Two-Layer Diffusion Adaptive Filters Over Directed Markovian Switching Networks

Siyu Xie, Die Gan, *Member, IEEE*, and Zhixin Liu, *Member, IEEE*

***Abstract*—We consider the problem of distributed adaptive filtering in this letter, where a set of nodes in the network is required to estimate an unknown parameter of interest from noisy measurements cooperatively. Based on normalized least mean squares (NLMS) adaptive filters, we focus on a two-layer diffusion strategy to diffuse data more thoroughly. We propose and analyze the two-layer diffusion NLMS algorithm, where the communications between nodes in the network are described by directed Markovian switching graphs. The directed graphs are not only used for the combination of local estimates but also used for the adaptation step in the two-layer strategy. The stability results of the proposed two-layer diffusion adaptive filters are established under a general cooperative information assumption, without independence and stationarity signal conditions which were widely used in the literature. Moreover, the stability result indicates that even if any node cannot estimate the unknown parameter individually, the whole network can still fulfill the estimation task through communications. Simulation results also show that the proposed two-layer diffusion NLMS algorithm has better performance compared with the consensus one.**

***Index Terms*—Distributed adaptive filtering, Markovian switching networks, two-layer diffusion strategy, stability.**

I. Introduction

DISTRIBUTED adaptive estimation problems have attracted increasing research attention recently, where the nodes in the network can be utilized to estimate or track a dynamic process of interest from noisy measurements cooperatively by sharing information with their neighbors through the network topology. Another solution to the estimation problem in the network is called the centralized method, in which all the nodes will send their information to a fusion center for data processing. The centralized method has the disadvantage of being non-robust to failure of the fusion center. Moreover, since at every time instant, all the nodes need to send data to the fusion center, the centralized method may require large amounts of communication and energy resources, and may suffer from serious data packet loss problem if some nodes are far away from the fusion center. In comparison, the distributed solution with in-network processing can save resources and is more robust to node and link failures in the network.

There are two main fully distributed strategies introduced and used in the literature, i.e., consensus [1], [2], [3], [4], [5], [6] and diffusion [7], [8], [9], [10], [11], [12], [13] strategies. Diffusion solutions may have better performance compared with consensus ones since data can be diffused more thoroughly in the network. Here in this letter, we consider one general class of diffusion least mean squares (LMS) algorithms for a stochastic time-varying linear regression model, which is a simplified system model compared with the linear system model considered in, e.g., [14], [15]. However, the observation vector in our model is stochastic, while most literature including [14], [15] considered deterministic observation vectors or matrices. Note that the signals are often stochastic since they are generated from dynamic systems affected by noises, and the linearized observation matrices in the extended Kalman filters are also stochastic, which are widely used to estimate the state of nonlinear engineering systems. There are many existing researches about the analysis of the diffusion LMS algorithms in the literature [8], [9], [10], [11], [12] to consider the stochastic regression vectors. However, almost all the existing theoretical analysis require independence and/or stationarity conditions of the system signals, which are not applicable to the signals generated from feedback systems.

In comparison to our prior work on the diffusion LMS algorithms [12], the current paper studies a more general class of diffusion strategies introduced in [7] of which [8], [9], [10], [11], [12] are some special cases, namely, a two-layer diffusion strategy. We subsequently establish the stability and performance of this two-layer diffusion adaptive filters under non-independent and non-stationary signal conditions, and show that even if any individual sensor cannot fulfill the estimation task, the whole network can still estimate the unknown parameter cooperatively. Moreover, our recent

Manuscript received 13 September 2023; revised 9 November 2023; accepted 15 November 2023. Date of publication 20 November 2023; date of current version 8 December 2023. This work was supported in part by the Major Key Project of Peng Cheng Laboratory under Grant PCL2023AS1-2, and in part by the China Postdoctoral Science Foundation under Grant 2023T160605. Recommended by Senior Editor V. Ugrinovskii. *(Corresponding author: Die Gan.)*

Siyu Xie is with the School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: syxie@uestc.edu.cn).

Die Gan is with Zhongguancun Laboratory, Beijing 100094, China (e-mail: gandie@amss.ac.cn).

Zhixin Liu is with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100045, China, and also with the School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: lzx@amss.ac.cn).

Digital Object Identifier 10.1109/LCSYS.2023.3334548

2475-1456 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 10,2023 at 03:56:46 UTC from IEEE Xplore. Restrictions apply.

work on diffusion LMS algorithms [12] focused on undirected and fixed communication networks. However, due to random failures and recoveries of links between sensors in practical applications, the network topologies may be randomly changing over time, which are usually modeled as an i.i.d. or a Markovian switching system [2], [3], [13]. Since in practice the randomly switching topologies are usually temporally correlated, we have consider directed Markovian switching communication topologies in this letter, which makes our theoretical results more general.

Compared with [6], one of the main contributions is that the random matrix analyzed in this letter is no longer symmetric which increases the difficulty for analysis. Also, the two-layer diffusion strategy cannot only use the estimated values of the neighbors, but also use their observed values to improve the accuracy of estimation. Therefore, the convergence rate is faster since the data may diffuse more thoroughly in the two-layer diffusion strategy which can be seen from the simulation results.

The remainder of this letter is organized as follows. In Section II, we introduce the problem formulation. In Section III, we establish the stability result of the proposed two-layer diffusion algorithms under some mild signal assumption. Finally, we provide some simulation results in Section IV, and conclude this letter with some remarks in Section V.

II. Problem Formulation

A. Graph Theory

We consider a network with n nodes and the communications among sensors are modeled as a weighted digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}_{\mathcal{G}})$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ is the set of vertexes, and $\mathcal{E}_{\mathcal{G}} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of directed arrows. An arrow (i, j) is considered to be directed from i to j, where j is called the head and i is called the tail of the arrow. In the estimation process, node i can obtain local estimate and local observation information from its in-neighbors through graph $\mathcal{G}$. The associated weighted adjacency matrix $\mathcal{A}_{\mathcal{G}} = \{a^{ij}_{\mathcal{G}}\}_{n\times n}$ is used to describe the structures of the corresponding digraph $\mathcal{G}$, where $a^{ij}_{\mathcal{G}} > 0$ if the arrow $(i, j) \in \mathcal{E}_{\mathcal{G}}$, and $a^{ij}_{\mathcal{G}} = 0$ otherwise. In this letter, we assume that the elements of the weighted matrix $\mathcal{A}_{\mathcal{G}}$ satisfy $\sum_{j=1}^{n} a^{ji}_{\mathcal{G}} = 1, \forall i = 1, \ldots, n$. Note that matrix $\mathcal{A}_{\mathcal{G}}$ maybe asymmetric. Also, the matrix $\mathcal{G}$ is called a balanced digraph, if $\sum_{j=1}^{n} a^{ij}_{\mathcal{G}} = \sum_{j=1}^{n} a^{ji}_{\mathcal{G}} = 1, \forall i = 1, \ldots, n$.

