \boldsymbol{\eta}_2,$$

and substitute this into (A.4); it is easy to see

$$\begin{aligned}2\boldsymbol{M}_1^T \boldsymbol{M}_2 = &-2\boldsymbol{\eta}_1^T \boldsymbol{H}_k \boldsymbol{\eta}_2 \leq \delta \boldsymbol{M}_1^T \boldsymbol{M}_1 + \frac{1}{\delta} \boldsymbol{M}_2^T \boldsymbol{M}_2 \\ &= \delta \boldsymbol{\eta}_1^T \boldsymbol{H}_k \boldsymbol{\eta}_1 + \frac{1}{\delta} \boldsymbol{\eta}_2^T \boldsymbol{H}_k \boldsymbol{\eta}_2.\end{aligned} \tag{A.5}$$

Then we can obtain

$$\text{(A.6)}\qquad \begin{aligned} s_3 =& 2\boldsymbol{\eta}_1^T \boldsymbol{H}_k \boldsymbol{\eta}_2 \geq -\delta \boldsymbol{\eta}_1^T \boldsymbol{H}_k \boldsymbol{\eta}_1 - \frac{1}{\delta}\boldsymbol{\eta}_2^T \boldsymbol{H}_k \boldsymbol{\eta}_2 \\ =& -\ \delta s_1 - \frac{1}{\delta} s_2. \end{aligned}$$

From (A.3) and (A.6), it is obvious that

$$\text{(A.7)}\qquad \boldsymbol{\eta}^T \boldsymbol{\Delta}_k \boldsymbol{\eta} \geq (1-\delta)s_1 + (1-\frac{1}{\delta})s_2 + s_4 + s_5 + s_6.$$

Now, we proceed to estimate s_i one by one.

Note that

$$\text{(A.8)}\qquad \begin{aligned} s_1 = \boldsymbol{\eta}_1^T \boldsymbol{H}_k \boldsymbol{\eta}_1 &= \left(\sum_{j=1}^m x_j \boldsymbol{\xi}_j\right)^T \boldsymbol{H}_k \left(\sum_{j=1}^m x_j \boldsymbol{\xi}_j\right) \\ &= \boldsymbol{X}^T \boldsymbol{\Xi}^T \boldsymbol{H}_k \boldsymbol{\Xi} \boldsymbol{X}, \end{aligned}$$

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

Since

$$\boldsymbol{\Xi} = \frac{1}{\sqrt{n}} \begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_m \\ \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_m \end{pmatrix}_{mn\times m},$$

it is easy to show that

$$\begin{aligned} \boldsymbol{H}_k \boldsymbol{\Xi} =& \frac{1}{\sqrt{n}} \begin{pmatrix} \boldsymbol{H}_k^1 \boldsymbol{e}_1 & \boldsymbol{H}_k^1 \boldsymbol{e}_2 & \cdots & \boldsymbol{H}_k^1 \boldsymbol{e}_m \\ \boldsymbol{H}_k^2 \boldsymbol{e}_1 & \boldsymbol{H}_k^2 \boldsymbol{e}_2 & \cdots & \boldsymbol{H}_k^2 \boldsymbol{e}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{H}_k^n \boldsymbol{e}_1 & \boldsymbol{H}_k^n \boldsymbol{e}_2 & \cdots & \boldsymbol{H}_k^n \boldsymbol{e}_m \end{pmatrix} \\ =& \frac{1}{\sqrt{n}} \begin{pmatrix} \boldsymbol{H}_k^1 & \boldsymbol{H}_k^2 & \cdots & \boldsymbol{H}_k^n \end{pmatrix}^T. \end{aligned}$$

Similarly, we can obtain

$$\boldsymbol{\Xi}^T \boldsymbol{H}_k \boldsymbol{\Xi} = \frac{1}{n} \begin{pmatrix} \boldsymbol{e}_1^T \boldsymbol{H}_k^1 + \boldsymbol{e}_1^T \boldsymbol{H}_k^2 + \cdots + \boldsymbol{e}_1^T \boldsymbol{H}_k^n \\ \boldsymbol{e}_2^T \boldsymbol{H}_k^1 + \boldsymbol{e}_2^T \boldsymbol{H}_k^2 + \cdots + \boldsymbol{e}_2^T \boldsymbol{H}_k^n \\ \vdots \\ \boldsymbol{e}_m^T \boldsymbol{H}_k^1 + \boldsymbol{e}_m^T \boldsymbol{H}_k^2 + \cdots + \boldsymbol{e}_m^T \boldsymbol{H}_k^n \end{pmatrix} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{H}_k^i.$$

