Decentralized Cooperative Online Estimation With Random Observation Matrices, Communication Graphs and Time Delays

Jiexiang Wang, Tao Li, *Senior Member, IEEE*, and Xiwei Zhang

***Abstract*— We analyze convergence of decentralized cooperative online estimation algorithms by a network of multiple nodes via information exchanging in an uncertain environment. Each node has a linear observation of an unknown parameter with randomly time-varying observation matrices. The underlying communication network is modeled by a sequence of random digraphs and is subjected to nonuniform random time-varying delays in channels. Each node runs an online estimation algorithm consisting of a consensus term taking a weighted sum of its own estimate and neighbours' delayed estimates, and an innovation term processing its own new measurement at each time step. By stochastic time-varying system, martingale convergence theories and the binomial expansion of random matrix products, we transform the convergence analysis of the algorithm into that of the mathematical expectation of random matrix products. Firstly, for the delay-free case, we show that the algorithm gains can be designed properly such that all nodes' estimates converge to the true parameter in mean square and almost surely if the observation matrices and communication graphs satisfy the *stochastic spatio-temporal persistence of excitation* condition. Secondly, for the case with time delays, we introduce *delay matrices* to model the random time-varying communication delays between nodes. It is shown that under the *stochastic spatio-temporal persistence of excitation* condition, for any given bounded delays, proper algorithm gains can be designed to guarantee mean square convergence for the case with conditionally balanced digraphs.**

***Index Terms*— Decentralized online estimation, cooperative estimation, random graph, random time delay, persistence of excitation.**

I. Introduction

ESTIMATION algorithms have important applications in many fields, e.g. navigation systems, space exploration, machine learning and power systems ([1]–[4]), etc. In a power system, measurement devices such as remote terminal units and phasor measurement units, send the measured active and reactive power flows, bus injection powers and voltage amplitudes to the Supervisory Control and Data Acquisition (SCDA) system, then the voltage amplitudes and phase angles at all buses are estimated for secure and stable operation of the system ([5], [6]). Generally speaking, there are mainly two categories of estimation algorithms in term of information structure, i.e. centralized and decentralized algorithms. In a centralized algorithm, a fusion center is used to collect all nodes's measurements and gives the global estimate. This structure heavily relies on the fusion center and lacks robustness and security. In a decentralized algorithm, a network of multiple nodes is employed to cooperatively estimate the unknown parameter via information exchanging, where each node is an entity with integrated capacity of sensing, computing and communication, and occasional node/link failures may not destroy the entire estimation task. Hence, decentralized cooperative estimation algorithms are more robust than centralized ones ([7], [8]).

There exist various kinds of uncertainties in real networks. For example, sensors are usually powered by chemical or solar cells, and the unpredictability of cell power leads to random node/link failures, which can be modeled by a sequence of random communication graphs. Besides, node sensing failures or measurement losses ([9]) can be modeled by a sequence of random observation matrices. There are lots of literature on decentralized online estimation problems with random graphs. Ugrinovskii [10] studied decentralized estimation with Markovian switching graphs. Kar & Moura [11] and Sahu *et al.* [12] considered decentralized estimation with i.i.d. graph sequences, where Kar & Moura [11] showed that the algorithm achieves weak consensus under a weak distributed detectability condition and Sahu *et al.* [12] proved that the algorithm converges almost surely if the mean graph is balanced and strongly connected. Simões & Xavier [13] proposed a decentralized estimation algorithm with i.i.d. undirected graphs and proved that the convergence rate of mean square estimation error is asymptotically equal to that of the centralized algorithm. Decentralized cooperative online estimation based on diffusion strategies was addressed in [14]–[18] with spatio-temporally independent observation matrices, i.e. the sequence of observation matrices of each node is an independent random process and those of different nodes are mutually independent.

Manuscript received August 22, 2019; revised November 17, 2020; accepted March 22, 2021. Date of publication April 5, 2021; date of current version May 20, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61977024 and in part by the Basic Research Project of Shanghai Science and Technology Commission under Grant 20JC1414000. This article was presented in part at the 2020 European Control Conference. *(Corresponding author: Tao Li.)*

Jiexiang Wang is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China (e-mail: 920331447@qq.com).

