Stability of FFLS-Based Diffusion Adaptive Filter Under Cooperative Excitation Condition

Die Gan, *Member, IEEE*, Siyu Xie, Zhixin Liu, *Member, IEEE*, and Jinhu Lü, *Fellow, IEEE*

***Abstract*—In this article, we consider the distributed filtering problem over sensor networks such that all sensors cooperatively track unknown time-varying parameters by using local information. A distributed forgetting factor least squares algorithm is proposed by minimizing a local cost function formulated as a linear combination of accumulative estimation error. Stability analysis of the algorithm is provided under a cooperative excitation condition which contains spatial union information to reflect the cooperative effect of all sensors. Furthermore, we generalize theoretical results to the case of Markovian switching directed graphs. The main difficulties of theoretical analysis lie in how to analyze properties of the product of nonindependent and nonstationary random matrices. Some techniques, such as stability theory, algebraic graph theory, and Markov chain theory are employed to deal with the above issue. Our theoretical results are obtained without relying on the independence or stationarity assumptions of regression vectors which are commonly used in existing literature. Finally, numerical simulations are provided to demonstrate the effectiveness of our theoretical results.**

***Index Terms*—Cooperative excitation condition, distributed forgetting factor least squares (FFLS), exponential stability, Markovian switching topology, stochastic dynamic systems.**

I. Introduction

OWING to the capability to process the collaborative data, wireless sensor networks have attracted increasing research attention in diverse areas, including consensus seeking [1], [2], resource allocation [3], [4], and formation control [5], [6]. How to design the distributed adaptive estimation and filtering algorithms to cooperatively estimate unknown parameters has become one of the most important research topics. Compared with centralized estimation algorithms where a fusion center is needed to collect and process information measured by all sensors, the distributed ones can estimate or track an unknown parameter process of interest cooperatively by using local noisy measurements. Therefore, the distributed algorithms are easier to be implemented because of their robustness to network link failure, privacy protection, and reduction on communication and computation costs.

Based on classical estimation algorithms and typical distributed strategies, such as the incremental, diffusion and consensus, a number of distributed adaptive estimation or filtering algorithms have been investigated (cf., [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]), e.g., the consensus-based least mean squares (LMS), the diffusion recursive least squares (RLS), the incremental LMS, and the diffusion forgetting factor least squares (FFLS). The performance analysis of the distributed algorithms is also studied under some information conditions. For deterministic signals or deterministic system matrices, Battistelli and Chisci in [7] provided the mean-square boundedness of the state estimation error of the distributed Kalman filter algorithm under a collectively observable condition. Chen et al. [8] studied the convergence of distributed adaptive identification algorithm under a cooperative persistent excitation (PE) condition. Javed et al. [9] presented stability analysis of the cooperative gradient algorithm for the deterministic regression vectors satisfying a cooperative PE condition. Note that the signals are often random since they are generated from dynamic systems affected by noises. For the random regression vector case, Barani et al. [10] studied the convergence of distributed stochastic gradient descent algorithm with independent and identically distributed (i.i.d.) signals. Schizas et al. [11] provided the stability analysis of a distributed LMS-type adaptive algorithm under the strictly stationary and ergodic regression vectors. Zhang et al. [12] studied the mean square performance of a diffusion FFLS algorithm with independent input signals. Mateos and Giannakis in [15] presented the stability and performance analysis of the distributed FFLS algorithm under the spatio-temporally white regression vectors condition.

We remark that most theoretical results mentioned in above literature were established by requiring regression vectors to be either deterministic and satisfy PE conditions, or random but satisfy independence, stationarity, and ergodicity conditions. In

Manuscript received 3 February 2024; accepted 2 April 2024. Date of publication 12 April 2024; date of current version 25 October 2024. This work was supported in part by the Natural Science Foundation of China under Grant T2293772, in part by the National Key R&D Program of China under Grant 2018YFA0703800, and in part by the Strategic Priority Research Program of Chinese Academy of Sciences under Grant XDA27000000. Recommended by Associate Editor G. A. Rovithakis. *(Corresponding author: Zhixin Liu.)*

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

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).

Zhixin Liu is with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China (e-mail: lzx@amss.ac.cn).

Jinhu Lü is with the School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China, and also with the Zhongguancun Laboratory, Beijing 100094, China (e-mail: jhlu@iss.ac.cn).

Digital Object Identifier 10.1109/TAC.2024.3388128

fact, the observed data are often random and hard to satisfy the above statistical assumptions, since they are generated by complex dynamic systems where feedback loops inevitably exist ([20]). The main difficulty in the performance analysis of distributed algorithms is to analyze the product of random matrices involved in estimation error equations. In order to relax the above stringent conditions on random regression vectors, some progress has been made on distributed adaptive estimation and filtering algorithms under undirected graphs. For estimating time-invariant parameters, the convergence analysis of the distributed SG algorithm and the distributed RLS algorithm is provided in [21] and [22] under cooperative excitation conditions. For tracking a time-varying parameter, Xie and Guo in [16] and [23] proposed the weakest possible cooperative information conditions to guarantee the stability and performance of consensus-based and diffusion-based LMS algorithms. As we know, an unfortunate property of the traditional RLS estimator is that the gain rapidly tends to zero, which makes the ability of RLS estimator to track time-varying parameters lost. The FFLS algorithm incorporates a forgetting factor to place more weight on recent observations which can overcome this deficiency in some sense. Moreover, compared with the LMS algorithm, the FFLS algorithm can generate more accurate estimates in the transient phase [24], but the stability analysis for the distributed FFLS algorithm is still lacking. This article focuses on the design and stability analysis of the distributed FFLS algorithm without relying on the independence, stationarity, or ergodicity assumptions on regression vectors.

The information exchange between sensors is an important factor for the performance of distributed estimation algorithms, and previous studies often assume that the networks are undirected and time-invariant (cf., [16], [21], [22], [23]). In practice, they might not be bidirectional or time-invariant due to the heterogeneity of sensors and signal losses caused by the temporary deterioration in the communication link. One approach is to model the networks which randomly change over time as an i.i.d. process, see, e.g., [25], [26]. However, the loss of connection usually occurs with correlations [27]. Another approach is to model the random switching process as a Markov chain whose states correspond to possible communication topologies, see [2], [27], [28], [29] among many others. Some studies on the distributed algorithms with deterministic or temporally independent measurement matrix under Markovian switching topologies are given in [30] and [31].

In this article, we consider the distributed filtering problem over sensor networks where all sensors aim at collectively tracking an unknown randomly time-varying parameter vector. We first introduce a forgetting factor into the local accumulative cost function formulated as a linear combination of local estimation errors between the observation signals and the prediction signals. By minimizing the local cost function, we propose the distributed FFLS algorithm based on the diffusion strategy over the fixed undirected graph. The stability analysis of the distributed FFLS algorithm is provided under a cooperative excitation condition. Moreover, we generalize the theoretical results to the case of Markovian switching directed sensor networks. The main contributions of this article can be summarized as the following aspects.

1) In comparison with [16] and [21], the random matrices in the error equation of the diffusion FFLS algorithm are not symmetric and the adaptive gain is no longer a scalar. We establish the exponential stability of the homogeneous part of the estimation error equation and the bound of the tracking error by virtue of the specific structure of the proposed diffusion FFLS algorithm and the stability theory of stochastic dynamic systems.
2) Different from the theoretical results of distributed FFLS algorithms in [12] and [15], where regression vectors are required to satisfy the independent or spatio-temporally uncorrelated assumptions, our theoretical analysis is obtained without relying on such stringent conditions, which makes it possible to be applied to the stochastic feedback systems.
3) The cooperative excitation condition introduced in this article is a temporal and spatial union information condition on the random regression vectors, which can reveal the cooperative effect of multiple sensors in a certain sense, i.e., the whole sensor network can cooperatively finish the estimation task, even if any individual sensor cannot due to lack of necessary information.

The rest of this article is organized as follows. In Section II, we give the problem formulation of this article. Section III presents the distributed FFLS algorithm. The stability of the proposed algorithm under fixed undirected graph and Markovian switching directed graphs are given in Sections IV and V, respectively. Simulation examples are provided in Section VI. Finally, Section VII concludes this article.

II. Problem Formulation

A. Matrix Theory

In this article, we use $\mathbb{R}^m$ to denote the set of m-dimensional real vectors, $\mathbb{R}^{m\times n}$ to denote the set of real matrices with m rows and n columns, and $\boldsymbol{I}_m$ to denote the m-dimensional square identity matrix. For a matrix $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\|\boldsymbol{A}\|$ denotes its Euclidean norm, i.e., $\|\boldsymbol{A}\|\triangleq(\lambda_{\max}(\boldsymbol{A}\boldsymbol{A}^T))^{\frac{1}{2}}$, where the notation T denotes the transpose operator and $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the matrix. Correspondingly, $\lambda_{\min}(\cdot)$ and $tr(\cdot)$ denote the smallest eigenvalue and the trace of the matrix, respectively. The notation $\mathrm{col}(\cdot,\ldots,\cdot)$ is used to denote a vector stacked by the specified vectors, and $\mathrm{diag}(\cdot,\ldots,\cdot)$ is used to denote a block matrix formed in a diagonal manner of the corresponding vectors or matrices.

For a matrix $\boldsymbol{A}=[a_{ij}]\in\mathbb{R}^{m\times m}$, if $\sum_{j=1}^m a_{ij}=1$ holds for all $i=1,\ldots,m$, then it is called stochastic. The Kronecker product of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is denoted by $\boldsymbol{A}\otimes\boldsymbol{B}$. For two real symmetric matrices $\boldsymbol{X}\in\mathbb{R}^{n\times n}$ and $\boldsymbol{Y}\in\mathbb{R}^{n\times n}$, $\boldsymbol{X}\geq\boldsymbol{Y}$ ($\boldsymbol{X}>\boldsymbol{Y}$, $\boldsymbol{X}\leq\boldsymbol{Y}$, $\boldsymbol{X}<\boldsymbol{Y}$) means that $\boldsymbol{X}-\boldsymbol{Y}$ is a semipositive (positive, seminegative, negative) definite matrix. For a matrix sequence $\{\boldsymbol{A}_t\}$ and a positive scalar sequence $\{a_t\}$, the equation $\boldsymbol{A}_t=O(a_t)$ means that there exists a positive constant

C independent of t and a_t such that $\|\boldsymbol{A}_t\| \leq Ca_t$ holds for all $t \geq 0$.

The matrix inversion formula is often used in this article and we list it as follows.

Lemma 2.1 (Matrix inversion formula [32]): For any matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, and $\boldsymbol{D}$ with suitable dimensions, the following formula:

$$(\boldsymbol{A}+\boldsymbol{BDC})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D}^{-1}+\boldsymbol{CA}^{-1}\boldsymbol{B})^{-1}\boldsymbol{CA}^{-1}$$

holds, provided that the relevant matrices are invertible.

B. Graph Theory

We use graphs to model the communication topology between sensors. A directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ is composed of a vertex set $\mathcal{V} = \{1, 2, 3, \ldots, n\}$ which stands for the set of sensors (i.e., nodes), $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the edge set, and $\mathcal{A} = [a_{ij}]_{1\leq i,j\leq n}$ is the weighted adjacency matrix. A directed edge $(i, j) \in \mathcal{E}$ means that the jth sensor can receive the data from the ith sensor, and sensors i and j are called the parent and child sensors, respectively. The elements of matrix $\mathcal{A}$ satisfy $a_{ij} > 0$ if $(i, j) \in \mathcal{E}$ and $a_{ij} = 0$ otherwise. The in-degree and out-degree of sensor i are defined by $\deg_{\text{in}}(i) = \sum_{j=1}^{n} a_{ji}$ and $\deg_{\text{out}}(i) = \sum_{j=1}^{n} a_{ij}$, respectively. The digraph $\mathcal{G}$ is called balanced if $\deg_{\text{in}}(i) = \deg_{\text{out}}(i)$ for $i = 1, \ldots, n$. Here, we assume that $\mathcal{A}$ is a stochastic matrix. The neighbor set of i is denoted as $\mathcal{N}_i = \{j \in \mathcal{V}, (j, i) \in \mathcal{E}\}$, and the sensor i is also included in this set. For a given positive integer k, the union of k digraphs $\{\mathcal{G}_j = (\mathcal{V}, \mathcal{E}_j, \mathcal{A}_j), 1 \leq j \leq k\}$ with the same node set is denoted by $\cup_{j=1}^{k}\mathcal{G}_j = (\mathcal{V}, \cup_{j=1}^{k}\mathcal{E}_j, \frac{1}{k}\sum_{j=1}^{k}\mathcal{A}_j)$. A directed path from i_1 to i_l consists of a sequence of sensors $i_1, i_2, \ldots i_l (l \geq 2)$, such that $(i_k, i_{k+1}) \in \mathcal{E}$ for $k = 1, \ldots, l-1$. The digraph $\mathcal{G}$ is said to be strongly connected if for any senor there exist directed paths from this sensor to all other sensors. For the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, if $a_{ij} = a_{ji}$ for all $i, j \in \mathcal{V}$, then it is called an undirected graph. The diameter $D_{\mathcal{G}}$ of the undirected graph $\mathcal{G}$ is defined as the maximum shortest length of paths between any two sensors.