From this, we have

$$\text{(A.9)}\qquad \boldsymbol{\Xi}^T \boldsymbol{H}_k \boldsymbol{\Xi} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{H}_k^i = \frac{1}{n}\boldsymbol{\Gamma}_k.$$

Substitute (A.9) into (A.8); it can be deduced that

$$\text{(A.10)}\qquad s_1 = \frac{1}{n}\boldsymbol{X}^T \boldsymbol{\Gamma}_k \boldsymbol{X} \geq \frac{\lambda_{min}(\boldsymbol{\Gamma}_k)}{n}\sum_{j=1}^m x_j^2.$$

Notice that

$$|s_2| \leq hc\|\boldsymbol{\eta}_2\|^2 = hc \sum_{j=m+1}^{mn} x_j^2. \tag{A.11}$$

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

$$s_4 = s_6 = 0. \tag{A.12}$$

For s_5, we know that

$$s_5 = h\nu \sum_{j=m+1}^{mn} l_j x_j^2 \geq h\nu l_{m+1} \sum_{j=m+1}^{mn} x_j^2. \tag{A.13}$$

Denote

$$y \triangleq \sum_{j=1}^{m} x_j^2.$$

Denote $\rho_k \triangleq \lambda_{min}(\boldsymbol{\Delta}_k)$ and $\lambda_k \triangleq \lambda_{min}(\boldsymbol{\Gamma}_k)$, and here we choose $\delta \in (0,1)$; we have by (A.3)

$$\begin{aligned} \rho_k \geq & \frac{1-\delta}{n}\lambda_k y + (1-\frac{1}{\delta})hc(1-y) + h\nu l_{m+1}(1-y) \\ = & \left[\frac{1-\delta}{n}\lambda_k - (h\nu l_{m+1} + hc - \frac{hc}{\delta})\right]y + (h\nu l_{m+1} + hc - \frac{hc}{\delta}), \quad y \in [0,1]. \end{aligned} \tag{A.14}$$

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

$$\rho_k \geq \left[\frac{0.5 l_{m+1} h\nu}{nhc + 0.5 l_{m+1}\nu nh}\lambda_k - 0.5 l_{m+1} h\nu\right] y + 0.5 l_{m+1} h\nu, \quad y \in [0,1]. \tag{A.15}$$

It is easy to obtain

$$\rho_k \geq \frac{0.5 l_{m+1} h\nu}{nhc + 0.5 l_{m+1}\nu nh}\lambda_k = \frac{l_{m+1}\nu}{2nc + l_{m+1}\nu n}\lambda_k = \sigma\lambda_k,$$

where $0 < \sigma = \frac{l_{m+1}\nu}{2nc + l_{m+1}\nu n}$. This completes the proof of Lemma 5.4.

Appendix B. Proof of Lemma 5.10. We should show that there exists a positive constant C such that

$$\sum_{i=1}^{n} \sum_{j=k+1}^{k+h} A_j^i \geq C I_m$$

$\forall k \geq 0$. We first prove that for any $k \geq 0$, the eigenvalues of matrices $\sum_{i=1}^{n}\sum_{j=k+1}^{k+h} A_j^i$ 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 $\sum_{i=1}^{n}\sum_{j=k^*+1}^{k^*+h} A_j^i$ is 0. Denote the unit eigenvector of 0 is β_{k^*}; then we have

$$\boldsymbol{\beta}_{k^*}^T \left(\sum_{i=1}^{n} \sum_{j=k^*+1}^{k^*+h} A_j^i\right) \boldsymbol{\beta}_{k^*} = 0. \tag{B.1}$$

Similar to the proof of Lemma 5.4, we know that $\mathcal{L}$ has only one zero eigenvalue whose unit eigenvector is $\frac{1}{\sqrt{n}}(1,\dots,1)^T$, i.e., $\frac{1}{\sqrt{n}}\mathbf{1}$, where $\mathbf{1}=(1,\dots,1)^T_{n\times 1}$. Correspondingly, $\mathcal{L}\otimes I_m$ has m zero eigenvalues whose orthogonal unit eigenvectors are