The set of in-neighbors of node i in graph $\mathcal{G}$ is denoted as $\mathcal{N}^{i}_{\mathcal{G}} = \{j \in \mathcal{V} | (j, i) \in \mathcal{E}_{\mathcal{G}}\}$, and any in-neighboring nodes of i have the ability to transmit information over the directed arrows to node i. The Laplacian matrix of the directed graph $\mathcal{G}$ is defined as $\mathcal{L}_{\mathcal{G}} = I_n - \mathcal{A}_{\mathcal{G}}$, where I_n denotes the n-dimensional identity matrix. From [18], we know that the smallest eigenvalue of the matrix $\mathcal{L}_{\mathcal{G}}$ always equals zero, with $\frac{1}{\sqrt{n}}\mathbf{1}_n^{\top}$ being the corresponding normalized eigenvector. A sequence of edges $(i, i_1), (i_1, i_2), \ldots, (i_{k-1}, j)$ is called a path from i to j. The graph $\mathcal{G}$ is called strongly connected if for any $i, j \in \mathcal{V}$, there is a path from i to j. In addition, if the directed graph contains a spanning tree, the matrix $\mathcal{L}_{\mathcal{G}}$ has only one zero eigenvalue with other eigenvalues have positive real parts.

For a given integer $k > 0$, by the union of a collection of digraphs $\{\mathcal{G}_j = (\mathcal{V}, \mathcal{E}_{\mathcal{G}_j}, \mathcal{A}_{\mathcal{G}_j}), j = 1, \ldots, k\}$, means the digraph $\sum_{j=1}^{k} \mathcal{G}_j$ with the vertex set $\mathcal{V}$, the edge set $\cup_{j=1}^{k} \mathcal{E}_{\mathcal{G}_j}$ equalling to the union of the edge sets of all the digraphs in the collection, and the corresponding adjacency matrix becomes $\frac{1}{k}\sum_{j=1}^{k} \mathcal{A}_{\mathcal{G}_j}$.

B. Observation Model

Assume that at each time instant k, each node i ($i \in \{1, \ldots, n\}$) in the network receives a noisy scalar measurement y^i_k and an regressor $\boldsymbol{\varphi}^i_k \in \mathbb{R}^m$. They are related by a stochastic time-varying linear regression model

$$y^i_k = \left(\boldsymbol{\varphi}^i_k\right)^{\top} \boldsymbol{\theta}_k + v^i_k, \quad k \geq 0, \tag{1}$$

where v^i_k is the scalar stochastic noise, and $\boldsymbol{\theta}_k \in \mathbb{R}^m$ is an unknown time-varying signal vector which needs to be estimated. Note that the variation of the unknown parameter $\boldsymbol{\theta}_k$ at time k can be denoted by $\Delta\boldsymbol{\theta}_k$, i.e.,

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

For every node i in the network, the objective is to use the data $\{y^i_k, \boldsymbol{\varphi}^i_k\}$ to track the unknown parameter vector $\boldsymbol{\theta}_k$. Many problems from different application areas, such as target localization, collaborative spectral sensing, signal processing and so on, can be cast as the observation model (1), see [16]. It can also include the well-known ARX model [17] with time-varying coefficients, which is widely used to analyze and predict the trend of time series data in many fields such as control systems, economics and finance systems.

Here in this letter, we assume that the ideal network graph is $\mathcal{G} = (\mathcal{V}, \mathcal{E}_{\mathcal{G}}, \mathcal{A}_{\mathcal{G}})$. Since some edges in the network are failure-prone with positive probability, we will use a Markov chain m^{ji}_k with the state space $\{0, 1\}$ to denote the evolution of the failure-prone edge $(j, i) \in \mathcal{E}_{\mathcal{G}}$, where state "0" means that the communication from j to i is lost, while "1" means that the communication is normal. Let us make a fixed ordering for all the edges in $\mathcal{E}_{\mathcal{G}}$, and we assume that the number of all the failure-prone edges is $n_{\mathcal{G}}$. Then the Markov chain of all the failure-prone edges can be denoted as a vector $m_k \in \mathbb{R}^{n_{\mathcal{G}}}$ with 0 or 1 as its element, which has a finite state space $\mathcal{S}$.

Label the states of $\mathcal{S}$ in sequence as a set $I = \{1, 2, \ldots, s\}$, where $s = 2^{n_{\mathcal{G}}}$, corresponding to the communication topology graph set $\{\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(s)}\}$, where $\mathcal{G}^{(t)} = (\mathcal{V}, \mathcal{E}_{\mathcal{G}^{(t)}}, \mathcal{A}_{\mathcal{G}^{(t)}})$ is a digraph. Without loss of generality, we assume the first state is $[1, \ldots, 1]^{\top}$, which means that $\mathcal{G}^{(1)} = \mathcal{G}$. Denote the Markovian communication graph at time k by $\mathcal{G}_k$, and the random process m_k completely describes dynamic changes of the communication topology.

C. Two-Layer Diffusion Adaptive Filtering Algorithm

In the following, we present the two-layer combine-then-adapt (CTA) diffusion strategy based on NLMS algorithms to estimate $\boldsymbol{\theta}_k$ in a cooperative way (see Algorithm 1), where the directed graphs are not only used for the combination of local estimates but also used for the adaptation step. Thus, the two-layer diffusion strategy can diffuse data more thoroughly compared with the traditional diffusion one [12].