Tao Li and Xiwei Zhang are with the Key Laboratory of Pure Mathematics and Mathematical Practice, School of Mathematical Sciences, East China Normal University, Shanghai 200241, China (e-mail: tli@math.ecnu.edu.cn; xwzhangmath@sina.com).

Communicated by R. Balan, Associate Editor for Detection and Estimation.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2021.3071167.

Digital Object Identifier 10.1109/TIT.2021.3071167

Piggott & Solo [19], [20] studied decentralized estimation with temporally correlated observation matrices and a fixed communication graph. Ishihara & Alghunaim [21] studied decentralized estimation with spatially independent observation matrices. Kar *et al.* [22] and Kar & Moura [23] proposed consensus+innovations decentralized estimation algorithms with random graphs and observation matrices, where the sequences of communication graphs and observation matrices are both i.i.d. They proved that the algorithm converges almost surely if the mean graph is balanced and strongly connected. Zhang & Zhang [24] considered decentralized estimation with finite Markovian switching graphs and i.i.d. observation matrices, and proved that the algorithm converges in mean square and almost surely if all graphs are balanced and jointly contain a spanning tree. Zhang *et al.* [25] proposed a robust decentralized estimation algorithm with the communication graphs and observation matrices being mutually independent with each other and both uncorrelated sequences. In summary, most existing literature on decentralized cooperative estimation algorithms required balanced mean graphs and special statistical properties of communication graphs and observation matrices, such as i.i.d. or Markovian switching graph sequences, spatially or temporally independent observation matrices with the fixed mathematical expectation, which are also independent of communication graphs.

Besides random communication graphs and observation matrices, random communication delays are also common in real systems ([26]–[28]). Due to congestions of communication links and external interferences, time delays are usually random and time-varying, whose probability distribution can be approximately estimated by statistical methods. However, to our best knowledge, there has been no literature on decentralized online estimation with general random time-varying communication delays. Zhang *et al.* [29] and Millán *et al.* [30] considered decentralized estimation with uniform deterministic time-invariant and time-varying communication delays, respectively, where Millán *et al.* [30] established a LMI type convergence condition by the Lyapunov-Krasovskii functional method.

In this paper, we analyze convergence of decentralized cooperative online parameter estimation algorithms with random observation matrices, communication graphs and time delays. Each node's algorithm consists of a consensus term taking a weighted sum of its own estimate and delayed estimates of its neighbouring nodes, and an innovation term processing its own new measurement at each time step. The sequences of observation matrices, communication graphs and time delays are not required to satisfy special statistical properties, such as mutual independence and spatio-temporal independence. Furthermore, neither the sample paths of the random graphs nor the mean graphs are necessarily balanced and connected at each time step. These relaxations together with the existence of random time-varying time delays bring essential difficulties to the convergence analysis, and most existing methods are not applicable. For example, the frequency domain approach ([29], [31]) is only suitable for deterministic uniform time-invariant time delays, and the Lyapunov-Krasovskii functional method leads to a non-explicit LMI type convergence condition ([30]). Liu *et al.* [32] and Liu *et al.* [33] addressed distributed consensus with deterministic time-varying communication delays and i.i.d. communication graphs. The analysis method therein required the mean graph to be time-invariant and connected at each time step, and is not applicable to time-varying mean graphs.

We introduce *delay matrices* to model the random time-varying communication delays between each pair of nodes. By stochastic time-varying system, martingale convergence theories and the binomial expansion of random matrix products, we transform the convergence analysis of the algorithm into that of the mathematical expectation of random matrix products. Firstly, for the delay-free case, we show that the algorithm gains can be designed properly such that all nodes' estimates converge to the true parameter in mean square and almost surely if the observation matrices and communication graphs satisfy the *stochastic spatio-temporal persistence of excitation* condition. Especially, it is shown that for Markovian switching communication graphs and observation matrices, this condition holds if the stationary graph is balanced with a spanning tree and the measurement model is *spatio-temporally jointly observable*. Secondly, for the case with time delays, we propose several conditions for mean square convergence, which explicitly relies on the conditional expectations of delay matrices, observation matrices and weighted adjacency matrices of communication graphs over a sequence of fixed-length time intervals. Furthermore, we show that if the communication graphs are conditionally balanced, then under the *stochastic spatio-temporal persistence of excitation* condition, for any given bounded delays, proper algorithm gains can be designed to guarantee mean square convergence of the algorithm. Compared with the existing literature, our contributions are summarized as below.