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

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 $l_{m+1},\dots,l_{mn}$ whose orthogonal unit eigenvectors are denoted as $\boldsymbol{\xi}_{m+1},\dots,\boldsymbol{\xi}_{mn}$, correspondingly. 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

$$\boldsymbol{H}^i_{k^*}=\sum_{j=k^*+1}^{k^*+h}A^i_j,\quad \boldsymbol{H}_{k^*}=\mathrm{diag}\{\boldsymbol{H}^1_{k^*},\dots,\boldsymbol{H}^n_{k^*}\}.$$

Then we denote

$$\boldsymbol{\Delta}_{k^*}=\boldsymbol{H}_{k^*}+h\nu(\mathcal{L}\otimes I_m). \tag{B.2}$$

Note also that

$$\boldsymbol{\Gamma}_{k^*}\triangleq\sum_{i=1}^{n}\boldsymbol{H}^i_{k^*}. \tag{B.3}$$

By (B.2), let us consider the following form:

$$\begin{aligned}\boldsymbol{\eta}^T\boldsymbol{\Delta}_{k^*}\boldsymbol{\eta}=&(\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)^T\boldsymbol{\Delta}_{k^*}(\boldsymbol{\eta}_1+\boldsymbol{\eta}_2)\\ =&\boldsymbol{\eta}_1^T\boldsymbol{H}_{k^*}\boldsymbol{\eta}_1+\boldsymbol{\eta}_2^T\boldsymbol{H}_{k^*}\boldsymbol{\eta}_2+2\boldsymbol{\eta}_1^T\boldsymbol{H}_{k^*}\boldsymbol{\eta}_2\\ &+\boldsymbol{\eta}_1^T[h\nu(\mathcal{L}\otimes I_m)]\boldsymbol{\eta}_1+\boldsymbol{\eta}_2^T[h\nu(\mathcal{L}\otimes I_m)]\boldsymbol{\eta}_2\\ &+2\boldsymbol{\eta}_1^T[h\nu(\mathcal{L}\otimes I_m)]\boldsymbol{\eta}_2\\ \triangleq&s_1^{k^*}+s_2^{k^*}+s_3^{k^*}+s_4+s_5+s_6.\end{aligned} \tag{B.4}$$

For two column vectors $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ with the same dimensions, we have the following inequality:

$$2\boldsymbol{M}_1^T\boldsymbol{M}_2\le\delta\boldsymbol{M}_1^T\boldsymbol{M}_1+\frac{1}{\delta}\boldsymbol{M}_2^T\boldsymbol{M}_2, \tag{B.5}$$

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

$$\boldsymbol{M}_1\triangleq\boldsymbol{H}_{k^*}^{1/2}\boldsymbol{\eta}_1,\quad \boldsymbol{M}_2\triangleq\boldsymbol{H}_{k^*}^{1/2}\boldsymbol{\eta}_2,$$

and substitute this into (B.5); it is easy to have

$$\begin{aligned}2\boldsymbol{M}_1^T\boldsymbol{M}_2&=2\boldsymbol{\eta}_1^T\boldsymbol{H}_{k_s}\boldsymbol{\eta}_2\\ &\le\delta\boldsymbol{M}_1^T\boldsymbol{M}_1+\frac{1}{\delta}\boldsymbol{M}_2^T\boldsymbol{M}_2\\ &=\delta\boldsymbol{\eta}_1^T\boldsymbol{H}_{k^*}\boldsymbol{\eta}_1+\frac{1}{\delta}\boldsymbol{\eta}_1^T\boldsymbol{H}_{k^*}\boldsymbol{\eta}_1.\end{aligned}$$

Then we can obtain

$$s_3^{k^*} = 2\boldsymbol{\eta}_1^T \boldsymbol{H}_{k^*} \boldsymbol{\eta}_2 \le \delta \boldsymbol{\eta}_1^T \boldsymbol{H}_{k^*} \boldsymbol{\eta}_1 + \frac{1}{\delta} \boldsymbol{\eta}_2^T \boldsymbol{H}_{k^*} \boldsymbol{\eta}_2 = \delta s_1^{k^*} + \frac{1}{\delta} s_2^{k^*}. \quad \text{(B.6)}$$