C. Observation Model

Consider a network consisting of n sensors (labeled $1, \ldots, n$) whose task is to estimate an unknown time-varying parameter $\boldsymbol{\theta}_t$ by cooperating with each other. We assume that the measurement $\{y_{t,i}, \boldsymbol{\varphi}_{t,i}\}$ at the sensor i obeys the following discrete-time stochastic regression model:

$$y_{t+1,i} = \boldsymbol{\varphi}_{t,i}^T \boldsymbol{\theta}_t + w_{t+1,i} \tag{1}$$

where $y_{t,i}$ is the scalar output of the sensor i at time t, $\boldsymbol{\varphi}_{t,i} \in \mathbb{R}^m$ is the random regression vector, $\{w_{t,i}\}$ is a noise process, and $\boldsymbol{\theta}_t$ is the unknown m-dimensional time-varying parameter whose variation at time t is denoted by $\Delta\boldsymbol{\theta}_t$, i.e.,

$$\Delta\boldsymbol{\theta}_t \triangleq \boldsymbol{\theta}_{t+1} - \boldsymbol{\theta}_t, t \geq 0. \tag{2}$$

Note that when $\Delta\boldsymbol{\theta}_t \equiv 0$, $\boldsymbol{\theta}_t$ becomes a constant vector. For the special case where $w_{t+1,i}$ is a moving average process and $\boldsymbol{\varphi}_{t,i}$ consists of current and past input–output data, i.e.,

$$\boldsymbol{\varphi}_{t,i}^T = [y_{t,i}, \ldots, y_{t-p,i}, u_{t,i}, \ldots, u_{t-q,i}]$$

with $u_{t,i}$ being the input signal of the sensor i at time t, then the model (1) can be reduced to the well-known autoregressive-moving average with exogenous input model with time-varying coefficients.

Algorithm 1: Standard noncooperative FFLS algorithm.

For any given sensor $i \in \{1, \ldots, n\}$, begin with an initial estimate $\hat{\boldsymbol{\theta}}_{0,i} \in \mathbb{R}^m$ and an initial positive definite matrix $\hat{\boldsymbol{P}}_{0,i} \in \mathbb{R}^{m\times m}$. The standard FFLS is recursively defined at time $t \geq 0$ as follows:

$$\hat{\boldsymbol{\theta}}_{t+1,i} = \hat{\boldsymbol{\theta}}_{t,i} + \frac{\hat{\boldsymbol{P}}_{t,i}\boldsymbol{\varphi}_{t,i}}{\alpha + \boldsymbol{\varphi}_{t,i}^T\hat{\boldsymbol{P}}_{t,i}\boldsymbol{\varphi}_{t,i}}(y_{t+1,i} - \boldsymbol{\varphi}_{t,i}^T\hat{\boldsymbol{\theta}}_{t,i}),$$

$$\hat{\boldsymbol{P}}_{t+1,i} = \frac{1}{\alpha}\left(\hat{\boldsymbol{P}}_{t,i} - \frac{\hat{\boldsymbol{P}}_{t,i}\boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}_{t,i}^T\hat{\boldsymbol{P}}_{t,i}}{\alpha + \boldsymbol{\varphi}_{t,i}^T\hat{\boldsymbol{P}}_{t,i}\boldsymbol{\varphi}_{t,i}}\right).$$

III. Distributed FFLS Algorithm

Tracking a time-varying signal is a fundamental problem in system identification and signal processing. The well-known RLS estimator with a constant forgetting factor $\alpha \in (0, 1)$ is often used to track time-varying parameters, which is defined by

$$\hat{\boldsymbol{\theta}}_{t+1,i} \triangleq \arg\min_{\boldsymbol{\beta}} \sum_{k=0}^{t} \alpha^{t-k}(y_{k+1,i} - \boldsymbol{\beta}^T\boldsymbol{\varphi}_{k,i})^2. \tag{3}$$

With some simple manipulations using the matrix inversion formula, we can obtain the following recursive FFLS algorithm (Algorithm 1) for an individual sensor.

However, due to the limited sensing ability of each sensor, it is often the case where the measurements obtained by each sensor can only reflect partial information of the unknown parameter. In such a case, if only local measurements of the sensor itself are utilized to perform the estimation task (see Algorithm 1), then at most part of the unknown parameter rather than the whole vector can be estimated. Thus, in this article, we aim at designing a distributed adaptive estimation algorithm such that all sensors cooperatively track the unknown time-varying parameter $\boldsymbol{\theta}_t$ by using random regression vectors and the observation signals from its neighbors. To simplify the analysis, in this section, we use a fixed undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ to model the communication topology of n sensors.

We first introduce the following local cost function $\sigma_{t+1,i}(\boldsymbol{\beta})$ for each sensor i at the time instant $t \geq 0$ recursively formulated as a linear combination of its neighbors' local estimation error between the observation signal and the prediction signal

$$\sigma_{t+1,i}(\boldsymbol{\beta}) = \sum_{j\in\mathcal{N}_i} a_{ij}\left(\alpha\sigma_{t,j}(\boldsymbol{\beta}) + (y_{t+1,j} - \boldsymbol{\beta}^T\boldsymbol{\varphi}_{t,j})^2\right). \tag{4}$$

with $\sigma_{0,i}(\boldsymbol{\beta}) = 0$. Set

$\boldsymbol{\sigma}_t(\boldsymbol{\beta}) = \text{col}\{\sigma_{t,1}(\boldsymbol{\beta}), \ldots, \sigma_{t,n}(\boldsymbol{\beta})\}$

$\boldsymbol{e}_{t+1}(\boldsymbol{\beta}) = \text{col}\{(y_{t+1,1} - \boldsymbol{\beta}^T\boldsymbol{\varphi}_{t,1})^2, \cdots, (y_{t+1,n} - \boldsymbol{\beta}^T\boldsymbol{\varphi}_{t,n})^2\}$.

Hence by (4), we have

$$\begin{aligned}\boldsymbol{\sigma}_{t+1}(\boldsymbol{\beta}) &= \alpha\mathcal{A}\boldsymbol{\sigma}_t(\boldsymbol{\beta}) + \mathcal{A}\boldsymbol{e}_{t+1}(\boldsymbol{\beta}) \\ &= \alpha^2\mathcal{A}^2\boldsymbol{\sigma}_{t-1}(\boldsymbol{\beta}) + \alpha\mathcal{A}^2\boldsymbol{e}_t(\boldsymbol{\beta}) + \mathcal{A}\boldsymbol{e}_{t+1}(\boldsymbol{\beta}) \\ &= \cdots \\ &= \alpha^{t+1}\mathcal{A}^{t+1}\boldsymbol{\sigma}_0(\boldsymbol{\beta}) + \sum_{k=0}^{t}\alpha^{t-k}\mathcal{A}^{t+1-k}\boldsymbol{e}_{k+1}(\boldsymbol{\beta}) \\ &= \sum_{k=0}^{t}\alpha^{t-k}\mathcal{A}^{t+1-k}\boldsymbol{e}_{k+1}(\boldsymbol{\beta})\end{aligned}$$

which implies that

$$\sigma_{t+1,i}(\boldsymbol{\beta}) = \sum_{j=1}^{n}\sum_{k=0}^{t}\alpha^{t-k}a_{ij}^{(t+1-k)}(y_{k+1,j} - \boldsymbol{\beta}^T\boldsymbol{\varphi}_{k,j})^2 \quad (5)$$

where $a_{ij}^{(t+1-k)}$ is the ith row, the jth column entry of the matrix $\mathcal{A}^{t+1-k}$.

By minimizing the local cost function $\sigma_{t+1,i}(\boldsymbol{\beta})$ in (5), we obtain the distributed FFLS estimate $\hat{\boldsymbol{\theta}}_{t+1,i}$ of the unknown time-varying parameter for sensor i, i.e.,

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{t+1,i} &\triangleq \arg\min_{\boldsymbol{\beta}}\sigma_{t+1,i}(\boldsymbol{\beta}) \\ &= \left[\sum_{j=1}^{n}\sum_{k=0}^{t}\alpha^{t-k}a_{ij}^{(t+1-k)}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T\right]^{-1} \\ &\quad \left(\sum_{j=1}^{n}\sum_{k=0}^{t}\alpha^{t-k}a_{ij}^{(t+1-k)}\boldsymbol{\varphi}_{k,j}y_{k+1,j}\right). \quad (6)\end{aligned}$$

Denote $\boldsymbol{P}_{t+1,i} = (\sum_{j=1}^{n}\sum_{k=0}^{t}\alpha^{t-k}a_{ij}^{(t+1-k)}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T)^{-1}$. Then, we write it into the following recursive form:

$$\boldsymbol{P}_{t+1,i}^{-1} = \sum_{j\in\mathcal{N}_i}a_{ij}(\alpha\boldsymbol{P}_{t,j}^{-1} + \boldsymbol{\varphi}_{t,j}\boldsymbol{\varphi}_{t,j}^T). \quad (7)$$

By (6), we similarly have

$$\hat{\boldsymbol{\theta}}_{t+1,i} = \boldsymbol{P}_{t+1,i}\sum_{j\in\mathcal{N}_i}a_{ij}(\alpha\boldsymbol{P}_{t,j}^{-1}\hat{\boldsymbol{\theta}}_{t,j} + \boldsymbol{\varphi}_{t,j}y_{t+1,j}). \quad (8)$$

Note that in the above derivation, we assume that the matrix $\sum_{j=1}^{n}\sum_{k=0}^{t}\alpha^{t-k}a_{ij}^{(t+1-k)}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T$ is invertible which is usually not satisfied for small t. To solve this problem, we take the initial matrix $\boldsymbol{P}_{0,i}$ to be positive definite. Then, (7) can be modified into the following equation:

$$\boldsymbol{P}_{t+1,i} = \left(\sum_{j=1}^{n}\sum_{k=0}^{t}\alpha^{t-k}a_{ij}^{(t+1-k)}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T + \sum_{j=1}^{n}\alpha^{t+1}a_{ij}^{(t+1)}\boldsymbol{P}_{0,j}^{-1}\right)^{-1}. \quad (9)$$

This modification can make the matrix $\boldsymbol{P}_{t+1,i}$ invertible for all $t \geq 0$, but it will not affect the asymptotic analysis and results of the algorithm since (9) has the same recursive form as (7).

To design the distributed algorithm, we denote

$$\bar{\boldsymbol{P}}_{t+1,i}^{-1} = \alpha\boldsymbol{P}_{t,i}^{-1} + \boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}_{t,i}^T. \quad (10)$$

By Lemma 2.1, we have $\bar{\boldsymbol{P}}_{t+1,i} = \frac{1}{\alpha}(\boldsymbol{P}_{t,i} - \frac{\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}_{t,i}^T\boldsymbol{P}_{t,i}}{\alpha + \boldsymbol{\varphi}_{t,i}^T\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}})$. Hence,

$$\begin{aligned}\bar{\boldsymbol{\theta}}_{t+1,i} &\triangleq \bar{\boldsymbol{P}}_{t+1,i}(\alpha\boldsymbol{P}_{t,i}^{-1}\hat{\boldsymbol{\theta}}_{t,i} + \boldsymbol{\varphi}_{t,i}y_{t+1,i}) \\ &= \hat{\boldsymbol{\theta}}_{t,i} + \frac{\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}}{\alpha + \boldsymbol{\varphi}_{t,i}^T\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}}(y_{t+1,i} - \boldsymbol{\varphi}_{t,i}^T\hat{\boldsymbol{\theta}}_{t,i}).\end{aligned}$$

Therefore, we get the following distributed FFLS algorithm of diffusion type, i.e., Algorithm 2.

Note that when $\mathcal{A} = \boldsymbol{I}_n$, the distributed FFLS algorithm will degenerate to the classical FFLS (i.e., Algorithm 1), and when $\alpha = 1$, the distributed FFLS algorithm will degenerate to the distributed LS in [22] which is used to estimate the time-invariant parameter. The quantity $1 - \alpha$ is usually referred to as the speed of adaption. Intuitively, when the parameter process $\{\boldsymbol{\theta}_t\}$ is slowly time-varying, the adaptation speed should also be slow (i.e., α is large). The purpose of this article is to establish the stability of the above diffusion FFLS-based adaptive filter without independence or stationarity assumptions on random regression vector $\{\boldsymbol{\varphi}_{t,i}\}$.

Algorithm 2: Distributed FFLS Algorithm.

Input: $\{\boldsymbol{\varphi}_{t,i}, y_{t+1,i}\}_{i=1}^n$, $t = 0, 1, 2, \cdots$

Output: $\{\hat{\boldsymbol{\theta}}_{t+1,i}\}_{i=1}^n$, $t = 0, 1, 2, \cdots$

Initialization: For each sensor $i \in \{1, \ldots, n\}$, begin with an any initial vector $\hat{\boldsymbol{\theta}}_{0,i}$ and an any initial positive-definite matrix $\boldsymbol{P}_{0,i} > 0$.

for each time $t = 0, 1, 2, \cdots$ **do**

for each sensor $i = 1, \ldots, n$ **do**

Step 1. Adaption (generate $\bar{\boldsymbol{\theta}}_{t+1,i}$ and $\bar{\boldsymbol{P}}_{t+1,i}$ based on $\hat{\boldsymbol{\theta}}_{t,i}$, $\boldsymbol{P}_{t,i}$, $\boldsymbol{\varphi}_{t,i}$ and $y_{t+1,i}$):

$$\bar{\boldsymbol{\theta}}_{t+1,i} = \hat{\boldsymbol{\theta}}_{t,i} + \frac{\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}}{\alpha + \boldsymbol{\varphi}_{t,i}^T\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}}(y_{t+1,i} - \boldsymbol{\varphi}_{t,i}^T\hat{\boldsymbol{\theta}}_{t,i}), \quad (11)$$

$$\bar{\boldsymbol{P}}_{t+1,i} = \frac{1}{\alpha}\left(\boldsymbol{P}_{t,i} - \frac{\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}_{t,i}^T\boldsymbol{P}_{t,i}}{\alpha + \boldsymbol{\varphi}_{t,i}^T\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}}\right), \quad (12)$$

Step 2. Combination (generate $\boldsymbol{P}_{t+1,i}^{-1}$ and $\hat{\boldsymbol{\theta}}_{t+1,i}$ by a convex combination of $\bar{\boldsymbol{\theta}}_{t+1,j}$ and $\bar{\boldsymbol{P}}_{t+1,j}$):

$$\boldsymbol{P}_{t+1,i}^{-1} = \sum_{j\in\mathcal{N}_i}a_{ij}\bar{\boldsymbol{P}}_{t+1,j}^{-1}, \quad (13)$$

$$\hat{\boldsymbol{\theta}}_{t+1,i} = \boldsymbol{P}_{t+1,i}\sum_{j\in\mathcal{N}_i}a_{ij}\bar{\boldsymbol{P}}_{t+1,j}^{-1}\bar{\boldsymbol{\theta}}_{t+1,j}. \quad (14)$$