- The delay-free case
  - We show that it is not necessary that the sequences of observation matrices and communication graphs be mutually independent or spatio-temporally independent. Also, the mean graphs are not necessarily time-invariant and balanced. We establish the *stochastic spatio-temporal persistence of excitation* condition under which the algorithm with random graphs and observation matrices converges in mean square and almost surely. For a network consisting of completely isolated nodes, the *stochastic spatio-temporal persistence of excitation* condition degenerates to a set of independent *stochastic persistence of excitation* conditions for centralized algorithms ([38]).
  - Especially, for the case with Markovian switching communication graphs and observation matrices, we prove that the *stochastic spatio-temporal persistence of excitation* condition holds if the stationary graph is balanced with a spanning tree and the measurement model is spatio-temporally jointly observable, implying that *neither local observability of each node nor instantaneous global observability of the entire measurement model is necessary*.

- The case with time delays
  - We introduce *delay matrices* to model the random time-varying time delays between each pair of nodes. By the method of binomial expansion of random matrix products, we obtain several conditions for mean square convergence, which explicitly relies on the conditional expectations of the delay matrices, observation matrices and weighted adjacency matrices of communication graphs over a sequence of fixed-length time intervals. These conditions show that for given algorithm gains, the communication graphs and observation matrices need to be persistently excited with enough intensity to mitigate the random time delays. We further show that if the *stochastic spatio-temporal persistence of excitation* condition holds, then for any given bounded delays, proper algorithm gains can be designed to guarantee mean square convergence of the algorithm for the case with conditionally balanced digraphs.
  - In this paper, the nonuniform random time-varying communication delays are more general, and correlated communication delays, graphs and observation matrices are allowed.

The rest of the paper is arranged as follows. In Section II, we formulate the problem. In Section III, we describe the decentralized cooperative online parameter estimation algorithm with random observation matrices, communication graphs and time delays. We make the convergence analysis for the delay-free case and the case with time delays in Sections IV and V, respectively. In Section VI, we give a numerical example to demonstrate the theoretical results. Finally, we conclude the paper and give some future topics in Section VII.

Notation and symbols: $\circ$: the Hadamard product; $\otimes$: the Kronecker product; $\mathrm{Tr}(A)$: the trace of matrix A; $\|A\|$: the 2-norm of matrix A; A^T: the transpose of matrix A; $\mathbb{P}\{A\}$: the probability of event A; I_n: the n dimensional identity matrix; $\rho(A)$: the spectral radius of matrix A; $|a|$: the absolute value of real number a; $\mathbb{R}^n$: the n dimensional real vector space; $A \geq B$: the matrix $A-B$ is positive semidefinite; $\lfloor x \rfloor$: the largest integer less than or equal to x; $\lceil x \rceil$: the smallest integer greater than or equal to x; $\mathbb{E}[\xi]$: the mathematical expectation of random variable ξ; $\lambda_{\min}(A)$: the minimum eigenvalue of real symmetric matrix A; $\mathbf{1}_n$: the n dimensional column vector with all entries being one; $\mathbf{0}_{n\times m}$: the $n\times m$ dimensional matrix with all entries being zero; $b_n = O(r_n)$: $\limsup_{n\to\infty} \frac{|b_n|}{r_n} < \infty$, where $\{b_n, n \geq 0\}$ is a sequence of real numbers, $\{r_n, n \geq 0\}$ is a sequence of real positive numbers; $b_n = o(r_n)$: $\lim_{n\to\infty} \frac{b_n}{r_n} = 0$; For a sequence of $n \times n$ dimensional matrices $\{Z(k), k \geq 0\}$ and a sequence of scalars $\{c(k), k \geq 0\}$, denote

$$\Phi_Z(j,i) = \begin{cases} Z(j)\cdots Z(i), & j \geq i \\ I_n, & j < i. \end{cases},$$

$$\prod_{k=i}^{j} c(k) = \begin{cases} c(j)\cdots c(i), & j \geq i \\ 1, & j < i. \end{cases}$$

For any nonnegative integers i and j, denote the Kronecker function by $\mathcal{I}_{i,j}$, satisfying $\mathcal{I}_{i,j} = 1$ if $i = j$ and $\mathcal{I}_{i,j} = 0$ otherwise.