We introduce the following notations for the analysis:

$$\boldsymbol{Y}_k \overset{\triangle}{=} \mathrm{col}\{y^1_k, \ldots, y^n_k\}, \qquad \boldsymbol{\Phi}_k \overset{\triangle}{=} \mathrm{diag}\{\boldsymbol{\varphi}^1_k, \ldots, \boldsymbol{\varphi}^n_k\},$$

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 10,2023 at 03:56:46 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 Two-Layer CTA Diffusion NLMS Algorithm

For any given node i ($i \in \{1\ldots,n\}$), begin with an initial estimate $\widehat{\theta}_0^i$. The algorithm is recursively defined for iteration $k \geq 0$ as follows, and $\mu \in (0,1)$ is the step size:

1. Combine local estimates:

$$\widehat{\beta}_k^i = \sum_{j\in\mathcal{N}_{\mathcal{G}_k}^i} d_{\mathcal{G}_k}^{ji} \widehat{\theta}_k^j,$$

2. Adapt the local estimates:

$$\widehat{\theta}_{k+1}^i = \widehat{\beta}_k^i + \mu \sum_{j\in\mathcal{N}_{\mathcal{G}_k}^i} d_{\mathcal{G}_k}^{ji} \frac{\varphi_k^j}{1+\|\varphi_k^j\|^2}[y_k^j - (\varphi_k^j)^\top \widehat{\beta}_k^j].$$

$$\begin{aligned}
&\boldsymbol{L}_k \triangleq \text{diag}\left\{\frac{\varphi_k^1}{1+\|\varphi_k^1\|^2},\ldots,\frac{\varphi_k^n}{1+\|\varphi_k^n\|^2}\right\},\\
&\boldsymbol{F}_k \triangleq \boldsymbol{L}_k\boldsymbol{\Phi}_k^\top, \quad \boldsymbol{V}_k \triangleq \text{col}\{v_k^1,\ldots,v_k^n\},\\
&\Delta\boldsymbol{\Theta}_k \triangleq \text{col}\{\underbrace{\Delta\boldsymbol{\theta}_k,\ldots,\Delta\boldsymbol{\theta}_k}_{n}\}, \quad \boldsymbol{\Theta}_k \triangleq \text{col}\{\underbrace{\boldsymbol{\theta}_k,\ldots,\boldsymbol{\theta}_k}_{n}\},\\
&\widehat{\boldsymbol{\Theta}}_k \triangleq \text{col}\{\widehat{\boldsymbol{\theta}}_k^1,\ldots,\widehat{\boldsymbol{\theta}}_k^n\},\ \widehat{\boldsymbol{\mathcal{B}}}_k \triangleq \text{col}\{\widehat{\boldsymbol{\beta}}_k^1,\ldots,\widehat{\boldsymbol{\beta}}_k^n\},\\
&\widetilde{\boldsymbol{\Theta}}_k \triangleq \text{col}\{\widetilde{\boldsymbol{\theta}}_k^1,\ldots,\widetilde{\boldsymbol{\theta}}_k^n\},\ \text{where } \widetilde{\boldsymbol{\theta}}_k^i = \widehat{\boldsymbol{\theta}}_k^i - \boldsymbol{\theta}_k,\\
&\mathscr{A}_{\mathcal{G}_k} \triangleq \mathcal{A}_{\mathcal{G}_k}\otimes I_m, \quad \mathscr{L}_{\mathcal{G}_k} \triangleq \mathcal{L}_{\mathcal{G}_k}\otimes I_m.
\end{aligned}$$

By (1) and (2), we have the vector form model $\boldsymbol{Y}_k = \boldsymbol{\Phi}_k^\top\boldsymbol{\Theta}_k + \boldsymbol{V}_k$, and $\Delta\boldsymbol{\Theta}_{k+1} = \boldsymbol{\Theta}_{k+1} - \boldsymbol{\Theta}_k$. For the two-layer CTA diffusion NLMS algorithm, we have $\widehat{\boldsymbol{\mathcal{B}}}_k = \mathscr{A}_{\mathcal{G}_k}^\top\widehat{\boldsymbol{\Theta}}_k$ and $\widehat{\boldsymbol{\Theta}}_{k+1} = \widehat{\boldsymbol{\mathcal{B}}}_k + \mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{L}_k(\boldsymbol{Y}_k - \boldsymbol{\Phi}_k^\top\widehat{\boldsymbol{\mathcal{B}}}_k)$. Denote $\widetilde{\boldsymbol{\Theta}}_k = \widehat{\boldsymbol{\Theta}}_k - \boldsymbol{\Theta}_k$, we can get

$$\begin{aligned}
\widetilde{\boldsymbol{\Theta}}_{k+1} &= \mathscr{A}_{\mathcal{G}_k}^\top\widehat{\boldsymbol{\Theta}}_k - \boldsymbol{\Theta}_k - \Delta\boldsymbol{\Theta}_{k+1}\\
&\quad + \mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{L}_k\Big[\boldsymbol{\Phi}_k^\top\boldsymbol{\Theta}_k + \boldsymbol{V}_k - \boldsymbol{\Phi}_k^\top\mathscr{A}_{\mathcal{G}_k}^\top\widehat{\boldsymbol{\Theta}}_k\Big],\\
&= \Big(I_{mn} - \mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k\Big)\mathscr{A}_{\mathcal{G}_k}^\top\widetilde{\boldsymbol{\Theta}}_k + \mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{L}_k\boldsymbol{V}_k - \Delta\boldsymbol{\Theta}_{k+1}\\
&= \Big(I_{mn} - \Big[\mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k + \mathscr{L}_{\mathcal{G}_k}^\top - \mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k\mathscr{L}_{\mathcal{G}_k}^\top\Big]\Big)\widetilde{\boldsymbol{\Theta}}_k\\
&\quad + \mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{L}_k\boldsymbol{V}_k - \Delta\boldsymbol{\Theta}_{k+1},
\end{aligned} \tag{3}$$

which will be analyzed in the following section. The random matrix $\mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k + \mathscr{L}_{\mathcal{G}_k}^\top - \mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k\mathscr{L}_{\mathcal{G}_k}^\top$ involved here may be non-symmetric, non-independent and non-stationary. In order to show stability properties of (3), we divide the problem into two steps: (1) Analyzing the exponential stability of the homogeneous part $I_{mn} - [\mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k + \mathscr{L}_{\mathcal{G}_k}^\top - \mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k\mathscr{L}_{\mathcal{G}_k}^\top]$ of (3); (2) Analyzing the cumulative effects of the noise $\boldsymbol{V}_k$ and the parameter variation $\Delta\boldsymbol{\Theta}_{k+1}$. The details will be provided in *Theorems 1* and *2*.

III. Stability Analysis

A. Definitions

For a matrix $X \in \mathbb{R}^{m\times n}$, $\|X\| = (\lambda_{\max}\{XX^\top\})^{\frac{1}{2}}$ denotes the spectral norm, and $\lambda_{\max}\{\cdot\}$ denotes the largest eigenvalue. Also, for a random matrix Y, $\|Y\|_{L_p} = \{\mathbb{E}[\|Y\|^p]\}^{\frac{1}{p}}$ denotes the L_p-norm, and $\mathbb{E}[\,\cdot\,]$ denotes the mathematical expectation operator. We need the following definitions for the (exponential) stability introduced in [19].