In order to analyze the distributed FFLS algorithm, we need to derive the estimation error equation. Denote $\widetilde{\boldsymbol{\theta}}_{t,i} \triangleq \boldsymbol{\theta}_t - \hat{\boldsymbol{\theta}}_{t,i}$, then from (13) and (14), we have

$$\begin{aligned}\widetilde{\boldsymbol{\theta}}_{t+1,i} &= \boldsymbol{\theta}_{t+1} - \boldsymbol{P}_{t+1,i}\sum_{j\in\mathcal{N}_i} a_{ij}\bar{\boldsymbol{P}}^{-1}_{t+1,j}\bar{\boldsymbol{\theta}}_{t+1,j}\\ &= \boldsymbol{P}_{t+1,i}\sum_{j\in\mathcal{N}_i} a_{ij}\bar{\boldsymbol{P}}^{-1}_{t+1,j}\boldsymbol{\theta}_{t+1} - \boldsymbol{P}_{t+1,i}\sum_{j\in\mathcal{N}_i} a_{ij}\bar{\boldsymbol{P}}^{-1}_{t+1,j}\bar{\boldsymbol{\theta}}_{t+1,j}\\ &= \boldsymbol{P}_{t+1,i}\sum_{j\in\mathcal{N}_i} a_{ij}\bar{\boldsymbol{P}}^{-1}_{t+1,j}(\boldsymbol{\theta}_{t+1} - \bar{\boldsymbol{\theta}}_{t+1,j}). \qquad (15)\end{aligned}$$

By (1), (2), (11), and (12), we can obtain the following equation:

$$\begin{aligned}&\boldsymbol{\theta}_{t+1} - \bar{\boldsymbol{\theta}}_{t+1,i}\\ &= \boldsymbol{\theta}_t + \Delta\boldsymbol{\theta}_t - \hat{\boldsymbol{\theta}}_{t,i} - \frac{\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}}{\alpha + \boldsymbol{\varphi}^T_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}}(y_{t+1,i} - \boldsymbol{\varphi}^T_{t,i}\hat{\boldsymbol{\theta}}_{t,i})\\ &= \left(\boldsymbol{I}_m - \frac{\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}\boldsymbol{\varphi}^T_{t,i}}{\alpha + \boldsymbol{\varphi}^T_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}}\right)\widetilde{\boldsymbol{\theta}}_{t,i} - \frac{\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}w_{t+1,i}}{\alpha + \boldsymbol{\varphi}^T_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}} + \Delta\boldsymbol{\theta}_t\\ &= \alpha\bar{\boldsymbol{P}}_{t+1,i}\boldsymbol{P}^{-1}_{t,i}\widetilde{\boldsymbol{\theta}}_{t,i} - \frac{\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}w_{t+1,i}}{\alpha + \boldsymbol{\varphi}^T_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}} + \Delta\boldsymbol{\theta}_t. \qquad (16)\end{aligned}$$

For convenience of analysis, we introduce the following set of notations:

$$\begin{aligned}
\boldsymbol{Y}_t &= \mathrm{col}\{y_{t,1},\ldots,y_{t,n}\}, && (n\times 1)\\
\boldsymbol{\Phi}_t &= \mathrm{diag}\{\boldsymbol{\varphi}_{t,1},\ldots,\boldsymbol{\varphi}_{t,n}\}, && (mn\times n)\\
\boldsymbol{W}_t &= \mathrm{col}\{w_{t,1},\ldots,w_{t,n}\}, && (n\times 1)\\
\boldsymbol{P}_t &= \mathrm{diag}\{\boldsymbol{P}_{t,1},\ldots,\boldsymbol{P}_{t,n}\}, && (mn\times mn)\\
\bar{\boldsymbol{P}}_t &= \mathrm{diag}\{\bar{\boldsymbol{P}}_{t,1},\ldots,\bar{\boldsymbol{P}}_{t,n}\}, && (mn\times mn)\\
\boldsymbol{\Theta}_t &= \mathrm{col}\{\underbrace{\boldsymbol{\theta}_t,\ldots,\boldsymbol{\theta}_t}_{n}\}, && (mn\times 1)\\
\Delta\boldsymbol{\Theta}_t &= \mathrm{col}\{\underbrace{\Delta\boldsymbol{\theta}_t,\ldots,\Delta\boldsymbol{\theta}_t}_{n}\}, && (mn\times 1)\\
\boldsymbol{L}_t &= \mathrm{diag}\{\boldsymbol{L}_{t,1},\ldots,\boldsymbol{L}_{t,n}\}, && (mn\times n)\\
&\quad \text{where}\boldsymbol{L}_{t,i} = \frac{\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}}{\alpha + \boldsymbol{\varphi}^T_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}},\\
\widetilde{\boldsymbol{\Theta}}_t &= \mathrm{col}\{\widetilde{\boldsymbol{\theta}}_{t,1},\ldots,\widetilde{\boldsymbol{\theta}}_{t,n}\}, && (mn\times 1)\\
\mathscr{A} &= \mathcal{A}\otimes\boldsymbol{I}_m. && (mn\times mn).
\end{aligned}$$

Hence, by (15) and (16), we have the following equation about estimation error:

$$\begin{aligned}&\widetilde{\boldsymbol{\Theta}}_{t+1}\\ &= \alpha\boldsymbol{P}_{t+1}\mathscr{A}\boldsymbol{P}^{-1}_t\widetilde{\boldsymbol{\Theta}}_t - \boldsymbol{P}_{t+1}\mathscr{A}\bar{\boldsymbol{P}}^{-1}_{t+1}(\boldsymbol{L}_t\boldsymbol{W}_{t+1} + \Delta\boldsymbol{\Theta}_t). \qquad (17)\end{aligned}$$

From (17), we see that the properties of product of random matrices, i.e., $\prod_t \alpha\boldsymbol{P}_{t+1}\mathscr{A}\boldsymbol{P}^{-1}_t$, play important roles in the stability analysis of the homogeneous part in the error equation.

As we all know, the analysis of the product of random matrices is generally a difficult mathematical problem if the random matrices do not satisfy the independence or stationarity assumptions. There is some work to study this problem, which focuses on either symmetric random matrix or scalar gain case. For example, the authors in [16] and[21] investigated the convergence of consensus-diffusion SG algorithm and the stability of the consensus normalized LMS algorithm, where the random matrices in error equations are symmetric. Note that the random matrices $\alpha\boldsymbol{P}_{t+1}\mathscr{A}\boldsymbol{P}^{-1}_t$ here are asymmetric. Although the authors in [23] studied the properties of the asymmetric random matrices in the LMS-based estimation error equation, the adaptive gain of distributed LMS algorithm in [23] is a scalar while the gain $\frac{\boldsymbol{P}_{t,i}}{\alpha + \boldsymbol{\varphi}^T_{t,i}\boldsymbol{P}_{t,i}\boldsymbol{\varphi}_{t,i}}$ in (11) of this article is a random matrix. Hence the methods used in existing literature including [16], [21], and [23] are no longer applicable to our case. One of the main purposes of this article is to overcome the above difficulties by using both the specific structure of the diffusion FFLS and some results of FFLS on the single sensor case (see [33]).

IV. Stability of Distributed FFLS Algorithm Under Fixed Undirected Graph

In this section, we will establish exponential stability for the homogeneous part of the error equation (17) and the tracking error bounds for the proposed distributed FFLS algorithm in Algorithm 2 without requiring statistical independence on the system signals. For this purpose, we need to introduce some definitions on the stability of random matrices (see [33]) and assumptions on the graph and random regression vectors.

A. Some Definitions

Definition 4.1: A random matrix sequence $\{\boldsymbol{A}_t, t\geq 0\}$ defined on the basic probability space $(\Omega,\mathscr{F},P)$ is called L_p-stable $(p>0)$ if $\sup_{t\geq 0}\mathbb{E}(\|\boldsymbol{A}_t\|^p)<\infty$, where $\mathbb{E}(\cdot)$ denotes the mathematical expectation operator. We define $\|\boldsymbol{A}_t\|_{L_p} \triangleq [\mathbb{E}(\|\boldsymbol{A}_t\|^p)]^{\frac{1}{p}}$ as the L_p-norm of the random matrix $\boldsymbol{A}_t$.

Definition 4.2: A sequence of $n\times n$ random matrices $\boldsymbol{A} = \{\boldsymbol{A}_t, t\geq 0\}$ is called L_p-exponentially stable $(p\geq 0)$ with parameter $\lambda\in[0,1)$, if it belongs to the following set:

$$S_p(\lambda) = \left\{\boldsymbol{A} : \left\|\prod_{j=k+1}^{t}\boldsymbol{A}_j\right\|_{L_p} \leq M\lambda^{t-k}\forall t\geq k,\right.$$
$$\left.\forall k\geq 0, \text{for some} M>0\right\}. \qquad (18)$$

As demonstrated by Guo in [33], $\{\boldsymbol{A}_t, t\geq 0\}\in S_p(\lambda)$ is in some sense the necessary and sufficient condition for stability of $\{\boldsymbol{x}_t\}$ generated by $\boldsymbol{x}_t = \boldsymbol{A}_t\boldsymbol{x}_t + \boldsymbol{\xi}_{t+1}, t\geq 0$. Also, the stability analysis of the matrix sequence may be reduced to that of a certain class of scalar sequence, which can be further analyzed based on some excitation conditions on the regressors. To this end, we introduce the following subset of $S_1(\lambda)$ for a scalar sequence $a = (a_t, t \geq 0)$:

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

The definition $S^0(\lambda)$ will be used when we convert the product of a random matrix to that of a scalar sequence.

Remark 4.1: It is clear that if there exists a constant $a_0 \in (0,1)$ such that $a_t \leq a_0$ for all t, then $a_t \in S^0(a_0)$. More properties about the set $S^0(\lambda)$ can be found in [34].

B. Assumptions

Assumption 4.1: The undirected graph $\mathcal{G}$ is connected.

Remark 4.2: For any $k > 1$, we denote $\mathcal{A}^k \triangleq (a_{ij}^{(k)})$ with $\mathcal{A}$ being the weighted adjacency matrix of the graph $\mathcal{G}$, i.e., $a_{ij}^{(k)}$ is the ith row, the jth column element of the matrix $\mathcal{A}^k$ (the kth power of the matrix $\mathcal{A}$). Under Assumption 4.1, it is clear that $\mathcal{A}^k$ is a positive matrix for $k \geq D_{\mathcal{G}}$, which means that $a_{ij}^{(k)} > 0$ for any i and j ([35]).

Assumption 4.2 (Cooperative Excitation Condition): For the adapted sequences $\{\varphi_{t,i}, \mathscr{F}_t, t \geq 0\}$, where $\mathscr{F}_t$ is a sequence of nondecreasing σ-algebras, there exists an integer $h > 0$ such that $\{1-\lambda_t\} \in S^0(\lambda)$ for some $\lambda \in (0,1)$, where λ_t is defined by

$$\lambda_t \triangleq \lambda_{\min}\left[\mathbb{E}\left(\frac{1}{n(1+h)}\sum_{i=1}^{n}\sum_{k=th+1}^{(t+1)h}\frac{\varphi_{k,i}\varphi_{k,i}^T}{1+\|\varphi_{k,i}\|^2}\Big|\mathscr{F}_{th}\right)\right]$$

with $\mathbb{E}(\cdot|\cdot)$ being the conditional mathematical expectation operator.

Remark 4.3: Assumption 4.2 is also used to guarantee the stability and performance of the distributed LMS algorithm (see [16] and [23]). We give some intuitive explanations for the above cooperative excitation condition about the following two aspects.

1) *"Why excitation:"* Let us consider an extreme case where all regression vectors $\varphi_{k,i}$ are equal to zero, then Assumption 4.2 cannot be satisfied. Moreover, from (1), we see that the unknown parameter θ_t cannot be estimated or tracked since the observations $y_{t,i}$ do not contain any information about the unknown parameter θ_t. In order to estimate θ_t, some nonzero information condition (named excitation condition) should be imposed on the regression vectors $\varphi_{t,i}$. In fact, Assumption 4.2 intuitively gives a lower bound (which may be changed over time) of the sequence $\{\lambda_t\}$. For the ϕ-mixing and bounded regressor sequence $\{\varphi_{t,i}\}$ ([16]), Assumption 4.2 can be equivalently written as a cleaner cooperative excitation condition

$$\inf_t \lambda_{\min}\left[\mathbb{E}\left(\sum_{i=1}^{n}\sum_{k=th+1}^{(t+1)h}\varphi_{k,i}\varphi_{k,i}^T\right)\right] > 0$$

which is actually a stochastic version of deterministic cooperative conditions ([9]). Moreover, for the typical case where the regressor sequence $\{\varphi_{k,i}\}$ is i.i.d. and bounded with positive-definite covariance matrix, Assumption 4.2 can be easily verified.

2) *"Why cooperative:"* Compared with the excitation condition for the FFLS algorithm of the single sensor case in [33], i.e., there exists a constant $h > 0$ such that

$$\{1-\lambda'_t, t \geq 0\} \in S^0(\lambda') \tag{19}$$

for some λ' where

$$\lambda'_t = \lambda_{\min}\left[\mathbb{E}\left(\frac{1}{1+h}\sum_{k=th+1}^{(t+1)h}\frac{\varphi_{k,i}\varphi_{k,i}^T}{1+\|\varphi_{k,i}\|^2}\Big|\mathscr{F}_{th}\right)\right].$$

Assumption 4.2 contains not only temporal union information but also spatial union information of all the sensors, which means that Assumption 4.2 is much weaker than the condition (19) since $\lambda_t \geq \lambda'_t$ when $n > 1$. Besides, we also note that Assumption 4.2 can be reduced to the condition (19) when $n = 1$. In fact, Assumption 4.2 can reflect the cooperative effect of multiple sensors in the sense that the estimation task can be still fulfilled by the cooperation of multiple sensors even if any of them cannot (see Example 6.1 in Section VI).