II. Problem Formulation

A. Measurement Model

Consider a network of N nodes. Each node is an estimator with integrated capacity of sensing, computing, storage and communication. The estimators/nodes cooperatively estimate an unknown parameter vector $x_0 \in \mathbb{R}^n$ via information exchanging. The relation between the measurement vector $z_i(k) \in \mathbb{R}^{n_i}$ of estimator i and the unknown parameter x_0 is represented by

$$z_i(k) = H_i(k)x_0 + v_i(k),\ i = 1, \cdots, N,\ k \geq 0. \tag{1}$$

Here, $H_i(k) \in \mathbb{R}^{n_i\times n}$ is the random observation (regression) matrix at time instant k with $n_i \leq n$, and $v_i(k) \in \mathbb{R}^{n_i}$ is the additive measurement noise. Denote $z(k) = [z_1^T(k), \cdots, z_N^T(k)]^T$, $H(k) = [H_1^T(k), \cdots, H_N^T(k)]^T$ and $v(k) = [v_1^T(k), \cdots, v_N^T(k)]^T$. Rewrite (1) by the compact form

$$z(k) = H(k)x_0 + v(k),\ k \geq 0. \tag{2}$$

Remark 1: In many real applicaitons, the relations between the unknown parameter and the measurements can be represented by (1). For example, in the decentralized multi-area state estimation in power systems, the grid is partitioned into multiple geographically non-overlapping areas, and each area is regarded as a node. The grid state x_0 to be estimated consists of voltage amplitudes and phase angles at all buses. The measurement $z_i(k)$ of each area/node consists of the active and reactive power flow, bus injection powers and voltage amplitude information measured by remote terminal units and phasor measurement units in the i-th area. By the DC power flow approximation ([34]), the grid state degenerates to the voltage phase angles at all buses and the relation between the measurement of each area and the grid state can be represented by (1). In decentralized parameter identification, each node's measurement equation is given by

$$\begin{aligned} z_i(k) &= \sum_{j=1}^{n} c_j z_i(k-j) + v_i(k) \\ &= [z_i(k-1), \cdots, z_i(k-n)][c_1, \cdots, c_n]^T + v_i(k). \end{aligned}$$

For this case, the unknown parameter $x_0 = [c_1, \cdots, c_n]^T$ and the observation matrix (generally called regressor) $H_i(k) = [z_i(k-1), \cdots, z_i(k-n)]$ is an n dimensional row vector. In addition, sensing failures in real networks can be modeled by a Markov chain or an i.i.d. sequence of Bernoulli variables $\{\delta_i(k), k \geq 0\}$. Then $H_i(k) = \delta_i(k)H_i'(k)$, where $\{H_i'(k), k \geq 0\}$ is the sequence of observation matrices without sensing failures.

B. Communication Models

Assume that there exist nonuniform random time-varying communication delays for the communication links between each pair of nodes. We use a sequence of random variables

$\{\lambda_{ji}(k) \in \{0,\cdots,d\},\ k \geq 0\}$ to represent the time delays associated with the link from node j to node i, where the positive integer d represents the maximum time delay. This sequence is subjected to the discrete probability distribution

$$\mathbb{P}\{\lambda_{ji}(k)=q\}=p_{ji,q}(k) \text{ with} \sum_{q=0}^{d} p_{ji,q}(k)=1. \tag{3}$$

We stipulate that $\mathbb{P}\{\lambda_{ii}(k)=0\}=1,\ i=1,\cdots,N,\ k\geq 0$. Denote the N dimensional matrices $\mathcal{I}(k,q)=[\mathcal{I}_{\lambda_{ji}(k),q}]_{1\leq j,i\leq N},\ 0\leq q\leq d,\ k\geq 0$, called *delay matrices*. By the definition of Kronecker function, we know that for each $q=0,1,...,d$, $\{\mathcal{I}(k,q),k\geq 0\}$ is a sequence of random matrices and its sample paths are sequences of $0-1$ matrices. By (3), we know that $\mathbb{E}[\mathcal{I}_{\lambda_{ji}(k),q}]=p_{ji,q}(k)$ and

$$\sum_{q=0}^{d}\mathcal{I}(k,q)=\mathbf{1}_N\mathbf{1}_N^T \text{ a.s.} \tag{4}$$