Definition 1: For a sequence of $d\times d$ random matrices $A = \{A_k, k\geq 0\}$, if sequence A belongs to the following set

$$S_p(\lambda) = \Bigg\{A : \Bigg\|\prod_{j=i+1}^{k}(I_d - A_j)\Bigg\|_{L_p} \leq M\lambda^{k-i},\ \forall k\geq i+1, \forall i\geq 0,\ \text{for some } M>0\Bigg\}, \tag{4}$$

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

For convenience of discussions, we introduce the following subclass of $S_1(\lambda)$ for a scalar sequence.

Definition 2: For a scalar sequence $a = \{a_k, k\geq 0\}$, define $S^0(\lambda)$ class as follows

$$S^0(\lambda) = \Bigg\{a :\ a_k\in[0,1], \mathbb{E}\Bigg[\prod_{j=i+1}^{k}(1-a_j)\Bigg] \leq M\lambda^{k-i},\ \forall k\geq i+1, \forall i\geq 0, \text{for some } M>0\Bigg\}, \tag{5}$$

where $\lambda\in[0,1)$.

Note that if the scalar random sequence a is uniformly bounded from below by a positive constant, then obviously it belongs to $S^0(\lambda)$ class.

Definition 3: 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.

B. Assumptions

In the theoretical analysis, we also need the following assumptions on the graph $\mathcal{G}$ and the regressors.

Assumption 1 (Network Topology Assumption):

A1) The all digraphs $\mathcal{G}^{(t)}$, $1\leq t\leq s$, are balanced.

A2) The union of the communication topology set $\mathcal{C} = \{\mathcal{G}^{(t)}, 1\leq t\leq s\}$, which is denoted by $\mathcal{G}_{\mathcal{C}}$, contains a spanning tree.

A3) $\{m_k, k\geq 0\}$ is a homogeneous ergodic Markov chain with the transition probability matrix $[p_{ij}]_{1\leq,i,j\leq s}$, where $p_{ij} = P\{m_{k+1} = j|m_k = i, m_0,\ldots,m_{k-1}, \varphi_t^l, \Delta\boldsymbol{\theta}_t, v_t^l, l = 1,\ldots,n,\ t\leq k\}$.

A4) There exists a constant $c\in(0,1)$ such that $\mathcal{L}_{\mathcal{G}^{(t)}}\mathcal{L}_{\mathcal{G}^{(t)}}^\top \leq (1-c)(\mathcal{L}_{\mathcal{G}^{(t)}} + \mathcal{L}_{\mathcal{G}^{(t)}}^\top)$, *a.s.*, $1\leq t\leq s$.

Remark 1: Since the digraph $\mathcal{G}^{(t)}$ is balanced, then the adjacency matrix $\mathcal{A}_{\mathcal{G}^{(t)}}$ is doubly stochastic. Hence, $\frac{1}{2}(\mathcal{A}_{\mathcal{G}^{(t)}} + \mathcal{A}_{\mathcal{G}^{(t)}}^\top)$ is also symmetric and doubly stochastic, which means that the corresponding mirror graph is undirected. Then, we know that all the eigenvalues of the Laplacian matrix of the mirror graph, i.e., $\frac{1}{2}(\mathcal{L}_{\mathcal{G}^{(t)}} + \mathcal{L}_{\mathcal{G}^{(t)}}^\top)$, are non-negative and real (less than 2). A2) is a joint connectivity condition on the communication topology. Intuitively, it means that if the communication connectivity relation among the sensors visits all digraphs of $\mathcal{C}$ in a certain time interval, then for any pair of sensors i and j, sensor i can influence sensor j in this time interval only by local interactions among sensors. A3) is on the ergodicity of the Markovian random switches of communication topologies. From Markov chain theory, a discrete time Markov chain with finite states is ergodic if and only if it is irreducible and aperiodic. Note that we need the above inequalities in A4) for our theoretical analysis, which is a requirement on the network topology.

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 10,2023 at 03:56:46 UTC from IEEE Xplore. Restrictions apply.

The following assumption is a suitable generalization of the excitation condition used in [19] from single sensor to sensor networks.

Assumption 2 (Cooperative Information Assumption): Let $\{\boldsymbol{\varphi}_k^i, \mathcal{F}_k, k \geq 0\}, i = 1, \ldots, n$, be n adapted sequences and there exists an integer $h > 0$ such that $\{\lambda_k, k \geq 0\} \in S^0(\lambda)$ class for some $\lambda \in (0, 1)$, where λ_k is defined by

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

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 2: We remark that the assumption on the conditional mathematical expectation in *Assumption 2* implies that the system signals will have some kind of "persistent excitations" since the prediction of the "future" is non-degenerate given the "past", which is required to track constantly changing unknown signals. Moreover, under *Assumption 2*, the two-layer diffusion filtering network can be shown to fulfill the estimation task cooperatively even if any individual filter cannot, which is a natural property for distributed filters.

C. Stability Results

Denote $\boldsymbol{A}_k = \mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k+\mathscr{L}_{\mathcal{G}_k}^\top-\mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k\mathscr{L}_{\mathcal{G}_k}^\top$ as the random matrix in the error equation. Now, we are in the position to give our first main result on stability. Since the network topology is directed and Markovian switching, the key difficulty is to establish the exponential stability under the assumption that the matrix $\boldsymbol{A}_k$ is non-symmetric, non-independent and non-stationary compared with [6].

Theorem 1: Consider the model (1) and the estimation error equation (3). Suppose that *Assumptions 1* and *2* are satisfied. Then for any $p \geq 1$, there exists a constant $\mu^* \in (0, 1)$, such that for all $\mu \in (0, \mu^*]$, $\{I_{mn} - [\mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k + \mathscr{L}_{\mathcal{G}_k}^\top - \mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k\mathscr{L}_{\mathcal{G}_k}^\top], \quad k \geq 1\}$ is L_p-exponentially stable.

Remark 3: The detailed proof of *Theorem 1* is given in the next section, and by *Theorem 1*, we can obtain a preliminary tracking error bound in the following theorem.