C. Main Results

In order to establish exponential stability of the product of random matrices $\alpha \boldsymbol{P}_{t+1}\mathscr{A}\boldsymbol{P}_t^{-1}$, we first analyze the properties of the random matrix $\boldsymbol{P}_t$ to obtain its upper bound.

Lemma 4.1: For $\{\boldsymbol{P}_t\}$ generated by (12) and (13), under Assumptions 4.1–4.2, we have

$$T_{t+1} \leq \frac{1}{\alpha^{h'}}(1-\beta_{t+1})(h'-D_{\mathcal{G}})tr(\boldsymbol{P}_{th'+1}) \tag{20}$$

where

$$T_t \triangleq \sum_{k=(t-1)h'+D_{\mathcal{G}}+1}^{th'} tr(\boldsymbol{P}_{k+1}), T_0 = 0$$

$$\beta_{t+1} \triangleq \frac{a_{\min}^2\gamma_{t+1}}{n(h'-D_{\mathcal{G}})\left(\alpha^{h'}+\lambda_{\max}\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)\right)tr(\boldsymbol{P}_{th'+1})}$$

$$\gamma_{t+1} \triangleq tr\left(\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)^2\sum_{k=th'+D_{\mathcal{G}}+1}^{(t+1)h'}\sum_{j=1}^{n}\frac{\varphi_{k,j}\varphi_{k,j}^T}{(1+\|\varphi_{k,j}\|^2)}\right)$$

$$a_{\min} \triangleq \min_{i,j\in\{1,\ldots,n\}} a_{ij}^{(D_{\mathcal{G}})} > 0$$

$$h' \triangleq 2h + D_{\mathcal{G}}$$

and h is given by Assumption 4.2.

Proof: Note that $a_{ij}^{(k)}$ is the ith row, the jth column element of the matrix $\mathcal{A}^k$, $k \geq 1$, where $a_{ij}^{(1)} = a_{ij}$. By (10), we have $\boldsymbol{P}_{k+1,i}^{-1} \geq \sum_{j=1}^{n} a_{ij}\alpha\boldsymbol{P}_{k,j}^{-1}$. Hence, by the inequality

$$\left(\sum_{j=1}^{n} a_{ij}\boldsymbol{A}_j\right)^{-1} \leq \sum_{j=1}^{n} a_{ij}\boldsymbol{A}_j^{-1} \tag{21}$$

with $\boldsymbol{A}_j \geq 0$ [36], we obtain for any $t \geq 0$, and any $k \in [th' + D_{\mathcal{G}} + 1, (t+1)h']$

$$\begin{aligned}\boldsymbol{P}_{k,i} &\leq \left(\sum_{j=1}^{n} a_{ij}\alpha \boldsymbol{P}_{k-1,j}^{-1}\right)^{-1} \leq \frac{1}{\alpha}\sum_{j=1}^{n} a_{ij}\boldsymbol{P}_{k-1,j} \\ &\leq \frac{1}{\alpha}\sum_{j=1}^{n} a_{ij}\left(\frac{1}{\alpha}\sum_{l=1}^{n} a_{jl}\boldsymbol{P}_{k-2,l}\right) \leq \cdots \\ &\leq \frac{1}{\alpha^{k-th'-1}}\sum_{j=1}^{n} a_{ij}^{(k-th'-1)}\boldsymbol{P}_{th'+1,j} \\ &\leq \frac{1}{\alpha^{h'-1}}\sum_{j=1}^{n} a_{ij}^{(k-th'-1)}\boldsymbol{P}_{th'+1,j}. \end{aligned} \tag{22}$$

Denote $\boldsymbol{Q}_i^{k,th'} = \sum_{j=1}^{n} a_{ij}^{(k-th'-1)}\boldsymbol{P}_{th'+1,j}$. Then, by (10), (13), (21), and (22), we have for $k \in [th' + D_{\mathcal{G}} + 1, (t+1)h']$

$$\begin{aligned}\boldsymbol{P}_{k+1,i} &= \left(\sum_{j=1}^{n} a_{ij}(\alpha\boldsymbol{P}_{k,j}^{-1} + \boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T)\right)^{-1} \\ &\leq \sum_{j=1}^{n} a_{ij}(\alpha\boldsymbol{P}_{k,j}^{-1} + \boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T)^{-1} \\ &\leq \sum_{j=1}^{n} a_{ij}\left(\alpha\left(\frac{1}{\alpha^{h'-1}}\boldsymbol{Q}_j^{k,th'}\right)^{-1} + \boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T\right)^{-1}. \end{aligned} \tag{23}$$

By Lemma 2.1 and (23), it follows that:

$$\begin{aligned}\boldsymbol{P}_{k+1,i} &\leq \frac{1}{\alpha^{h'}}\sum_{j=1}^{n} a_{ij}\left(\boldsymbol{Q}_j^{k,th'} - \frac{\boldsymbol{Q}_j^{k,th'}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T\boldsymbol{Q}_j^{k,th'}}{\alpha^{h'} + \boldsymbol{\varphi}_{k,j}^T\boldsymbol{Q}_j^{k,th'}\boldsymbol{\varphi}_{k,j}}\right) \\ &= \frac{1}{\alpha^{h'}}\sum_{j=1}^{n} a_{ij}^{(k-th')}\boldsymbol{P}_{th'+1,j} \\ &\quad - \frac{1}{\alpha^{h'}}\sum_{j=1}^{n} a_{ij}\frac{\boldsymbol{Q}_j^{k,th'}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T\boldsymbol{Q}_j^{k,th'}}{\alpha^{h'} + \boldsymbol{\varphi}_{k,j}^T\boldsymbol{Q}_j^{k,th'}\boldsymbol{\varphi}_{k,j}} \\ &\leq \frac{1}{\alpha^{h'}}\sum_{j=1}^{n} a_{ij}^{(k-th')}\boldsymbol{P}_{th'+1,j} - \\ &\quad \frac{1}{\alpha^{h'}}\sum_{j=1}^{n}\frac{a_{ij}\boldsymbol{Q}_j^{k,th'}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T\boldsymbol{Q}_j^{k,th'}}{\alpha^{h'} + \lambda_{\max}(\boldsymbol{Q}_j^{k,th'})(1 + \|\boldsymbol{\varphi}_{k,j}\|^2)}. \end{aligned} \tag{24}$$

Then, by (24), we have

$$\begin{aligned} tr(\boldsymbol{P}_{k+1}) &= tr\left(\sum_{i=1}^{n}\boldsymbol{P}_{k+1,i}\right) \\ &\leq \frac{1}{\alpha^{h'}}tr\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^{(k-th')}\boldsymbol{P}_{th'+1,j}\right) \\ &\quad - \frac{1}{\alpha^{h'}}tr\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\frac{\boldsymbol{Q}_j^{k,th'}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T\boldsymbol{Q}_j^{k,th'}}{\alpha^{h'} + \lambda_{\max}(\boldsymbol{Q}_j^{k,th'})(1 + \|\boldsymbol{\varphi}_{k,j}\|^2)}\right) \\ &= \frac{1}{\alpha^{h'}}\left(tr(\boldsymbol{P}_{th'+1}) - \sum_{j=1}^{n}\frac{tr\left(\boldsymbol{Q}_j^{k,th'}\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T\boldsymbol{Q}_j^{k,th'}\right)}{\alpha^{h'} + \lambda_{\max}(\boldsymbol{Q}_j^{k,th'})(1 + \|\boldsymbol{\varphi}_{k,j}\|^2)}\right). \end{aligned}$$

Hence, combining this with the inequality $\sum_{j=1}^{n}\frac{a_j}{b_j} \geq \frac{\sum_{j=1}^{n} a_j}{\sum_{j=1}^{n} b_j}$, where $a_j \geq 0$ and $b_j \geq 0$, we obtain that

$$\begin{aligned} &tr(\boldsymbol{P}_{k+1}) \\ &\leq \frac{1}{\alpha^{h'}}\left(tr(\boldsymbol{P}_{th'+1}) - \frac{tr\left(\sum_{j=1}^{n}\left(\boldsymbol{Q}_j^{k,th'}\right)^2\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T}{(1+\|\boldsymbol{\varphi}_{k,j}\|^2)}\right)}{\sum_{j=1}^{n}\left(\alpha^{h'} + \lambda_{\max}\left(\boldsymbol{Q}_j^{k,th'}\right)\right)}\right). \end{aligned} \tag{25}$$

By Remark 4.2, we know that $a_{ij}^{(k)} \geq a_{\min}$ holds for all $k \geq D_{\mathcal{G}}$. Thus, by (25), we have for $k \in [th' + D_{\mathcal{G}} + 1, (t+1)h']$

$$\begin{aligned} tr(\boldsymbol{P}_{k+1}) &\leq \frac{1}{\alpha^{h'}}\Bigg(tr(\boldsymbol{P}_{th'+1}) \\ &\quad - \frac{a_{\min}^2 tr\left(\sum_{j=1}^{n}\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)^2\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T}{(1+\|\boldsymbol{\varphi}_{k,j}\|^2)}\right)}{n\left(\alpha^{h'} + \lambda_{\max}\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)\right)}\Bigg). \end{aligned} \tag{26}$$

Summing up both sides of (26) from $th' + D_{\mathcal{G}} + 1$ to $(t+1)h'$, by the definition of β_{t+1}, we have

$$\begin{aligned} T_{t+1} &= \sum_{k=th'+D_{\mathcal{G}}+1}^{(t+1)h'} tr(\boldsymbol{P}_{k+1}) \\ &\leq \frac{1}{\alpha^{h'}}(1 - \beta_{t+1})(h' - D_{\mathcal{G}})tr(\boldsymbol{P}_{th'+1}). \end{aligned}$$

This completes the proof of the lemma. ■

Before giving the boundness of the random matrix $\boldsymbol{P}_t$, we first introduce two lemmas in [33].

Lemma 4.2 ([33]): Let $\{1 - \xi_t\} \in S^0(\lambda)$, and $0 < \xi_t \leq \xi^* < 1$, where ξ^* is a positive constant. Then, for any $\varepsilon \in (0,1)$, $\{1 - \varepsilon\xi_t\} \in S^0(\lambda^{(1-\xi^*)\varepsilon})$.

Lemma 4.3 ([33]): Let $\{x_t, \mathscr{F}_t\}$ be an adapted process, and $x_{t+1} \leq \xi_{t+1}x_t + \eta_{t+1}, t \geq 0, \mathbb{E}x_0^2 < \infty$, where $\{\xi_t, \mathscr{F}_t\}$ and $\{\eta_t, \mathscr{F}_t\}$ are two adapted nonnegative process with properties: $\xi_t \geq \varepsilon_0 > 0 \forall t; \mathbb{E}(\eta_{t+1}^2|\mathscr{F}_t) \leq N < \infty \forall t; \|\prod_{k=j}^{t}\mathbb{E}(\xi_{k+1}^4|\mathscr{F}_k)\| \leq M\eta^{t-j+1}\forall t \geq j \forall j,$ where ε_0, M, N and $\eta \in (0,1)$ are constants. Then, we have

$$(i) \qquad \left\|\prod_{k=j}^{t}\xi_k\right\|_{L_2} \leq M^{\frac{1}{4}}\eta^{\frac{1}{4}(t-j+1)}\forall t \geq j \forall j;$$

$$(ii) \qquad \sup_t \mathbb{E}(\|x_t\|) < \infty.$$

The following lemma proves the boundedness of the random matrix sequence $\{\boldsymbol{P}_t\}$.

Lemma 4.4: For $\{\boldsymbol{P}_t\}$ generated by (12) and (13), under Assumptions 4.1–4.2, we have for any $p \geq 1$, $\boldsymbol{P}_t$ is L_p stable, i.e.,

$$\sup_{t\geq 0}\mathbb{E}(\|\boldsymbol{P}_t\|^p)<\infty$$

provided that $\lambda^{\frac{a_{\min}^2}{32pmh(4h+D_{\mathcal{G}}-1)}}<\alpha<1$, where λ and h are given by Assumption 4.2, and m is the dimension of $\boldsymbol{\varphi}_{t,i}$.