From (B.4) and (B.6), it is obvious that

$$\boldsymbol{\eta}^T \boldsymbol{\Delta}_{k^*} \boldsymbol{\eta} \le (1+\delta) s_1^{k^*} + (1+\frac{1}{\delta}) s_2^{k^*} + s_4 + s_5 + s_6. \quad \text{(B.7)}$$

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

$$\begin{aligned} s_1^{k^*} &= \boldsymbol{\eta}_1^T \boldsymbol{H}_{k^*} \boldsymbol{\eta}_1 = \Big(\sum_{j=1}^m x_j \boldsymbol{\xi}_j\Big)^T \boldsymbol{H}_{k^*} \Big(\sum_{j=1}^m x_j \boldsymbol{\xi}_j\Big) \\ &= \boldsymbol{X}^T \boldsymbol{\Xi}^T \boldsymbol{H}_{k^*} \boldsymbol{\Xi} \boldsymbol{X} \\ &= \frac{1}{n} \boldsymbol{X}^T \Big(\sum_{i=1}^n \boldsymbol{H}_{k^*}^i\Big) \boldsymbol{X} \\ &= \frac{1}{n} \cdot \boldsymbol{X}^T \boldsymbol{\Gamma}_{k^*} \boldsymbol{X}, \end{aligned} \quad \text{(B.8)}$$

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

Notice that

$$|s_2^{k^*}| \le hc \|\boldsymbol{\eta}_2\|^2 = hc \sum_{j=m+1}^{mn} x_j^2. \quad \text{(B.9)}$$

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

$$s_4 = s_6 = 0. \quad \text{(B.10)}$$

For s_5, we know that

$$s_5 = h\nu \sum_{j=m+1}^{mn} l_j x_j^2 \le h\nu l_{mn} \sum_{j=m+1}^{mn} x_j^2, \quad \text{(B.11)}$$

where l_{mn} is the largest eigenvalue of matrix $\mathcal{L} \otimes I_m$.

Denote

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

By (B.7)(B.8)(B.9)(B.10)(B.11) and denoting $\rho_{k^*} = \lambda_{min}(\Delta_{k^*})$, we know that for any $y \in [0,1]$,

$$\rho_{k^*} \le \frac{1+\delta}{n} \cdot \boldsymbol{X}^T \boldsymbol{\Gamma}_{k^*} \boldsymbol{X} + (1+\frac{1}{\delta}) hc(1-y) + h\nu l_{mn}(1-y). \quad \text{(B.12)}$$

We can take $\boldsymbol{X} = \boldsymbol{\beta}_{k^*}$; then we have

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

which contradicts with $\rho_{k^*} > 0$. Then for any $k \geq 0$, the eigenvalues of matrices $\sum_{i=1}^{n}\sum_{j=k+1}^{k+h} A_j^i$ are all positive.

Secondly, we prove that all of the eigenvalues of the above matrix must have a uniform lower bound $C > 0$ with respect to $k \geq 0$. This can be done through contradiction by assuming that there exists an eigenvalue σ_k and a sequence $\{k_s\}_{s=1}^{\infty}$ such that $\lim_{s\to\infty}\sigma_{k_s} = 0$. Denote the unit eigenvector of σ_{k_s} as $\boldsymbol{\beta}_{k_s} \in \mathbb{R}^m$. Then we have

$$\lim_{s\to\infty} \boldsymbol{\beta}_{k_s}^T \left\{ \sum_{i=1}^{n} \sum_{j=k_s+1}^{k_s+h} A_j^i \right\} \boldsymbol{\beta}_{k_s} = \lim_{s\to\infty} \boldsymbol{\beta}_{k_s}^T \sigma_{k_s} \boldsymbol{\beta}_{k_s} = 0. \tag{B.13}$$

Similar to the above proof, we can take $\boldsymbol{X} = \boldsymbol{\beta}_{k_s}$; then we have

$$\rho_{k_s} \leq (1+\delta)\frac{\mu}{n} \cdot \boldsymbol{X}^T \boldsymbol{\Gamma}_{k_s} \boldsymbol{X}.$$

It is obvious that

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

which contradicts with (5.12). Therefore, we conclude that there exists a positive constant C such that

$$\sum_{i=1}^{n} \sum_{j=k+1}^{k+h} A_j^i \geq C I_m$$

$\forall k \geq 0$. This completes the proof.