Theorem 2: Consider the model (1) and the estimation error equation (3). Suppose that *Assumptions 1* and *2* are satisfied. If for some $p \geq 1$ and $\beta > 1$, $\sigma_p \triangleq \sup_k \|\xi_k \log^\beta(e+\xi_k)\|_{L_p} < \infty$, hold, where $\xi_k = \|\boldsymbol{V}_k\| + \|\Delta\boldsymbol{\Theta}_{k+1}\|$, then for Algorithm 1 with any initial estimate satisfies $\|\widetilde{\boldsymbol{\Theta}}_0\|_{L_p} < \infty$, there exists a constant $\mu^* \in (0, 1)$, such that for all $\mu \in (0, \mu^*]$, $\{\widetilde{\boldsymbol{\Theta}}_k, k \geq 1\}$ is L_p-stable and

$$\limsup_{k\to\infty}\|\widetilde{\boldsymbol{\Theta}}_k\|_{L_p} \leq c\left[\sigma_p\log\left(e+\sigma_p^{-1}\right)\right], \tag{6}$$

where c is a positive constant.

Remark 4: Since the upper bound for the error can be derived in a similar way as that of [19, Th. 4.2], details will be omitted here. Intuitively, by *Theorem 2* 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, independent, and stationary properties is required in this letter.

D. Proof of Theorem 1

We first prove that the random matrices $\boldsymbol{A}_k$ satisfy the inequality restriction in [12, Lemma 5.8].

Lemma 1: Under *Assumption 1*, there exist constants $\mu^* \in (0, 1)$ and $\varepsilon \in (0, 1)$, such that for any $\mu \in (0, \mu^*]$,

$$\boldsymbol{A}_k^\top\boldsymbol{A}_k \leq (1-\varepsilon)\left(\boldsymbol{A}_k+\boldsymbol{A}_k^\top\right), \ a.s. \tag{7}$$

Proof: For simplicity, we omit the subscript k, $\mathcal{G}_k$, the dimension mn, and $a.s.$ for sample paths hereafter. We have

$$\begin{aligned}\boldsymbol{A}^\top\boldsymbol{A} = {}& \mu^2\boldsymbol{F}\mathscr{A}\mathscr{A}^\top\boldsymbol{F} + \mathscr{L}\mathscr{L}^\top + \mu^2\mathscr{L}\boldsymbol{F}\mathscr{A}\mathscr{A}^\top\boldsymbol{F}\mathscr{L}^\top \\ & + \mu\boldsymbol{F}\mathscr{A}\mathscr{L}^\top + \mu\mathscr{L}\mathscr{A}^\top\boldsymbol{F} \\ & - \mu^2\boldsymbol{F}\mathscr{A}\mathscr{A}^\top\boldsymbol{F}\mathscr{L}^\top - \mu^2\mathscr{L}\boldsymbol{F}\mathscr{A}\mathscr{A}^\top\boldsymbol{F} \\ & - \mu\mathscr{L}\mathscr{A}^\top\boldsymbol{F}\mathscr{L}^\top - \mu\mathscr{L}\boldsymbol{F}\mathscr{A}\mathscr{L}^\top.\end{aligned}$$

Also, we have $\boldsymbol{A}+\boldsymbol{A}^\top = 2\mu\boldsymbol{F}+\mathscr{L}+\mathscr{L}^\top-\mu\mathscr{L}^\top\boldsymbol{F}-\mu\boldsymbol{F}\mathscr{L}-\mu\mathscr{L}\boldsymbol{F}-\mu\boldsymbol{F}\mathscr{L}^\top+\mu\mathscr{L}^\top\boldsymbol{F}\mathscr{L}^\top+\mu\mathscr{L}\boldsymbol{F}\mathscr{L}$.

By *Assumption 1*, denote $\mu_1 = \sqrt{c/2}$, since $0 \leq \boldsymbol{F} \leq I$, then for any $\mu \in (0, \mu_1]$, we have

$$\begin{aligned}& \mu^2\boldsymbol{F}\mathscr{A}\mathscr{A}^\top\boldsymbol{F} + \mathscr{L}\mathscr{L}^\top + \mu^2\mathscr{L}\boldsymbol{F}\mathscr{A}\mathscr{A}^\top\boldsymbol{F}\mathscr{L}^\top \\ & \leq \mu^2\boldsymbol{F}^2 + \mathscr{L}\mathscr{L}^\top + \mu^2\mathscr{L}\mathscr{L}^\top \\ & \leq \mu^2\boldsymbol{F} + (1-c)\left(\mathscr{L}+\mathscr{L}^\top\right) + \mu^2(1-c)\left(\mathscr{L}+\mathscr{L}^\top\right) \\ & \leq \mu^2\boldsymbol{F} + (1-c)\left(\mathscr{L}+\mathscr{L}^\top\right) + \mu^2\left(\mathscr{L}+\mathscr{L}^\top\right) \\ & \leq \left(1-\frac{c}{4}\right)\left(2\mu\boldsymbol{F}+\mathscr{L}+\mathscr{L}^\top\right) - \frac{c}{4}\left(2\mu\boldsymbol{F}+\mathscr{L}+\mathscr{L}^\top\right).\end{aligned}$$

Now we choose $\varepsilon = c/4$ and denote

$$\begin{aligned}\boldsymbol{G} \triangleq {}& \mu\left(\boldsymbol{F}\mathscr{A}\mathscr{L}^\top + \mathscr{L}\mathscr{A}^\top\boldsymbol{F}\right) \\ & - \mu^2\left(\boldsymbol{F}\mathscr{A}\mathscr{A}^\top\boldsymbol{F}\mathscr{L}^\top + \mathscr{L}\boldsymbol{F}\mathscr{A}\mathscr{A}^\top\boldsymbol{F}\right) \\ & - \mu\left(\mathscr{L}\mathscr{A}^\top\boldsymbol{F}\mathscr{L}^\top + \mathscr{L}\boldsymbol{F}\mathscr{A}\mathscr{L}^\top\right) \\ & + \mu(1-\varepsilon)\left(\mathscr{L}^\top\boldsymbol{F}+\boldsymbol{F}\mathscr{L}\right) + \mu(1-\varepsilon)\left(\mathscr{L}\boldsymbol{F}+\boldsymbol{F}\mathscr{L}^\top\right) \\ & - \mu(1-\varepsilon)\left(\mathscr{L}^\top\boldsymbol{F}\mathscr{L}^\top + \mathscr{L}\boldsymbol{F}\mathscr{L}\right).\end{aligned} \tag{8}$$