We use a sequence of random communication graphs $\{\mathcal{G}(k)=\langle\mathcal{V},\ \mathcal{A}_{\mathcal{G}(k)}\rangle,\ k\geq 0\}$ to describe the possible link failures among nodes, where $\mathcal{V}=\{1,\cdots,N\}$ is the node set and $\mathcal{A}_{\mathcal{G}(k)}=[a_{ij}(k)]_{1\leq i,j\leq N}$ is the weighted adjacency matrix of the communication graph, in which $a_{ii}(k)=0$ a.s. for all $i\in\mathcal{V}$ and $k\geq 0$ and $a_{ij}(k)\neq 0$ if and only if the link from node j to node i exists at time instant k for all $i\neq j$. The neighborhood of node i is $\mathcal{N}_i(k)=\{j|a_{ij}(k)\neq 0\}$. The degree matrix of the graph is $\mathcal{D}_{\mathcal{G}(k)}=diag(\sum_{j=1}^N a_{1j}(k),\cdots,\sum_{j=1}^N a_{Nj}(k))$ and the Laplacian matrix of the graph is $\mathcal{L}_{\mathcal{G}(k)}=\mathcal{D}_{\mathcal{G}(k)}-\mathcal{A}_{\mathcal{G}(k)}$ ([36] [37]). Denote $\widehat{\mathcal{L}}_{\mathcal{G}(k)}=\frac{\mathcal{L}_{\mathcal{G}(k)}+\mathcal{L}_{\mathcal{G}(k)}^T}{2}$. Specifically, if $\mathcal{G}(k)$ is balanced, then $\widehat{\mathcal{L}}_{\mathcal{G}(k)}$ is the Laplacian matrix of the symmetrized graph of $\mathcal{G}(k),\ k\geq 0$ ([37]). Let

$$\overline{A}(k,q)=(\mathcal{A}_{\mathcal{G}(k)}\circ\mathcal{I}(k,q))\otimes I_n. \tag{5}$$

Then, by (4) and the above, we have

$$\sum_{q=0}^{d}\overline{A}(k,q)=\mathcal{A}_{\mathcal{G}(k)}\otimes I_n. \tag{6}$$

III. Decentralized Cooperative Online Estimation Algorithm

Let $x_i(k)\in\mathbb{R}^n$ be the estimate by node i for the unknown parameter x_0 at time instant $k,k\geq -d$ with the initial estimates $x_i(k),-d\leq k\leq 0$ being any given real vectors. Starting at the initial estimate, at any time instant $k\geq 0$, node i takes a weighted sum of its own estimate and delayed estimates received from its neighbours, and then adds a correction term based on the local measurement information (innovation) to update the estimate $x_i(k+1)$. Specifically, the decentralized cooperative online parameter estimation algorithm with random observation matrices, communication graphs and time delays, motivated by a baseline version without time delays in [23], is given by

$$\begin{aligned}&x_i(k+1)\\&=x_i(k)+a(k)H_i^T(k)(z_i(k)-H_i(k)x_i(k))\\&+b(k)\sum_{j\in\mathcal{N}_i(k)}a_{ij}(k)(x_j(k-\lambda_{ji}(k))-x_i(k)),\\&i\in\mathcal{V},\ k\geq 0,\end{aligned} \tag{7}$$

where $a(k)$ and $b(k)$ are the innovation and consensus algorithm gains, respectively.

Denote the $\sigma-$fileds $\mathcal{F}(k)=\sigma(\mathcal{A}_{\mathcal{G}(s)},v(s),H_i(s),\lambda_{ji}(s),\ j,i\in\mathcal{V},0\leq s\leq k),\ k\geq 0$, with $\mathcal{F}(-1)=\{\Omega,\emptyset\}$. For the algorithm (7), we have the following assumptions.

A1.a The sequence $\{v(k),k\geq 0\}$ is independent of $\{H(k),\ k\geq 0\}$, $\{\mathcal{A}_{\mathcal{G}(k)},\ k\geq 0\}$ and $\{\lambda_{ji}(k),\ j,i\in\mathcal{V},\ k\geq 0\}$.