REFERENCES

[1] Y. Bar-Shalom, *Multitarget-Multisensor Tracking: Advanced Applications*, Vol. 1, Artech House, Boston, 1990.

[2] R. L. Berger, *A necessary and sufficient condition for reaching a consensus using Degroot's method*, J. Amer. Statist. Assoc., 76 (1981), pp. 415–418.

[3] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, *Distributed Kalman filtering based on consensus strategies*, IEEE J. Sel. Areas Commun., 26 (2008), pp. 622–633.

[4] F. S. Cattivelli and A. H. Sayed, *Diffusion LMS strategies for distributed estimation*, IEEE Trans. Signal Process., 58 (2010), pp. 1035–1048.

[5] F. S. Cattivelli and A. H. Sayed, *Diffusion strategies for distributed Kalman filtering and smoothing*, IEEE Trans. Automat. Control, 55 (2010), pp. 2069–2084.

[6] F. S. Cattivelli and A. H. Sayed, *Analysis of spatial and incremental lms processing for distributed estimation*, IEEE Trans. Signal Process., 59 (2011), pp. 1465–1480.

[7] C. Chen, Z. Liu, and L. Guo, *Performance analysis of distributed adaptive filters*, Commun. Inf. Syst., 15 (2015), pp. 453–476.

[8] C. Chen, Z. Liu, and L. Guo, *Performance bounds of distributed adaptive filters with cooperative correlated signals*, Sci. China Inf. Sci., 59 (2016), 112202.

[9] C. Chen, Z. Liu, and L. Guo, *Stability of diffusion adaptive filters*, in Proceedings of the 19th World Congress of the International Federation of Automatic Control (IFAC), Cape Town, South Africa, August 24-29, 2014, pp. 10409–10414.

[10] W. Chen, C. Wen, S. Hua, and C. Sun, *Distributed cooperative adaptive identification and control for a group of continuous-time systems with a cooperative PE condition via consensus*, IEEE Trans. Automat. Control, 59 (2014), pp. 91–106.

[11] Y. S. Chow and H. Teicher, *Probability Theory*, Springer, New York, 1978.

[12] M. H. DeGroot, *Reaching a consensus*, J. Amer. Statist. Assoc., 69 (1974), pp. 118–121.

[13] C. Godsil and G. Royle, *Algebraic Graph Theory*, Springer-Verlag, New York, 2001.

[14] L. Guo, *Stability of recursive stochastic tracking algorithms*, SIAM J. Control Optim., 32 (1994), pp. 1195–1225.