Then to prove (7), we need only to prove that there exists a constant $\mu^* \in (0, 1)$, such that for any $\mu \in (0, \mu^*]$,

$$\boldsymbol{G} \leq \varepsilon\left(2\mu\boldsymbol{F}+\mathscr{L}+\mathscr{L}^\top\right). \tag{9}$$

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

$$\begin{aligned}& x^\top\boldsymbol{G}x \\ & \leq 2\mu\|x^\top\boldsymbol{F}\|\cdot\|\mathscr{A}\|\cdot\|\mathscr{L}^\top x\| \\ & \quad + 2\mu^2\|x^\top\boldsymbol{F}\|\cdot\|\mathscr{A}\|\cdot\|\mathscr{A}^\top\|\cdot\|\boldsymbol{F}\|\cdot\|\mathscr{L}^\top x\| \\ & \quad + 2\mu\|x^\top\mathscr{L}\|\cdot\|\mathscr{A}^\top\|\cdot\|\boldsymbol{F}\|\cdot\|\mathscr{L}^\top x\| \\ & \quad + 2\mu(1-\varepsilon)\|x^\top\boldsymbol{F}\|\cdot\|\mathscr{L}x\| + 2\mu(1-\varepsilon)\|x^\top\boldsymbol{F}\|\cdot\|\mathscr{L}^\top x\| \\ & \quad + 2\mu(1-\varepsilon)\|x^\top\mathscr{L}\|\cdot\|\boldsymbol{F}\|\cdot\|\mathscr{L}x\| \\ & \leq \left(2\mu(2-\varepsilon)+2\mu^2\right)\|x^\top\boldsymbol{F}^{\frac{1}{2}}\|\cdot\|\mathscr{L}^\top x\| \\ & \quad + 2\mu\|x^\top\mathscr{L}\|\cdot\|\mathscr{L}^\top x\| + 2\mu(1-\varepsilon)\|x^\top\boldsymbol{F}^{\frac{1}{2}}\|\cdot\|\mathscr{L}x\| \\ & \quad + 2\mu(1-\varepsilon)\|x^\top\mathscr{L}\|\cdot\|\mathscr{L}x\|\end{aligned}$$

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 10,2023 at 03:56:46 UTC from IEEE Xplore. Restrictions apply.

$$\leq \left[\sqrt{\frac{\mu}{2}}(2-\varepsilon)+\sqrt{\frac{\mu^3}{2}}\right]\left(2\sqrt{2\mu}\|x^\top \boldsymbol{F}^{\frac{1}{2}}\|\cdot\|\mathscr{L}^\top x\|\right)$$
$$+\mu\left(2\|x^\top\mathscr{L}\|\cdot\|\mathscr{L}^\top x\|\right)$$
$$+\sqrt{\frac{\mu}{2}}(1-\varepsilon)\left(2\sqrt{2\mu}\|x^\top \boldsymbol{F}^{\frac{1}{2}}\|\cdot\|\mathscr{L}x\|\right)$$
$$+\mu(1-\varepsilon)\left(2\|x^\top\mathscr{L}\|\cdot\|\mathscr{L}x\|\right)$$
$$\leq \left[\sqrt{\frac{\mu}{2}}(3-2\varepsilon)+\sqrt{\frac{\mu^3}{2}}\right]x^\top(2\mu\boldsymbol{F})x$$
$$+\left[\sqrt{\frac{\mu}{2}}(3-2\varepsilon)+\sqrt{\frac{\mu^3}{2}}+2\mu(2-\varepsilon)\right]$$
$$\cdot x^\top\left(\mathscr{L}+\mathscr{L}^\top\right)x. \tag{10}$$

Here if we choose μ to satisfy

$$\begin{cases}\sqrt{\frac{\mu}{2}}(3-2\varepsilon)\leq\frac{c}{8},\ \sqrt{\frac{\mu^3}{2}}\leq\frac{c}{8},\\ \sqrt{\frac{\mu}{2}}(3-2\varepsilon)\leq\frac{c}{12},\ \ \sqrt{\frac{\mu^3}{2}}\leq\frac{c}{12},\\ 2\mu(2-\varepsilon)\leq\frac{c}{12},\end{cases}$$

then (9) holds. Hence, we can choose

$$\mu^*=\min\left\{\frac{c^2}{18(6-c)^2},\sqrt[3]{\frac{c^2}{72}},\frac{c}{6(8-c)}\right\}, \tag{11}$$

where $c\in(0,1)$ is a constant which is related to the Laplacian matrix of the network graph. Consequently, there exists a constant $\mu^*\in(0,1)$ such that for any $\mu\in(0,\mu^*]$, (7) holds. This completes the proof. ■

To accomplish the proof of *Theorem* 1, we also need the following lemma.

Lemma 2: Suppose that *Assumptions 1* and 2 are satisfied, then there exist constants $\mu^*\in(0,1)$ and $\varepsilon\in(0,1)$, such that for any $\mu\in(0,\mu^*]$, $\gamma_k\in S^0(\gamma)$, where

$$\gamma_k\triangleq\lambda_{\min}\left\{\mathbb{E}\left[\frac{1}{1+4(1-\varepsilon)h}\sum_{j=k+1}^{k+h}\left(\boldsymbol{A}_j+\boldsymbol{A}_j^\top\right)\Big|\mathcal{F}_k\right]\right\}, \tag{12}$$

and $\gamma=\lambda^\nu,\nu=\frac{h\ell_{m+1}\mu}{(2+\ell_{m+1})(1+h)[1+4(1-\varepsilon)h]}$, where $\ell_{m+1}>0$ is the $(m+1)$th smallest eigenvalue of matrix $\mathbb{E}[\frac{1}{2}(\mathscr{L}_{\mathcal{G}_k}+\mathscr{L}_{\mathcal{G}_k}^\top)|\mathcal{F}_k]$ and μ^* is defined by (13).