Proof: For any $t\geq 0$, there exists an integer $z_t=\lfloor\frac{th'+D_{\mathcal{G}}}{h}\rfloor+1$ such that

$$(z_t-1)h\leq(th'+D_{\mathcal{G}}+1)\leq z_th+1. \tag{27}$$

By the definition of β_{t+1} in Lemma 4.1, it is clear that

$$\begin{aligned}&\beta_{t+1}\\ &\geq\frac{a_{\min}^2tr\left(\left(\sum_{l=1}^n\boldsymbol{P}_{th'+1,l}\right)^2\sum_{k=z_th+1}^{(z_t+1)h}\sum_{j=1}^n\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T}{(1+\|\boldsymbol{\varphi}_{k,j}\|^2)}\right)}{n(h'-D_{\mathcal{G}})\left(\alpha^{h'}+\lambda_{\max}\left(\sum_{l=1}^n\boldsymbol{P}_{th'+1,l}\right)\right)tr(\boldsymbol{P}_{th'+1})}\\ &\triangleq b_{t+1}.\end{aligned}\tag{28}$$

Hence by Lemma 4.1 and (28), we obtain

$$T_{t+1}\leq\frac{1}{\alpha^{h'}}(1-b_{t+1})(h'-D_{\mathcal{G}})tr(\boldsymbol{P}_{th'+1}). \tag{29}$$

By the inequality $\boldsymbol{P}_{k,i}\leq\frac{1}{\alpha}\sum_{j=1}^n a_{ij}\boldsymbol{P}_{k-1,j}$ used in (22) it follows that

$$\begin{aligned}(h'-D_{\mathcal{G}})tr(\boldsymbol{P}_{th'+1})&=\sum_{k=(t-1)h'+D_{\mathcal{G}}+1}^{th'}tr(\boldsymbol{P}_{th'+1})\\ &=\sum_{k=(t-1)h'+D_{\mathcal{G}}+1}^{th'}\sum_{i=1}^n tr(\boldsymbol{P}_{th'+1,i})\\ &\leq\sum_{k=(t-1)h'+D_{\mathcal{G}}+1}^{th'}\sum_{i=1}^n tr\left(\frac{1}{\alpha^{th'-k}}\sum_{j=1}^n a_{ij}^{(th'-k)}\boldsymbol{P}_{k+1,j}\right)\\ &\leq\frac{1}{\alpha^{h'-D_{\mathcal{G}}-1}}\sum_{k=(t-1)h'+D_{\mathcal{G}}+1}^{th'}tr(\boldsymbol{P}_{k+1})=\frac{1}{\alpha^{h'-D_{\mathcal{G}}-1}}T_t.\end{aligned}$$

Hence, by (29), we have

$$T_{t+1}\leq\frac{1}{\alpha^{2h'-D_{\mathcal{G}}-1}}(1-b_{t+1})T_t. \tag{30}$$

For $p\geq 1$, denote

$$c_{t+1}=\frac{1}{\alpha^{p(2h'-D_{\mathcal{G}}-1)}}\left(1-\frac{b_{t+1}}{2}\right)I_{\{tr(\boldsymbol{P}_{th'+1})\geq 1\}} \tag{31}$$

where $I_{\{\cdot\}}$ denotes the indicator function, whose value is 1 if its argument (a formula) is true, and 0, otherwise. Then, by (29) and (30), we have

$$\begin{aligned}T_{t+1}^p&\leq T_{t+1}^p\left(I_{\{tr(\boldsymbol{P}_{th'+1})\geq 1\}}+I_{\{tr(\boldsymbol{P}_{th'+1})\leq 1\}}\right)\\ &\leq\frac{1}{\alpha^{p(2h'-D_{\mathcal{G}}-1)}}(1-b_{z_t+1})^pT_t^pI_{\{tr(\boldsymbol{P}_{th'+1})\geq 1\}}\\ &\quad+T_{t+1}^pI_{\{tr(\boldsymbol{P}_{th'+1})\leq 1\}}\\ &\leq c_{t+1}T_t^p+\frac{1}{\alpha^{ph'}}(h'-D_{\mathcal{G}})^p.\end{aligned}\tag{32}$$

Denote

$$\boldsymbol{H}_{z_t}=\mathbb{E}\left(\sum_{k=z_th+1}^{(z_t+1)h}\sum_{j=1}^n\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\middle|\mathscr{F}_{z_th}\right).$$

By the fact $\boldsymbol{P}_{th'+1,l}\in\mathscr{F}_{th'}\subset\mathscr{F}_{z_th}$ and the inequality $tr((\sum_{l=1}^n\boldsymbol{P}_{th'+1,l})^2)\geq m^{-1}(tr(\sum_{l=1}^n\boldsymbol{P}_{th'+1,l}))^2$, from the definition of b_{t+1} in (28), we can conclude the following inequality:

$$\begin{aligned}&\mathbb{E}(b_{t+1}|\mathscr{F}_{z_th})\\ &=\frac{a_{\min}^2tr\left[\left(\sum_{l=1}^n\boldsymbol{P}_{th'+1,l}\right)^2\boldsymbol{H}_{z_t}\right]}{n(h'-D_{\mathcal{G}})\left(\alpha^{h'}+\lambda_{\max}\left(\sum_{l=1}^n\boldsymbol{P}_{th'+1,l}\right)\right)tr(\boldsymbol{P}_{th'+1})}\\ &\geq\frac{a_{\min}^2\left(tr(\boldsymbol{P}_{th'+1})\right)^2\lambda_{\min}(\boldsymbol{H}_{z_t})}{mn(h'-D_{\mathcal{G}})\left(\alpha^{h'}+\lambda_{\max}\left(\sum_{l=1}^n\boldsymbol{P}_{th'+1,l}\right)\right)tr(\boldsymbol{P}_{th'+1})}\\ &\geq\frac{a_{\min}^2\left(tr(\boldsymbol{P}_{th'+1})\right)\lambda_{z_t}(1+h)}{m(h'-D_{\mathcal{G}})\left(\alpha^{h'}+\lambda_{\max}\left(\sum_{l=1}^n\boldsymbol{P}_{th'+1,l}\right)\right)}\\ &\geq\frac{a_{\min}^2\left(tr(\boldsymbol{P}_{th'+1})\right)\lambda_{z_t}(1+h)}{m(h'-D_{\mathcal{G}})\left(1+tr(\boldsymbol{P}_{th'+1})\right)}\\ &\geq\frac{a_{\min}^2\lambda_{z_t}(1+h)}{2m(h'-D_{\mathcal{G}})}\text{on}\{tr(\boldsymbol{P}_{th'+1})\geq 1\}.\end{aligned}\tag{33}$$

Hence, by the definition of c_{t+1} in (31) and (33)

$$\begin{aligned}&\mathbb{E}(c_{t+1}|\mathscr{F}_{z_th})\\ &=\frac{1}{\alpha^{p(2h'-D_{\mathcal{G}}-1)}}\left(1-\frac{\mathbb{E}(b_{t+1}|\mathscr{F}_{z_th})}{2}\right)I_{\{tr(\boldsymbol{P}_{th'+1})\geq 1\}}\\ &\leq\frac{1}{\alpha^{p(2h'-D_{\mathcal{G}}-1)}}\left(1-\frac{a_{\min}^2\lambda_{z_t}(1+h)}{4m(h'-D_{\mathcal{G}})}\right)I_{\{tr(\boldsymbol{P}_{th'+1})\geq 1\}}.\end{aligned}\tag{34}$$

Denote

$$d_{t+1}=\begin{cases}c_{t+1}, & tr(\boldsymbol{P}_{th'+1})\geq 1;\\ \frac{1}{\alpha^{p(2h'-D_{\mathcal{G}}-1)}}\left(1-\frac{a_{\min}^2\lambda_{z_t}(1+h)}{4m(h'-D_{\mathcal{G}})}\right), & \text{otherwise.}\end{cases}$$

Then, by (32) and (34), we have

$$T_{t+1}^p\leq d_{t+1}T_t^p+\frac{1}{\alpha^{ph'}}(h'-D_{\mathcal{G}})^p. \tag{35}$$

Since $\lambda_{z_t}\leq\frac{h}{1+h}$ and $b_{t+1}\leq\frac{a_{\min}^2h}{h'-D_{\mathcal{G}}}$, we know that $d_{t+1}\geq\varepsilon_0$ with ε_0 being a positive constant. Denote $\mathscr{B}_t\triangleq\mathscr{F}_{z_th}$, then by the definition of z_t, it is clear that $z_{t+1}\geq z_t+2$. Thus, we obtain that $d_{t+1}\in\mathscr{F}_{(z_t+1)h}\subset\mathscr{B}_{t+1}$. Similar to the analysis of (34), we have

$$\mathbb{E}(c_{t+1}^4|\mathscr{B}_t)\leq\frac{1}{\alpha^{4p(2h'-D_{\mathcal{G}}-1)}}\left(1-\frac{a_{\min}^2\lambda_{z_t}(1+h)}{4m(h'-D_{\mathcal{G}})}\right). \tag{36}$$

Hence, by the definition of d_{t+1}, it follows that

$$\left\|\prod_{k=j}^t\mathbb{E}(d_{k+1}^4|\mathscr{B}_k)\right\|_{L_1}$$

$$\leq \Big\| \prod_{k=j}^{t} \left(\frac{1}{\alpha^{4p(2h'-D_{\mathcal{G}}-1)}} \left(1 - \frac{a_{\min}^2 \lambda_{z_k}(1+h)}{8mh} \right) \right) \Big\|_{L_1}. \tag{37}$$

By Assumption 4.2 and the fact $\lambda_{z_k} \leq \frac{h}{1+h}$, applying Lemma 4.2, we obtain $\{1 - \frac{a_{\min}^2 \lambda_{z_k}(1+h)}{8mh}\} \in S^0(\lambda^{\frac{a_{\min}^2}{8mh}})$. By (37), we see that there exists a positive constant N such that

$$\Big\| \prod_{k=j}^{t} \mathbb{E}(d_{k+1}^4 | \mathscr{B}_k) \Big\|_{L_1} \leq N \lambda_1^{t-j+1}$$

where $\lambda_1 = \frac{1}{\alpha^{4p(2h'-D_{\mathcal{G}}-1)}} \lambda^{\frac{a_{\min}^2}{8mh}} \in (0,1)$. Furthermore, by Lemma 4.3, we have $\sup_t \mathbb{E}(T_t^p) < \infty$, which implies that $\sup_{t\geq 0} \mathbb{E}(\|\boldsymbol{P}_t\|^p) < \infty$. This completes the proof. ■

We then establish the exponential stability of the homogeneous part of the error equation (17).

Theorem 4.1: Consider the distributed FFLS algorithm in Algorithm 2. If the forgetting factor α satisfies $\lambda^{\frac{a_{\min}^2}{32pmh(4h+D_{\mathcal{G}}-1)}} < \alpha < 1$ and for any $i \in \{1, \ldots, n\}$, $\sup_t \|\boldsymbol{\varphi}_{t,i}\|_{L_{6p}} < \infty$, then under Assumptions 4.1 and 4.2, for any $p \geq 1$, $\{\alpha \boldsymbol{P}_{t+1} \mathscr{A} \boldsymbol{P}_t^{-1}\}$ is L_p-exponentially stable.

Proof: By (10) and (13), we have

$$\boldsymbol{P}_{t+1,i}^{-1} = \sum_{j=1}^{n} a_{ij} (\alpha \boldsymbol{P}_{t,j}^{-1} + \boldsymbol{\varphi}_{t,j} \boldsymbol{\varphi}_{t,j}^T).$$

Then, we can obtain the following equation:

$$\begin{aligned} tr(\boldsymbol{P}_{t+1}^{-1}) &= tr\left(\sum_{i=1}^{n} \boldsymbol{P}_{t+1,i}^{-1} \right) \\ &= tr\left(\sum_{j=1}^{n} (\alpha \boldsymbol{P}_{t,j}^{-1} + \boldsymbol{\varphi}_{t,j} \boldsymbol{\varphi}_{t,j}^T) \right) \\ &= \alpha tr(\boldsymbol{P}_t^{-1}) + \sum_{j=1}^{n} \|\boldsymbol{\varphi}_{t,j}\|^2. \end{aligned}$$

By the Mikowski inequality, it follows that

$$\begin{aligned} \|tr(\boldsymbol{P}_{t+1}^{-1})\|_{L_{3p}} &\leq \alpha \|tr(\boldsymbol{P}_t^{-1})\|_{L_{3p}} + O\left(\sum_{j=1}^{n} \|\boldsymbol{\varphi}_{t,j}\|_{L_{6p}}^2 \right) \\ &= \alpha^{t+1} \|tr(\boldsymbol{P}_0^{-1})\|_{L_{3p}} + O\left(\sum_{k=0}^{t} \alpha^j \right). \end{aligned}$$

Hence, we have

$$\sup_t \|\boldsymbol{P}_{t+1}^{-1}\|_{L_{3p}} < \infty. \tag{38}$$

By Lemma 4.4, we derive that

$$\begin{aligned} \Big\| \prod_{k=j}^{t} \alpha \boldsymbol{P}_{k+1} \mathscr{A} \boldsymbol{P}_k^{-1} \Big\|_{L_p} &= \left[\mathbb{E}\left(\Big\| \prod_{k=j}^{t} \alpha \boldsymbol{P}_{k+1} \mathscr{A} \boldsymbol{P}_k^{-1} \Big\|^p \right) \right]^{\frac{1}{p}} \\ &= \left[\mathbb{E}\left(\|\alpha^{t-j+1} \boldsymbol{P}_{t+1} \mathscr{A}^{t-j+1} \boldsymbol{P}_j^{-1}\|^p \right) \right]^{\frac{1}{p}} \\ &\leq \alpha^{t-j+1} \|\boldsymbol{P}_{t+1}\|_{L_{2p}} \|\boldsymbol{P}_j^{-1}\|_{L_{2p}} = O(\alpha^{t-j+1}). \end{aligned}$$

This completes the proof of the theorem. ■

Based on Theorem 4.1, we further establish the tracking error bound of Algorithm 2 under some conditions on the noises and parameter variation.