[15] L. Guo and L. Ljung, *Exponential stability of general tracking algorithms*, IEEE Trans. Automat. Control, 40 (1995), pp. 1376–1387.
[16] L. Guo and L. Ljung, *Performance analysis of general tracking algorithms*, IEEE Trans. Automat. Control, 40 (1995), pp. 1388–1402.
[17] L. Guo, L. Ljung, and G.-J. Wang, *Necessary and sufficient conditions for stability of LMS*, IEEE Trans. Automat. Control, 42 (1997), pp. 761–770.
[18] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ 1996.
[19] S. Kar, J. M. Moura, and H. V. Poor, *Distributed linear parameter estimation: Asymptotically efficient adaptive strategies*, SIAM J. Control Optim., 51 (2013), pp. 2200–2229.
[20] S. Kar, J. M. Moura, and K. Ramanan, *Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication*, IEEE Trans. Inform. Theory, 58 (2012), pp. 3575–3605.
[21] H. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer-Verlag, New York, 2003.
[22] A. J. Laub, *Matrix Analysis for Scientists and Engineers*, SIAM, Philadelphia, 2005.
[23] L. Ljung and S. Gunnarsson, *Adaptation and tracking in system identification - a survey*, Automatica, 26 (1990), pp. 7–21.
[24] C. G. Lopes and A. H. Sayed, *Diffusion least-mean squares over adaptive networks: Formulation and performance analysis*, IEEE Trans. Signal Process., 56 (2008), pp. 3122–3136.
[25] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission*, Wiley, New York, 1995.
[26] G. Mateos and G. B. Giannakis, *Distributed recursive least-squares: Stability and performance analysis*, IEEE Trans. Signal Process., 60 (2012), pp. 3740–3754.
[27] G. Mateos, I. D. Schizas, and G. B. Giannakis, *Distributed recursive least-squares for consensus-based in-network adaptive estimation*, IEEE Trans. Signal Process., 57 (2009), pp. 4583–4588.
[28] 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., 63 (2015), pp. 4300–4314.
[29] R. Olfati-Saber, *Distributed Kalman filtering for sensor networks*, in Decision and Control, IEEE, New York, 2007, pp. 5492–5498.
[30] M. J. Piggott and V. Solo, *Diffusion LMS with correlated regressors* I*: Realization-wise stability*, IEEE Trans. Signal Process., 64 (2016), pp. 5473–5484.
[31] M. J. Piggott and V. Solo, *Stability of distributed adaptive algorithms* II*: Diffusion algorithms*, in Proceedings of the 54th IEEE Conference on Decision and Control (CDC), Osaka, Japan, December 15-18, 2015, pp. 7428–7433.
[32] A. Rastegarnia, M. A. Tinati, and A. Khalili, *A distributed incremental LMS algorithm with reliability of observation consideration*, in Proceedings of the IEEE International Conference on Communication Systems (ICCS), 2010, pp. 67–70.
[33] A. H. Sayed, *Adaptive networks*, Proc. IEEE, 102 (2014), pp. 460–497.
[34] A. H. Sayed, *Diffusion adaptation over networks*, in Academic Press Library in Signal Processing 3, R. Chellapa and S. Theodoridis, eds., 2014, pp. 323–454.
[35] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. J. Towfic, *Diffusion strategies for adaptation and learning over networks: An examination of distributed strategies and network behavior*, IEEE Signal Process. Mag., 30 (2013), pp. 155–171.
[36] I. D. Schizas, G. Mateos, and G. B. Giannakis, *Distributed LMS for consensus-based in-network adaptive processing*, IEEE Trans. Signal Process., 57 (2009), pp. 2365–2382.
[37] V. Solo, *Stability of distributed adaptive algorithms* I*: Consensus algorithms*, in Proceedings of the 54th IEEE Conference on Decision and Control (CDC), Osaka, Japan, December 15-18, 2015, pp. 7422–7427.
[38] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*, Prentice-Hall, Upper Saddle River, NJ, 1995.
[39] M. A. Tinati, A. Rastegarnia, and A. Khalili, *An incremental least-mean square algorithm with adaptive combiner*, in Proceedings of the IET 3rd International Conference on Wireless, Mobile and Multimedia Networks, IET, 2010, pp. 266–269.
[40] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
[41] L. Xiao, S. Boyd, and S. Lall, *A scheme for robust distributed sensor fusion based on average consensus*, in Proceedings of the 4th International Symposium on Information Processing in Sensor Networks, Los Angeles, CA, 2005, pp. 63–70.
[42] S. Xie and L. Guo, *Analysis of distributed adaptive filters based on diffusion strategies over sensor networks*, IEEE Trans. Automat. Control, preprint, 2018.

[43] S. Xie and L. Guo, *A necessary and sufficient condition for stability of LMS-based consensus adaptive filters*, Automatica, 93 (2018), pp. 12–19.
[44] S. Xie and L. Guo, *Stability of distributed LMS under cooperative stochastic excitation*, in Proceedings of the 34th Chinese Control Conference (CCC), Hangzhou, China, July 28-30, 2015, pp. 7445–7450.
[45] S. Xie and L. Guo, *Exponential stability of LMS-based distributed adaptive filters*, in Proceedings of the 20th World Congress of the International Federation of Automatic Control (IFAC), Toulouse, France, July 9-14, 2017, pp. 13764–13769.
[46] G. G. Yin and V. Krishnamurthy, *LMS algorithms for tracking slow Markov chains with applications to hidden Markov estimation and adaptive multiuser detection*, IEEE Trans. Inform. Theory, 51 (2005), pp. 2475–2490.
[47] Y. Zhu, Z. You, J. Zhao, K. Zhang, and X. R. Li, *The optimality for the distributed Kalman filtering fusion with feedback*, Automatica, 37 (2001), pp. 1489–1493.