Proof: By *Lemma 1*, we know that there exists constants μ^* which is defined in (11) and $\varepsilon=c/4$, such that for any $\mu\in(0,\mu^*]$, (7) holds. Then by [12, *Lemma 5.5*], we know that for any $\mu\in(0,\mu^*]$, $\gamma_k\in[0,1]$ holds. By the notations in *Lemma 1*, we then have $\boldsymbol{A}_j+\boldsymbol{A}_j^\top=2\mu\boldsymbol{F}_j+\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top-\mu\mathscr{L}_{\mathcal{G}_j}^\top\boldsymbol{F}_j-\mu\boldsymbol{F}_j\mathscr{L}_{\mathcal{G}_j}-\mu\mathscr{L}_{\mathcal{G}_j}\boldsymbol{F}_j-\mu\boldsymbol{F}_j\mathscr{L}_{\mathcal{G}_j}^\top+\mu\mathscr{L}_{\mathcal{G}_j}^\top\boldsymbol{F}_j\mathscr{L}_{\mathcal{G}_j}^\top+\mu\mathscr{L}_{\mathcal{G}_j}\boldsymbol{F}_j\mathscr{L}_{\mathcal{G}_j}$. Similar to the proof of *Lemma 1*, we have

$$\mu\mathscr{L}_{\mathcal{G}_j}^\top\boldsymbol{F}_j+\mu\boldsymbol{F}_j\mathscr{L}_{\mathcal{G}_j}+\mu\mathscr{L}_{\mathcal{G}_j}\boldsymbol{F}_j+\mu\boldsymbol{F}_j\mathscr{L}_{\mathcal{G}_j}^\top$$
$$-\mu\mathscr{L}_{\mathcal{G}_j}^\top\boldsymbol{F}_j\mathscr{L}_{\mathcal{G}_j}^\top-\mu\mathscr{L}_{\mathcal{G}_j}\boldsymbol{F}_j\mathscr{L}_{\mathcal{G}_j}$$
$$\leq\sqrt{\frac{\mu}{2}}\left(2\mu\boldsymbol{F}_j+\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top\right)+\sqrt{\frac{\mu}{2}}\left(2\mu\boldsymbol{F}_j+\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top\right)$$
$$+\mu\left(\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top+\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top\right)$$
$$\leq5\sqrt{\frac{\mu}{2}}\left(2\mu\boldsymbol{F}_j+\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top\right).$$

Then we can obtain

$$\boldsymbol{A}_j+\boldsymbol{A}_j^\top\geq\left(1-5\sqrt{\frac{\mu}{2}}\right)\left(2\mu\boldsymbol{F}_j+\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top\right)$$
$$\geq\left(1-5\sqrt{\frac{\mu}{2}}\right)\mu\left(2\boldsymbol{F}_j+\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top\right).$$

Here we choose μ to satisfy $5\sqrt{\frac{\mu}{2}}\leq0.5$, then there exists a new μ^* satisfying

$$\mu^*=\min\left\{\frac{c^2}{18(6-c)^2},\sqrt[3]{\frac{c^2}{72}},\frac{c}{6(8-c)},\frac{1}{50}\right\}, \tag{13}$$

such that for any $\mu\in(0,\mu^*]$, $\boldsymbol{A}_j+\boldsymbol{A}_j^\top\geq0.5\mu(2\boldsymbol{F}_j+\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top)$. Denote

$$\rho_k\triangleq\lambda_{\min}\left\{\mathbb{E}\left[a\sum_{j=k+1}^{k+h}\left(\boldsymbol{F}_j+\frac{1}{2}(\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top)\right)\Big|\mathcal{F}_k\right]\right\}.$$

where $a=\frac{\mu}{1+4(1-\varepsilon)h}$. Since $0\leq\rho_k\leq\gamma_k\leq1$, we know that to prove (12), we need only to prove $\rho_k\in S^0(\lambda^\nu)$.

By *Assumption 1*, there exists an integer q such that for any initial time $k\geq0$ and initial state, the Markov chain m_t will visit all its states in the time interval $[k,k+q]$ with positive probability $p_0>0$, which does not depend on time instant k. Here we assume that $h\geq q$. Since the communication topology is balanced and connected, we know that $\mathbb{E}[\sum_{j=k+1}^{k+h}\frac{1}{2}(\mathcal{L}_{\mathcal{G}_j}+\mathcal{L}_{\mathcal{G}_j}^\top)|\mathcal{F}_k]$ has only one zero eigenvalue whose unit eigenvector is $\frac{1}{\sqrt{n}}(1,\ldots,1)^\top$, i.e., $\frac{1}{\sqrt{n}}\mathbf{1}$ where $\mathbf{1}=(1,\ldots,1)_{n\times1}^\top$. Correspondingly, $\mathbb{E}[\sum_{j=k+1}^{k+h}\frac{1}{2}(\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top)|\mathcal{F}_k]$ has m zero eigenvalues, and the other eigenvalues of $\mathbb{E}[\sum_{j=k+1}^{k+h}\frac{1}{2}(\mathscr{L}_{\mathcal{G}_j}+\mathscr{L}_{\mathcal{G}_j}^\top)|\mathcal{F}_k]$ are $\ell_{m+1}\leq\cdots\leq\ell_{mn}$ whose orthogonal unit eigenvectors are denoted as $\boldsymbol{\xi}_{m+1},\ldots,\boldsymbol{\xi}_{mn}$ correspondingly.

The following proof is similar to [12, Lemma 5.10], then we can obtain that $\{\rho_k\}\in S^0(\rho)$, where $\rho=\lambda^\nu$.

Proof of Theorem 1: By *Lemmas 1* and 2, we know that there exists a constant $\mu^*\in(0,1)$, such that for any $\mu\in(0,\mu^*]$ $\{\mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k+\mathscr{L}_{\mathcal{G}_k}^\top-\mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k\mathscr{L}_{\mathcal{G}_k}^\top\}\in S_p(\gamma^\alpha)$, where and

$$\alpha=\begin{cases}\frac{\varepsilon}{8h[1+4(1-\varepsilon)h]^2}, & 1\leq p\leq2;\\ \frac{\varepsilon}{4h[1+4(1-\varepsilon)h]^2p}, & p>2.\end{cases} \tag{14}$$

Then by *Definition 1*, it is obvious that $\{I_{mn}-(\mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k+\mathscr{L}_{\mathcal{G}_k}^\top-\mu\mathscr{A}_{\mathcal{G}_k}^\top\boldsymbol{F}_k\mathscr{L}_{\mathcal{G}_k}^\top),k\geq1\}$ is L_p-exponentially stable $(p\geq1)$. ■

IV. Simulation Results

We take $n=3$ for example. The state space of m_k is $\mathcal{S}=\{1,2,3\}$, and transition probability matrix is taken as

$$P=\begin{pmatrix}0.2 & 0.3 & 0.5\\ 0.5 & 0.1 & 0.4\\ 0.6 & 0.2 & 0.2\end{pmatrix},$$

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 10,2023 at 03:56:46 UTC from IEEE Xplore. Restrictions apply.