Theorem 4.2: Consider the model (1) and the error equation (17). Under the conditions of Theorem 4.1, if for some $p \geq 1$, $\sigma_{3p} \triangleq \sup_t (\|\boldsymbol{W}_t\|_{L_{3p}} + \|\Delta \boldsymbol{\Theta}_t\|_{L_{3p}}) < \infty$, then there exists a constant c such that

$$\limsup_{t\to\infty} \|\widetilde{\boldsymbol{\Theta}}_t\|_{L_p} \leq c \sigma_{3p}.$$

Proof: For convenience of analysis, let the state transition matrix $\boldsymbol{\Psi}(t,k)$ be recursively defined by

$$\boldsymbol{\Psi}(t+1,k) = \alpha \boldsymbol{P}_{t+1} \mathscr{A} \boldsymbol{P}_t^{-1} \boldsymbol{\Psi}(t,k), \boldsymbol{\Psi}(k,k) = \boldsymbol{I}_{mn}. \tag{39}$$

It is clear that $\boldsymbol{\Psi}(t+1,k) = \alpha^{t-k+1} \boldsymbol{P}_{t+1} \mathscr{A}^{t-k+1} \boldsymbol{P}_k^{-1}$. From the definition of $\boldsymbol{L}_t$ and (10), we have $\bar{\boldsymbol{P}}_{t+1}^{-1} \boldsymbol{L}_t = \boldsymbol{\Phi}_t$. Then, by (17), we have

$$\widetilde{\boldsymbol{\Theta}}_{t+1} = \alpha \boldsymbol{P}_{t+1} \mathscr{A} \boldsymbol{P}_t^{-1} \widetilde{\boldsymbol{\Theta}}_t - \boldsymbol{P}_{t+1} \mathscr{A} (\boldsymbol{\Phi}_t \boldsymbol{W}_{t+1} + \bar{\boldsymbol{P}}_{t+1}^{-1} \Delta \boldsymbol{\Theta}_t).$$

Hence, by the Hölder inequality, we have

$$\begin{aligned} \|\widetilde{\boldsymbol{\Theta}}_{t+1}\|_{L_p} &= \Big\| \boldsymbol{\Psi}(t+1,0) \widetilde{\boldsymbol{\Theta}}_0 \\ &\quad - \sum_{k=0}^{t} \boldsymbol{\Psi}(t+1,k+1) (\boldsymbol{P}_{k+1} \mathscr{A} (\boldsymbol{\Phi}_k \boldsymbol{W}_{k+1} + \bar{\boldsymbol{P}}_{k+1}^{-1} \Delta \boldsymbol{\Theta}_k)) \Big\|_{L_p} \\ &\leq \|\alpha^{t+1} \boldsymbol{P}_{t+1} \mathscr{A}^{t+1} \boldsymbol{P}_0^{-1} \widetilde{\boldsymbol{\Theta}}_0\|_{L_p} \\ &\quad + \Big\| \sum_{k=0}^{t} \alpha^{t-k} \boldsymbol{P}_{t+1} \mathscr{A}^{t-k+1} (\boldsymbol{\Phi}_k \boldsymbol{W}_{k+1} + \bar{\boldsymbol{P}}_{k+1}^{-1} \Delta \boldsymbol{\Theta}_k) \Big\|_{L_p} \\ &\leq O(\alpha^{t+1} \|\boldsymbol{P}_{t+1}\|_{L_{2p}}) \\ &\quad + \sum_{k=0}^{t} \alpha^{t-k} \|\boldsymbol{P}_{t+1}\|_{L_{3p}} \|\boldsymbol{\Phi}_k\|_{L_{3p}} \|\boldsymbol{W}_{k+1}\|_{L_{3p}} \\ &\quad + \sum_{k=0}^{t} \alpha^{t-k} \|\boldsymbol{P}_{t+1}\|_{L_{3p}} \|\bar{\boldsymbol{P}}_{k+1}^{-1}\|_{L_{3p}} \|\Delta \boldsymbol{\Theta}_k\|_{L_{3p}}. \end{aligned}$$

Hence, by Lemma 4.4 and (38), it follows that $\limsup_{t\to\infty} \|\widetilde{\boldsymbol{\Theta}}_t\|_{L_p} \leq c\sigma_{3p}$, where c is a positive constant depending on α and the upper bounds of $\{\boldsymbol{P}_t\}$, $\{\boldsymbol{\Phi}_t\}$, and $\{\boldsymbol{P}_t^{-1}\}$. This completes the proof. ■

Remark 4.4: Following the proof line of Theorems 4.1 and 4.2, we can see that if the forgetting factor α is selected to be uncoordinated for different sensors, i.e., we replace α with α_i in Algorithm 2, the key inequalities (22) and (23) still hold by replacing α with $\alpha_{\min} \triangleq \min\{\alpha_1, \ldots, \alpha_n\}$. Based on this, the results of Theorems 4.1 and 4.2 also hold only if the condition $\lambda^{\frac{a_{\min}^2}{32pmh(4h+D_{\mathcal{G}}-1)}} < \alpha$ is replaced with $\lambda^{\frac{a_{\min}^2}{32pmh(4h+D_{\mathcal{G}}-1)}} < \alpha_{\min}$.

V. Stability of Distributed FFLS Algorithm Over Unreliable Directed Networks

In Section IV, we have studied the stability of the distributed FFLS algorithm under the fixed undirected graph. However, in practical engineering applications, the information exchange between sensors might not be bidirectional. Moreover, it is often interfered by many uncertain random factors due to the distance, obstacle, and interference, which will lead to the interruption or reconstruction of communication links. Thus, in this section, we model the communication links between sensors as time-varying random switching directed communication topologies $\mathcal{G}_{r(t)} = (\mathcal{V}, \mathcal{E}_{r(t)}, \mathcal{A}_{r(t)})$. The switching process is governed by a homogeneous Markov chain $r(t)$ whose states belong to a finite set $\mathbb{S} = \{1, 2, \ldots, s\}$, and the corresponding set of communication topology graph is denoted by $\mathcal{C} = \{\mathcal{G}_1, \ldots, \mathcal{G}_s\}$. The communication graph is switched just at the instant that the value of $r(t)$ is changed. Thus, the corresponding adjacency matrix and the neighbor set of the sensor i are denoted as $\mathcal{A}_{r(t)} = [a_{ij,r(t)}]_{1\le i,j\le n}$ and $\mathcal{N}_{i,r(t)}$, respectively. For the distributed FFLS algorithm over the Markovian switching directed topologies, we just modify Step 2 in Algorithm 2 as follows:

$$\boldsymbol{P}_{t+1,i}^{-1} = \sum_{j\in\mathcal{N}_{i,r(t)}} a_{ji,r(t)} \bar{\boldsymbol{P}}_{t+1,j}^{-1}, \tag{40}$$

$$\hat{\boldsymbol{\theta}}_{t+1,i} = \boldsymbol{P}_{t+1,i} \sum_{j\in\mathcal{N}_{i,r(t)}} a_{ji,r(t)} \bar{\boldsymbol{P}}_{t+1,j}^{-1} \bar{\boldsymbol{\theta}}_{t+1,j}. \tag{41}$$

To analyze the stability of algorithm (11), (12), (40), and (41), we introduce the following assumptions.

Assumption 5.1: All possible digraphs $\{\mathcal{G}_1, \ldots, \mathcal{G}_s\}$ are balanced and the union of all those digraphs, i.e, $\cup_{j=1}^s \mathcal{G}_j = (\mathcal{V}, \cup_{j=1}^s \mathcal{E}_j, \frac{1}{s}\sum_{j=1}^s \mathcal{A}_j)$ is strongly connected.

Assumption 5.2: The Markov chain $\{r(t), t\ge 0\}$ is irreducible and aperiodic with the transition probability matrix $\boldsymbol{P} = [p_{ij}]_{1\le i,j\le s}$ where $p_{ij} = \Pr(r(t+1) = j | r(t) = i)$ with $\Pr(\cdot|\cdot)$ being the conditional probability.

According to the Markov chain theory [37], a discrete-time homogeneous Markov chain with finite states is ergodic if and only if it is irreducible and aperiodic. Hence, Assumption 5.2 means that the l-step transition matrix $\boldsymbol{P}^l$ has a limit with identical rows.

In the following, we will analyze the properties of the strongly connected directed graph. For convenience, we denote the ith row, the jth column element of the matrix $\boldsymbol{A}$ as $\boldsymbol{A}(i,j)$.

Lemma 5.1: Let $\mathcal{G}_k = (\mathcal{V}, \mathcal{E}_k, \mathcal{A}_k), (1\le k\le n)$ be n strongly connected graph with $\mathcal{V} = \{1, 2, \ldots, n\}$. Then, $\mathcal{A}_1\mathcal{A}_2\cdots\mathcal{A}_n$ is a positive matrix, i.e., every element of the matrix $\mathcal{A}_1\mathcal{A}_2\cdots\mathcal{A}_n$ is positive.

Proof: We just prove that the graph $\mathcal{G}_1^n$ corresponding to the matrix $\mathcal{A}_1\mathcal{A}_2\cdots\mathcal{A}_n$ is a complete graph. Denote the child node set of the node i in graph $\mathcal{G}_k$ as $\mathcal{O}_k(i)$. The corresponding child node set of the node i in graph $\mathcal{G}_1^n$ is denoted by $\mathcal{O}_1^n(i)$. For any $i\in\mathcal{V}$ and $j\in\mathcal{O}_1(i)$, we have

$$(\mathcal{A}_1\mathcal{A}_2)(i,j) = \sum_{k=1}^n \mathcal{A}_1(i,k)\mathcal{A}_2(k,j) \ge \mathcal{A}_1(i,j)\mathcal{A}_2(j,j) > 0. \tag{42}$$

Since $\mathcal{G}_2$ is strongly connected, if $\mathcal{O}_1(i)\ne\mathcal{V}$, then there exists two nodes $j_1\in\mathcal{V}\backslash\mathcal{O}_1(i)$ and $j_2\in\mathcal{O}_1(i)$ such that $(j_2, j_1)\in\mathcal{E}_2$, hence

$$(\mathcal{A}_1\mathcal{A}_2)(i,j_1) = \sum_{k=1}^n \mathcal{A}_1(i,k)\mathcal{A}_2(k,j_1) \ge \mathcal{A}_1(i,j_2)\mathcal{A}_2(j_2,j_1) > 0. \tag{43}$$

By (42) and (43), it is clear that $\{j_1\}\cup\mathcal{O}_1(i)\subset\mathcal{O}_1^2(i)$. Hence, for any $j\in\{j_1\}\cup\mathcal{O}_1(i)$, we have

$$(\mathcal{A}_1\mathcal{A}_2\mathcal{A}_3)(i,j) = \sum_{k=1}^n (\mathcal{A}_1\mathcal{A}_2)(i,k)\mathcal{A}_3(k,j) \ge (\mathcal{A}_1\mathcal{A}_2)(i,j)\mathcal{A}_3(j,j) > 0. \tag{44}$$

Since $\mathcal{G}_3$ is strongly connected, if $\{j_1\}\cup\mathcal{O}_1(i)\ne\mathcal{V}$, then there exists two nodes $j_2\in\mathcal{V}\backslash(\{j_1\}\cup\mathcal{O}_1(i))$ and $j_3\in\{j_1\}\cup\mathcal{O}_1(i)$ such that $(j_3, j_2)\in\mathcal{E}_3$, hence

$$(\mathcal{A}_1\mathcal{A}_2\mathcal{A}_3)(i,j_2) = \sum_{k=1}^n (\mathcal{A}_1\mathcal{A}_2)(i,k)\mathcal{A}_3(k,j_2) \ge (\mathcal{A}_1\mathcal{A}_2)(i,j_3)\mathcal{A}_3(j_3,j_2) > 0. \tag{45}$$

By (44) and (45), we can see that $\{j_2\}\cup\{j_1\}\cup\mathcal{O}_1(i)\subset\mathcal{O}_1^3(i)$. We repeat the above process until $\mathcal{O}_1^n(i) = \mathcal{V}$. The lemma can be proved by the arbitrariness of the node ■

Compared with the undirected graph case, the key difference is that the adjacency matrix in this section is an asymmetric and random matrix. Hence we need to deal with the effect of asymmetric random adjacency matrices. Here we assume the Markov chain $\{r(t), t\ge 0\}$ is independent of $\mathscr{F}_t$. By using the above lemma and Markov chain theory, we establish the stability of the algorithm (11), (12), (40), (41) under Markovian switching topology.

Theorem 5.1: Under Assumptions 4.2, 5.1, and 5.2, if for any $i\in\{1,\ldots,n\}$, $\sup_t \|\boldsymbol{\varphi}_{t,i}\|_{L_{6p}} < \infty$ and $\sigma_{3p} \triangleq \sup_t(\|\boldsymbol{W}_t\|_{L_{3p}} + \|\Delta\boldsymbol{\Theta}_t\|_{L_{3p}}) < \infty$ hold, then there exists a constant c' such that

$$\limsup_{t\to\infty} \|\widetilde{\boldsymbol{\Theta}}_t\|_{L_p} \le c'\sigma_{3p}.$$

Proof: Following the proof line of Theorem 4.2 in Section IV-C, it can be seen that we need to prove equation (33) holds under the assumptions of the theorem. By Assumption 5.2, there exists a positive integer q_0 such that

$$\Pr(r(t+q_0) = a | r(t) = b) > 0 \tag{46}$$

holds for all t and all states $a, b\in\mathbb{S}$. Denote $\Pi_k^t = \mathscr{A}_{r(t)}\mathscr{A}_{r(t-1)}\cdots\mathscr{A}_{r(k)}$. Then the i-th row, j-th column element of the matrix Π_k^t is denoted by $\Pi_k^t(i,j)$. Following Lemmas 4.1 and 4.4, we may abuse some notations $h' = 2h + nsq_0$, $z_t = \lfloor\frac{th'+nsq_0}{h}\rfloor + 1$ and

$$b_{t+1} = \frac{tr\left(\sum_{k=z_th+1}^{(z_t+1)h}\sum_{j=1}^n\left(\sum_{l=1}^n \Pi_{th'+1}^{k-1}(j,l)\boldsymbol{P}_{th'+1,l}\right)^2 \frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\right)}{n(h'-nsq_0)\left(\alpha^{h'} + \lambda_{\max}\left(\sum_{l=1}^n \boldsymbol{P}_{th'+1,l}\right)\right) tr(\boldsymbol{P}_{th'+1})}.$$

In the following we analyze the term $\mathbb{E}(b_{t+1}|\bar{\mathscr{F}}_{z_t h})$. By (46), we can see that there exists a positive constant p_0 such that for all t,

$$\begin{aligned}&\Pr\left(r(t+nsq_0)=s, r(t+(ns-1)q_0)=s-1,\ldots,\right.\\&\quad r(t+((n-1)s+1)q_0)=1;\\&\quad \cdots r(t+2sq_0)=s, r(t+(2s-1)q_0)=s-1,\ldots,\\&\quad r(t+(s+1)q_0)=1;\\&\quad r(t+sq_0)=s, r(t+(s-1)q_0)=s-1\cdots,\\&\quad \left. r(t+q_0)=1\middle| r(t)\right)\\&=\Pr\left(r(t+nsq_0)=s\middle| r(t+(ns-1)q_0)=s-1\right)\cdots\\&\Pr\left(r(t+((n-1)s+1)q_0)=1\middle| r(t+((n-1)s)q_0)=s\right)\\&\quad\cdots\Pr\left(r(t+q_0)=1|r(t)\right)\geq p_0>0. \qquad (47)\end{aligned}$$

By (47), we know that the Markov chain $\{r(t), t\geq 0\}$ can visit all states in $\mathbb{S}$ with n times in a positive probability during the time interval $[t+q_0, t+nsq_0]$. Hence for $k\in[z_t h+1,(z_t+1)h)]$, by Assumption 5.1 and Lemma 5.1, there exists a positive constant $\sigma>0$ such that the following inequality holds for all $j,l\in\{1,\ldots,n\}$,

$$\begin{aligned}&\mathbb{E}\left(\Pi_{th'+1}^{k-1}(j,l)\middle|\bar{\mathscr{F}}_k\right)=\mathbb{E}\left(\Pi_{th'+1}^{k-1}(j,l)\middle| r(th')\right)\\&\geq\sigma_0\mathbb{E}\left(\Pi_{th'+q_0}^{k-1}(j,l)\middle| r(th')\right)\geq\sigma, a.s.\end{aligned}$$

where $\bar{\mathscr{F}}_k$ is a σ-algebra generated by $\mathscr{F}_k$ and $\{r(1),\ldots, r(th')\}$. Then by the convexity of the matrix in [38] we have

$$\begin{aligned}&\mathbb{E}\left(\left(\sum_{l=1}^{n}\Pi_{th'+1}^{k-1}(j,l)\boldsymbol{P}_{th'+1,l}\right)^2\middle|\bar{\mathscr{F}}_k\right)\\&\geq\left(\mathbb{E}\left(\sum_{l=1}^{n}\Pi_{th'+1}^{k-1}(j,l)\boldsymbol{P}_{th'+1,l}\right)\middle|\bar{\mathscr{F}}_k\right)^2\\&=\left(\sum_{l=1}^{n}\mathbb{E}\left(\Pi_{th'+1}^{k-1}(j,l)\middle|\bar{\mathscr{F}}_k\right)\boldsymbol{P}_{th'+1,l}\right)^2\\&\geq\sigma^2\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)^2.\end{aligned}$$

By $\mathscr{F}_{z_t h}\subset\mathscr{F}_k$ and $\boldsymbol{\varphi}_{k,j}\in\mathscr{F}_k$, we conclude that

$$\begin{aligned}&\mathbb{E}\left(\left(\sum_{l=1}^{n}\Pi_{th'+1}^{k-1}(j,l)\boldsymbol{P}_{th'+1,l}\right)^2\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\middle|\bar{\mathscr{F}}_{z_t h}\right)\\&=\mathbb{E}\left(\left(\mathbb{E}\left(\sum_{l=1}^{n}\Pi_{th'+1}^{k-1}(j,l)\boldsymbol{P}_{th'+1,l}\right)^2\middle|\bar{\mathscr{F}}_k\right)\right.\\&\quad\left.\cdot\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\middle|\bar{\mathscr{F}}_{z_t h}\right)\\&\geq\sigma^2\mathbb{E}\left(\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)^2\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\middle|\bar{\mathscr{F}}_{z_t h}\right), \qquad (48)\end{aligned}$$

where $\bar{\mathscr{F}}_{z_t h}$ is a σ-algebra generated by $\mathscr{F}_{z_t h}$ and $\{r(1),\ldots, r(th')\}$. From the above analysis, we can obtain the following inequality

$$\begin{aligned}&\mathbb{E}(b_{t+1}|\bar{\mathscr{F}}_{z_t h})\geq\\&\frac{\sum\limits_{k=z_t h+1}^{(z_t+1)h}\sum\limits_{j=1}^{n}\sigma^2 tr\left(\mathbb{E}\left(\left(\sum\limits_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)^2\frac{\boldsymbol{\varphi}_{k,j}\boldsymbol{\varphi}_{k,j}^T}{1+\|\boldsymbol{\varphi}_{k,j}\|^2}\middle|\bar{\mathscr{F}}_{z_t h}\right)\right)}{n(h'-nsq_0)\left(\alpha^{h'}+\lambda_{\max}\left(\sum\limits_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)\right)tr(\boldsymbol{P}_{th'+1})}\\&=\frac{\sigma^2 tr\left[\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)^2\boldsymbol{H}_{z_t}\right]}{n(h'-nsq_0)\left(\alpha^{h'}+\lambda_{\max}\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)\right)tr(\boldsymbol{P}_{th'+1})}\\&\geq\frac{\sigma^2\lambda_{\min}(\boldsymbol{H}_{z_t})tr\left[\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)^2\right]}{n(h'-nsq_0)\left(\alpha^{h'}+\lambda_{\max}\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)\right)tr(\boldsymbol{P}_{th'+1})}.\end{aligned}$$

Then by the above inequality and $tr((\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l})^2)\geq m^{-1}(tr(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}))^2=m^{-1}(tr(\boldsymbol{P}_{th'+1}))^2$, we have

$$\begin{aligned}&\mathbb{E}(b_{t+1}|\bar{\mathscr{F}}_{z_t h})\\&\geq\frac{\sigma^2\lambda_{\min}(\boldsymbol{H}_{z_t})tr(\boldsymbol{P}_{th'+1})}{mn(h'-nsq_0)\left(\alpha^{h'}+\lambda_{\max}\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)\right)}\\&\geq\frac{\sigma^2(1+h)\lambda_{z_t}tr(\boldsymbol{P}_{th'+1})}{m(h'-nsq_0)\left(\alpha^{h'}+\lambda_{\max}\left(\sum_{l=1}^{n}\boldsymbol{P}_{th'+1,l}\right)\right)}\\&\geq\frac{\sigma^2(1+h)\lambda_{z_t}tr(\boldsymbol{P}_{th'+1})}{m(h'-nsq_0)\left(1+tr\left(\boldsymbol{P}_{th'+1}\right)\right)}\\&\geq\frac{\sigma^2(1+h)\lambda_{z_t}}{2m(h'-nsq_0)}\text{on}\{tr(\boldsymbol{P}_{th'+1})\geq 1\}.\end{aligned}$$

Hence by following the rest part proof of Lemma 4.4 and replacing the notation $D_{\mathcal{G}}$ with nsq_0, we can obtain $\sup_{t\geq 0}\mathbb{E}(\|\boldsymbol{P}_t\|^p)<\infty$. Furthermore, by Assumption 5.1, we know that equation (38) still hold, which yields the exponential stability of the homogeneous part of the error equation. Then by following the proof of Theorem 4.2, we can obtain $\limsup_{t\to\infty}\|\widetilde{\boldsymbol{\Theta}}_t\|_{L_p}\leq c'\sigma_{3p}$, which completes the proof of Theorem 5.1. ■

Remark 5.1: From Theorem 5.1, (also Theorems 4.1 and 4.2), we see that our results are obtained without using the independence or stationarity assumptions on the regression signals by virtue of some powerful techniques including stochastic stability theory, Markov chain theory and some matrix inequalities. Thus, it is possible to apply the distributed FFLS algorithm to practical feedback systems.

VI. Simulation Results

In this section, we provide two examples to illustrate the theoretical results where the regression signals do not satisfy the i.i.d. condition.

Example 6.1: Let us consider a network composed of $n=5$ sensors whose dynamics obey the model (1) with the dimension

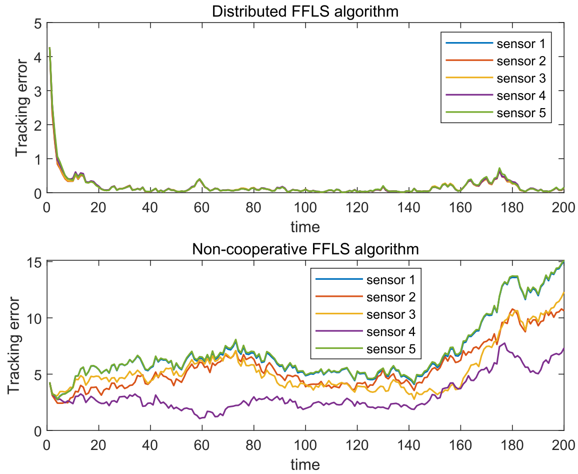

Fig. 1. Tracking errors of noncooperative FFLS algorithm and distributed FFLS algorithm under the undirected graph.

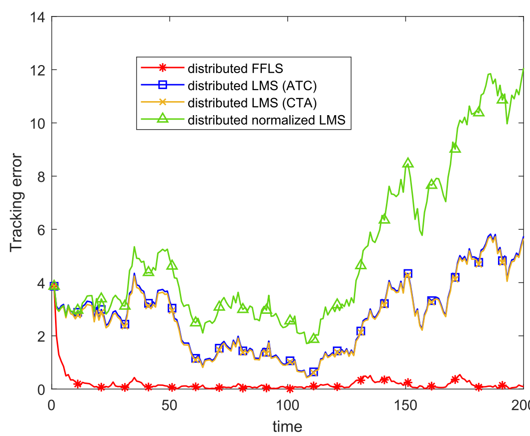

Fig. 2. Tracking errors of several distributed algorithms under the undirected graph.

$m=4$. The noise sequence $\{w_{t,i}, t\geq 1, i=1,\ldots,5\}$ in (1) is independent and identically distributed with $w_{t,i}\sim\mathcal{N}(0,0.1)$ (Gaussian distribution with zero mean and variance 0.1). Let the regression vectors $\boldsymbol{\varphi}_{t,i}$ be generated by the following state space model:

$$\begin{aligned}\boldsymbol{x}_{t,i} &= \boldsymbol{A}_i\boldsymbol{x}_{t-1,i}+\boldsymbol{B}_i\varepsilon_{t,i}\\ \boldsymbol{\varphi}_{t,i} &= \boldsymbol{C}_i\boldsymbol{x}_{t,i}\end{aligned} \tag{49}$$

where $\boldsymbol{x}_{t,i}\in\mathbb{R}^4$ is the state with the initial value $\boldsymbol{x}_{0,i}=[1,1,1,1]^T$, $\boldsymbol{A}_i=\text{diag}\{1/2,3/4,4/5,2/3\}$, $(i=1,\ldots,5)$, $\boldsymbol{B}_1=\boldsymbol{B}_5=[1,0,0,0]^T$, $\boldsymbol{B}_2=[0,1,0,0]^T$, $\boldsymbol{B}_3=[0,0,1,0]^T$, $\boldsymbol{B}_4=[0,0,0,1]^T$, $\boldsymbol{C}_1=\boldsymbol{C}_5=\text{diag}\{1,0,0,0\}$, $\boldsymbol{C}_2=\text{diag}\{0,1,0,0\}$, $\boldsymbol{C}_3=\text{diag}\{0,0,1,0\}$, and $\boldsymbol{C}_4=\text{diag}\{0,0,0,1\}$. The noise sequence $\{\varepsilon_{t,i}, t\geq 1, i=1,\ldots,n\}$ in (49) is independent and identically distributed with $\varepsilon_{t,i}\sim\mathcal{N}(0,1.2)$. It is clear that the regression vectors $\boldsymbol{\varphi}_{t,i}$ generated by (49) are non-i.i.d..

Assume that the unknown time-varying parameter $\boldsymbol{\theta}_t$ satisfies $\boldsymbol{\theta}_t=\boldsymbol{\theta}_{t-1}+\gamma\boldsymbol{v}_t$, with each element of $\boldsymbol{v}_t\sim\mathcal{N}(0,1)$. Here, we set $\gamma=0.08$. The adjacency matrix of the undirected network

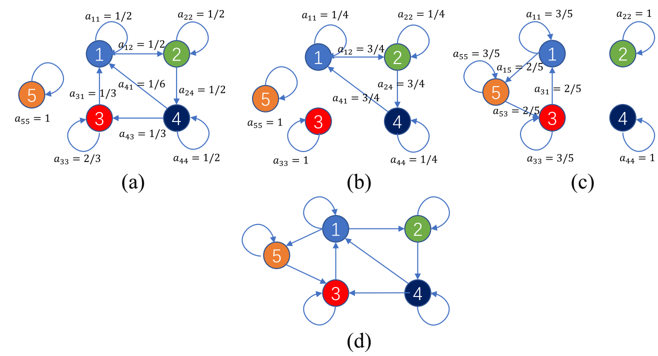

Fig. 3. (a)–(c) are the topology of $\mathcal{G}_1$, $\mathcal{G}_2$, and $\mathcal{G}_3$, respectively. (d) is the topology of the union graph. The union graph of these three graphs is strongly connected while non of them is.

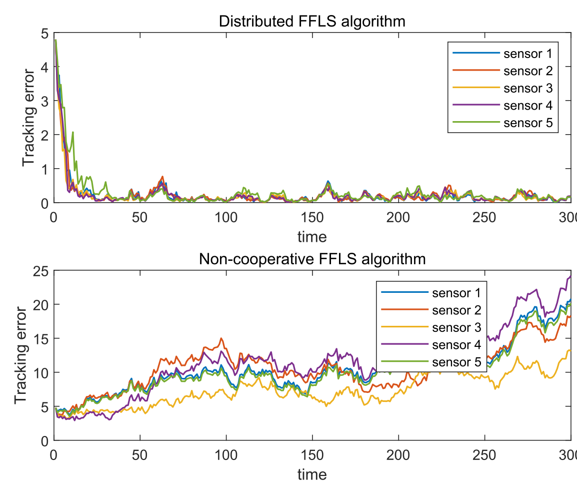

Fig. 4. Tracking errors of noncooperative FFLS algorithm and distributed FFLS algorithm under Markovian switching directed graphs.