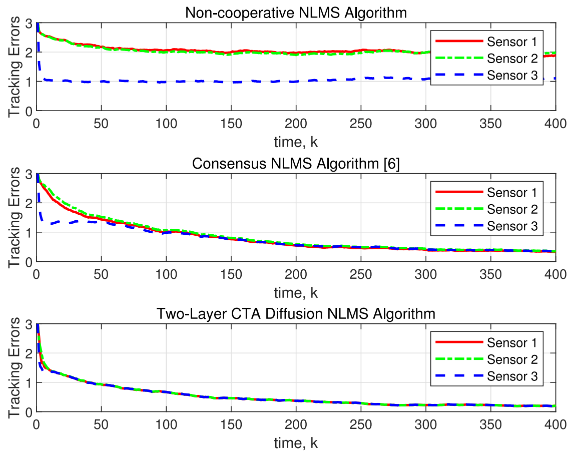

Fig. 1. Tracking errors of the three sensors with γ=0.1.

where each state stands for the following adjacency matrices

$$\mathcal{A}_{\mathcal{G}^{(1)}} = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/6 & 2/3 & 1/6 \\ 1/2 & 0 & 1/2 \end{pmatrix}, \mathcal{A}_{\mathcal{G}^{(2)}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{pmatrix},$$

$$\mathcal{A}_{\mathcal{G}^{(3)}} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

then the corresponding digraphs $\mathcal{G}^{(t)}, t = 1, 2, 3$ satisfy *Assumption 1*.

We will estimate or track an unknown 3-dimensional signal $\boldsymbol{\theta}_k$ with $\Delta\boldsymbol{\theta}_k = \gamma\boldsymbol{\omega}_k$ in (2), where $\gamma = 0.1$ and the parameter variation $\boldsymbol{\omega}_k \sim N(0, 0.1, 3, 1)$ (Gaussian distribution). In both cases, the observation noises $\{v_k^i, k \geq 1, i = 1, 2, 3\}$ are independent and identically distributed with $v_k^i \sim N(0, 0.1, 1, 1)$ in (1), where $\boldsymbol{\varphi}_k^i (i = 1, 2, 3)$ are generated by a state space model in [12]. Then *Assumption 2* is satisfied with $h = 2$. For numerical simulations, let $\boldsymbol{\theta}_0 = (1, 1, 1)^\top, \hat{\boldsymbol{\theta}}_0^i = (0, 0, 0)^\top (i = 1, 2, 3)$, $\mu = 0.8$. 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 $\{\|\hat{\boldsymbol{\theta}}_k^{i,j} - \boldsymbol{\theta}_k^j\|^2, k = 1, 2, \ldots, 400\}$ $(j = 1, \ldots, m)$, where the superscript j denotes the j-th simulation result. We use $\frac{1}{m}\sum_{j=1}^m \|\hat{\boldsymbol{\theta}}_k^{i,j} - \boldsymbol{\theta}_k^j\|^2$ $(i = 1, 2, 3, k = 1, 2, \ldots, 400)$ to denote the tracking errors of the three sensors in Fig. 1.

In Fig. 1, the upper one is the individual situation in which the tracking errors of the three sensors keep large, and the middle and lower ones are the distributed situation in which all the tracking errors converge nicely as k increases. This indicates that even if any node cannot estimate the unknown parameter individually, the tracking task can still be fulfilled through local communications among nodes in the network. Also, compared with the consensus NLMS algorithm, the convergence rate is faster since the data diffuse more thoroughly in the two-layer diffusion strategy.

V. Concluding Remarks

This letter has analyzed a general class of two-layer diffusion adaptive filtering algorithms over directed Markovian switching graphs, which are used for local estimate and measurement information communications. We have established the stability analysis under a general cooperative information assumption on the system regressors, without independence and stationarity considerations. Our main results also demonstrate that the two-layer diffusion NLMS filters studied in this letter can track a dynamic process of interest from noisy measurements by a set of sensors working cooperatively, in the natural scenario where any sensor cannot fulfill the estimation task individually.

References

[1] 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.

[2] Q. Zhang and J. F. Zhang, "Distributed parameter estimation over unreliable networks with Markovian switching topologies," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2545–2560, Oct. 2012.

[3] Q. Y Liu, Z. D. Wang, X. He, and D. H. Zhou, "Event-based distributed filtering over Markovian switching topologies," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1595–1602, Apr. 2019.

[4] S. Y. Xie and L. Guo, "Analysis of NLMS-based consensus adaptive filters under a general information condition," *SIAM J. Control Optim.*, vol. 56, no. 5, pp. 3404–3431, Sep. 2018.

[5] S. Y. Xie and L. Guo, "A necessary and sufficient condition for stability of LMS-based consensus adaptive filters," *Automatica*, vol. 93, pp. 12–19, Jul. 2018.

[6] S. Xie and L. Y. Wang, "Analysis of consensus adaptive filters over directed and Markovian switching networks," in *Proc. 13th Asian Control Conf. (ASCC)*, Jeju, South Korea, 2022, pp. 2030–2035.

[7] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[8] 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.

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

[11] 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.

[12] S. Y. Xie and L. Guo, "Analysis of distributed adaptive filters based on diffusion strategies over sensor networks," *IEEE Trans. Autom. Control*, vol. 63, no. 11, pp. 3643–3658, Nov. 2018.

[13] H. C. Yan, Q. Yang, H. Zhang, F. W. Yang, and X. S. Zhan, "Distributed H_∞ state estimation for a class of filtering networks with time-varying switching topologies and packet losses," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 12, pp. 2047–2057, Dec. 2018.

[14] S. Wang and W. Ren, "On the convergence conditions of distributed dynamic state estimation using sensor networks: A unified framework," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 4, pp. 1300–1316, Jul. 2018.

[15] Q. Liu, Z. Wang, X. He, and D. H. Zhou, "On Kalman-consensus filtering with random link failures over sensor networks," *IEEE Trans. Autom. Control*, vol. 63, no. 8, pp. 2701–2708, Aug. 2018.

[16] 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.*, vol. 30, no. 3, pp. 155–171, May 2013.

[17] J. Sjöberg et al., "Nonlinear black-box modeling in system identification: A unified overview," *Automatica*, vol. 31, pp. 1691–1724, Dec. 1995.

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

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

Authorized licensed use limited to: CAS Academy of Mathematics & Systems Science. Downloaded on December 10,2023 at 03:56:46 UTC from IEEE Xplore. Restrictions apply.