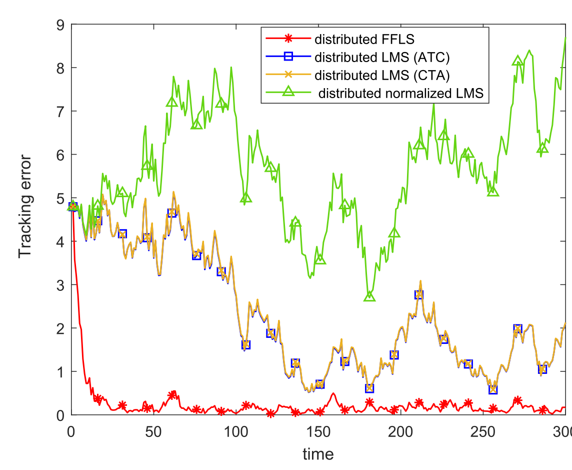

Fig. 5. Tracking errors of several distributed algorithms under Markovian switching directed graphs.

is taken as

$$\mathcal{A}=\begin{pmatrix}\frac{3}{4} & 0 & \frac{1}{4} & 0 & 0\\ 0 & \frac{1}{4} & \frac{1}{12} & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{4} & \frac{1}{12} & \frac{1}{6} & \frac{1}{4} & \frac{1}{4}\\ 0 & \frac{1}{3} & \frac{1}{4} & \frac{1}{12} & \frac{1}{3}\\ 0 & \frac{1}{3} & \frac{1}{4} & \frac{1}{3} & \frac{1}{12}\end{pmatrix}.$$

By the definition of adjacency matrix, it is easy to see that the corresponding undirected graph is connected. For the system settings, we can verify that for each sensor i $(i = 1 \cdots, 5)$, the regression signals $\boldsymbol{\varphi}_{t,i}$ [generated by (49)] cannot satisfy the excitation condition (19) for any single sensors, but they can cooperatively satisfy Assumption 4.2 with $h = 1$. We repeat the simulations for $\mathrm{s} = 200$ times with the same initial states.

1) We estimate the unknown parameter $\boldsymbol{\theta}_t$ by using the standard noncooperative FFLS algorithm (i.e., Algorithm 1) and distributed FFLS algorithm (i.e., Algorithm 2) with forgetting factor $\alpha = 0.9$. Fig. 1 shows tracking errors by the noncooperative FFLS algorithm and the distributed FFLS algorithm for the time-varying unknown parameter. From Fig. 1, we can see that if we use the standard noncooperative FFLS algorithm to estimate $\boldsymbol{\theta}_t$, the tracking errors of all five sensors are large because all the sensors do not satisfy the information condition (19), while the tracking errors of all five sensors in the distributed FFLS algorithm lie in a small neighborhood of 0 since all sensors cooperatively satisfy Assumption 4.2, which can reveal the cooperative effect of sensors in a sense that the estimation or tracking task can be fulfilled through exchanging information between sensors even though any individual sensor cannot.

2) We compare our algorithm (Algorithm 2) with three types of distributed LMS algorithms in [16] and [23] [i.e., distributed normalized LMS, combination then adaption (CTA) type distributed LMS, and adaption then combination (ATC) type distributed LMS]. We conduct the simulations by using the same regressors and initial states as above. The average tracking errors on the whole network of different distributed algorithms are shown in Fig. 2, from which we see that the distributed FFLS algorithm proposed in this article has better tracking performance than the three types of distributed LMS algorithms.

We next construct another simulation example to demonstrate the performance of the distributed FFLS algorithm under Markovian switching directed graphs.

Example 6.2: Take the same system settings including initial states and regression vectors as those in Example 6.1. The directed graphs $\mathcal{G}_{r(t)} = (\mathcal{V}, \mathcal{E}_{r(t)}, \mathcal{A}_{r(t)})$ with $r_t \in \{1, 2, 3\}$ are depicted in Fig. 3 . The transition probability matrix of the Markov chain r_t is chosen to be $\boldsymbol{P} = \begin{pmatrix} \frac{1}{5} & \frac{3}{10} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{10} & \frac{2}{5} \\ \frac{3}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}$. It is clear that Assumptions 5.1 and 5.2 are satisfied. Then, we get the tracking errors by using the noncooperative FFLS algorithm and the distributed FFLS algorithm under the Markovian switching directed graphs, see Fig. 4 . Moreover, Fig. 5 compares the tracking performance of the distributed FFLS algorithm with distributed LMS algorithms under the Markovian switching directed graphs. By Figs. 4 and 5, we can obtain similar results as those in Example 6.1.

VII. Conclusion

This article proposed a distributed FFLS algorithm to collaboratively track an unknown time-varying parameter by minimizing a local loss function with a forgetting factor. By introducing a spatio-temporal cooperative excitation condition, we established the stability of the proposed distributed FFLS algorithm for fixed undirected graph case. Then, the theoretical results were generalized to the case of Markovian switching directed graphs. The cooperative excitation condition revealed that the sensors can collaboratively accomplish the tracking task even though any individual sensor cannot. Simulation results showed that the distributed FFLS algorithm has better tracking performance than the well-investigated distributed LMS algorithms. We note that our theoretical results are established without using independence or stationarity conditions of the regression vectors. Thus, a relevant research topic is how to combine the distributed adaptive estimation with the distributed control. How to establish the stability analysis of the distributed algorithms for more complex cases, such as considering quantization effect or time-delay in communication channels is another interesting research topic.

References

[1] W. Ren and R. Beard, "Consensus seeking in multiagent systems under dynamically changing interaction topologies," *IEEE Trans. Autom. Control*, vol. 50, no. 5, pp. 655–661, May 2005.

[2] Y. Wang, L. Cheng, W. Ren, Z.-G. Hou, and M. Tan, "Seeking consensus in networks of linear agents: Communication noises and markovian switching topologies," *IEEE Trans. Autom. Control*, vol. 60, no. 5, pp. 1374–1379, May 2015.

[3] K. Lu, H. Xu, and Y. Zheng, "Distributed resource allocation via multiagent systems under time-varying networks," *Automatica*, vol. 136, 2022, Art. no. 110059.

[4] B. Wang, Q. Fei, and Q. Wu, "Distributed time-varying resource allocation optimization based on finite-time consensus approach," *IEEE Control Syst. Lett.*, vol. 5, no. 2, pp. 599–604, Apr. 2021.

[5] Z. Lin, L. Wang, Z. Han, and M. Fu, "Distributed formation control of multi-agent systems using complex laplacian," *IEEE Trans. Autom. Control*, vol. 59, no. 7, pp. 1765–1777, Jul. 2014.

[6] Y. Zhi, L. Liu, B. Guan, B. Wang, Z. Cheng, and H. Fan, "Distributed robust adaptive formation control of fixed-wing UAVs with unknown uncertainties and disturbances," *Aerosp. Sci. Technol.*, vol. 126, 2022, Art. no. 107600.

[7] G. Battistelli and L. Chisci, "Kullback–Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability," *Automatica*, vol. 50, no. 3, pp. 707–718, 2014.

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

[9] M. U. Javed, J. I. Poveda, and X. Chen, "Excitation conditions for uniform exponential stability of the cooperative gradient algorithm over weakly connected digraphs," *IEEE Contr. Syst. Lett.*, vol. 6, pp. 67–72, 2022.

[10] F. Barani, A. Savadi, and H. S. Yazdi, "Convergence behavior of diffusion stochastic gradient descent algorithm," *Signal Process.*, vol. 183, 2021, Art. no. 108014.

[11] 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.

[12] L. Zhang, Y. Cai, C. Li, and R. C. de Lamare, "Variable forgetting factor mechanisms for diffusion recursive least squares algorithm in sensor networks," *EURASIP J. Adv. Signal Process.*, vol. 2017, 2017, Art. no. 57.

[13] 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.

[14] J. Lei and H. Chen, "Distributed estimation for parameter in heterogeneous linear time-varying models with observations at network sensors," *Commun. Inf. Syst.*, vol. 15, no. 4, pp. 423–451, 2015.

[15] G. Mateos and G. B. Giannakis, "Distributed recursive least-squares: Stability and performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3740–3754, Jul. 2012.

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

[17] D. Gan and Z. Liu, “Performance analysis of the compressed distributed least squares algorithm,” *Syst. Control Lett.*, vol. 164, 2022, Art. no. 105228.

[18] D. Gan and Z. Liu, “Distributed order estimation of ARX model under cooperative excitation condition,” *SIAM J. Control Optim.*, vol. 60, no. 3, pp. 1519–1545, 2022.

[19] D. Gan, S. Xie, and Z. Liu, “Stability of the distributed Kalman filter using general random coefficients,” *Sci. China Inf. Sci.*, vol. 64, pp. 172204:1–172204:14, 2021.

[20] L. Guo, “Estimation, control, and games of dynamical systems with uncertainty,” *Scientia Sinica Informationis*, vol. 50, no. 9, pp. 1327–1344, 2020.

[21] D. Gan and Z. Liu, “Convergence of the distributed SG algorithm under cooperative excitation condition,” *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, 2022, doi 10.1109/TNNLS.2022.3213715.

[22] S. Xie, Y. Zhang, and L. Guo, “Convergence of a distributed least squares,” *IEEE Trans. Autom. Control*, vol. 66, no. 10, pp. 4952–4959, Oct. 2021.

[23] S. 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.

[24] O. Macchi and E. Eweda, “Compared speed and accuracy of RLS and LMS algorithms with constant forgetting factors,” *Traitement Signal*, vol. 22, pp. 255–267, 1988.

[25] Y. Hatano and M. Mesbahi, “Agreement over random networks,” *IEEE Trans. Autom. Control*, vol. 50, no. 11, pp. 1867–1872, Nov. 2005.

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

[27] I. Matei, N. Martins, and J. S. Baras, “Almost sure convergence to consensus in Markovian random graphs,” in *Proc. IEEE 47th Conf. Decis. Control*, 2008, pp. 3535–3540.

[28] K. You, Z. Li, and L. Xie, “Consensus condition for linear multi-agent systems over randomly switching topologies,” *Automatica*, vol. 49, no. 10, pp. 3125–3132, 2013.

[29] M. Meng, L. Liu, and G. Feng, “Adaptive output regulation of heterogeneous multiagent systems under Markovian switching topologies,” *IEEE Trans. Cybern.*, vol. 48, no. 10, pp. 2962–2971, Oct. 2018.

[30] 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.

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

[32] G. Zielke, “Inversion of modified symmetric matrices,” *J. Assoc. Comput. Machinery*, vol. 15, no. 3, pp. 402–408, 1968.

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

[34] L. Guo, *Time-Varying Stochastic Systems, Stability and Adaptive Theory*, 2nd ed., Beijing, China: Science Press, 2020.

[35] C. Godsil and G. Royle, *Algebraic Graph Theory*. Berln, Germay: Springer, 2001.

[36] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[37] S. Karlin and H. Taylor, *A Second Course in Stochastic Processes*. Cambridge, MA, USA: Academic, 1981.

[38] M. Rahman and M. Ahsanullah, “A note on the expected values of powers of a matrix,” *Can. J. Statistics/La Revue Canadienne de Statistique*, vol. 1, no. 1, pp. 123–125, 1973.

Die Gan (Member, IEEE) received the B.S. degree in mathematics from Shandong University, Jinan, China, in 2017, and the Ph.D. degree in systems theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2022.

She is currently a Postdoctoral Fellow with the Zhongguancun Laboratory, Beijing, China. Her research interests include system identification, machine learning and distributed adaptive filter.

Dr. Gan was the recipient of the China National Scholarship for graduate students in 2020, and the outstanding graduate title of Beijing in 2022.

Siyu Xie received the B.S. degree in information and computing science (systems control) from Beihang University, Beijing, China, in 2013, and the Ph.D. degree in control theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2018.

From 2019 to 2022, she was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI, USA. She is currently a Professor with the School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu, China. Her research interests include networked systems, distributed optimization problems for power systems, distributed adaptive filters, identification, and distributed control.

Dr. Xie was the recipient of the IEEE CSS Beijing Chapter Young Author Prize in 2015, the China National Scholarship for graduate students in 2017, the President Scholarship of Chinese Academy of Sciences in 2018, and the Excellent Doctoral dissertation of Chinese Academy of Sciences in 2019.

Zhixin Liu (Member, IEEE) received the B.S. degree in mathematics from Shandong University, Jinan, China, in 2002, and the Ph.D. degree in control theory from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2007.

She had visiting positions with the KTH Royal Institute of Technology, University of New South Wales, Canberra and University of Maryland, College Park. She is currently a Professor with AMSS, CAS. She has coauthored the SIGEST paper in *SIAM Review* in 2014. Her research interests include complex systems and multiagent systems.

Dr. Liu was the corecipient of the Best Theory Paper Award at the 13rd World Congress on Intelligent Control and Automation.

Jinhu Lü (Fellow, IEEE) received the Ph.D. degree in applied mathematics from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2002.

He was a Professor with RMIT University, Melbourne, VIC, Australia, and a Visiting Fellow with Princeton University, Princeton, NJ, USA. He is currently a Vice-President for scientific research with Beihang University, Beijing, China. He is also an Adjunct Professor with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He is a Chief Scientist with the National Key Research and Development Program of China and a Leading Scientist with Innovative Research Groups of the National Natural Science Foundation of China. His research interests include complex networks, industrial Internet, network dynamics, and cooperation control.

Dr. Lü is the Fellow of Chinese Automation Society, Operations Research Society of China, and Chinese Institute of Command and Control. He was the recipient of the prestigious Ho Leung Ho Lee Foundation Award in 2015; the National Innovation Competition Award in 2020; the State Natural Science Award three times from the Chinese Government in 2008, 2012, and 2016, respectively; the Australian Research Council Future Fellowships Award in 2009; the National Natural Science Fund for Distinguished Young Scholars Award; and the Highly Cited Researcher Award in engineering from 2014 to 2020, and 2022. He is/was an Editor in various ranks for 15 SCI journals, including the Co-Editor-in-Chief of IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS. He served as a member in the Fellows Evaluating Committee of the IEEE International Conference on Artificial Intelligence Circuits and Systems, the IEEE Computational Intelligence Society, and the IEEE Industrial Electronics Society. He was the General Co-Chair of the Annual Conference of the IEEE Industrial Electronics Society 2017.