A1.b The sequence $\{v(k),\mathcal{F}(k),k\geq 0\}$ is a martingale difference sequence and there exists a constant $\beta_v>0$ such that $\sup_{k\geq 0}\mathbb{E}[\|v(k)\|^2|\mathcal{F}(k-1)]\leq\beta_v$ a.s.

A2.a $\sup_{k\geq 0}\|H(k)\|<\infty$ a.s. and $\sup_{k\geq 0}\|\mathcal{A}_{\mathcal{G}(k)}\|<\infty$ a.s.

A2.b There exist positive constants β_a and β_H such that

$$\max_{i,j\in\mathcal{V}}\sup_{k\geq 0}|a_{ij}(k)|\leq\beta_a \text{ a.s.}, \max_{i\in\mathcal{V}}\sup_{k\geq 0}\|H_i(k)\|\leq\beta_H \text{ a.s.}$$

For the algorithm gains, we make the following conditions.

C1.a The sequences $\{a(k),k\geq 0\}$ and $\{b(k),k\geq 0\}$ are positive real sequences monotonically decreasing to zero, satisfying $a(k)=O(b(k))$.

C1.b $b^2(k)=o(a(k)),a(k)=O(a(k+1))$ and $\sum_{k=0}^{\infty}a(k)=\infty$.

C1.c $\sum_{k=0}^{\infty}b^2(k)<\infty$.

Remark 2: Note that, in Assumption **A1.a**, neither mutual independence nor spatio-temporal independence is assumed on the observation matrices, communication graphs and time delays.

Remark 3: It is easy to find $\{a(k),\ k\geq 0\}$ and $\{b(k),\ k\geq 0\}$ satisfying Conditions **C1.a–C1.c**. If $a(k)=\frac{1}{(k+1)^{\tau_1}}$, $b(k)=\frac{1}{(k+1)^{\tau_2}},\ k\geq 0,\ 0.5<\tau_2\leq\tau_1\leq 1$, then these conditions hold.

By the definition of $\mathcal{I}_{\lambda_{ji}(k),q}$, we know that $x_j(k-\lambda_{ji}(k))=\sum_{q=0}^{d}x_j(k-q)\mathcal{I}_{\lambda_{ji}(k),q}$. Then by (7), we have

$$\begin{aligned}&x_i(k+1)\\&=x_i(k)+a(k)H_i^T(k)[z_i(k)-H_i(k)x_i(k)]+b(k)\\&\times\sum_{j\in\mathcal{N}_i(k)}a_{ij}(k)\left[\sum_{q=0}^{d}x_j(k-q)\mathcal{I}_{\lambda_{ji}(k),q}-x_i(k)\right],\\&i\in\mathcal{V}.\end{aligned} \tag{8}$$

Denote $\mathcal{H}(k)=diag\{H_1(k),\cdots,H_N(k)\}$ and $x(k)=[x_1^T(k),\cdots,x_N^T(k)]^T$. By (5), rewrite (8) as

$$\begin{aligned}&x(k+1)\\&=[I_{Nn}-b(k)\mathcal{D}_{\mathcal{G}(k)}\otimes I_n-a(k)\mathcal{H}^T(k)\mathcal{H}(k)]x(k)\\&+b(k)\sum_{q=0}^{d}\overline{A}(k,q)x(k-q)+a(k)\mathcal{H}^T(k)z(k).\end{aligned} \tag{9}$$

Denote the overall estimation error vector $e(k)=x(k)-\mathbf{1}_N\otimes x_0$. Note that $(\mathcal{L}_{\mathcal{G}(k)}\otimes I_n)(\mathbf{1}_N\otimes x_0)=0$. By (2) and (6), subtracting $\mathbf{1}_N \otimes x_0$ on both sides of (9) leads to

$$\begin{aligned}
&e(k+1)\\
&= [I_{Nn} - b(k)\mathcal{D}_{\mathcal{G}(k)} \otimes I_n - a(k)\mathcal{H}^T(k)\mathcal{H}(k)]x(k)\\
&+b(k)\sum_{q=0}^{d}\overline{A}(k,q)x(k-q) + a(k)\mathcal{H}^T(k)z(k)\\
&-\mathbf{1}_N \otimes x_0\\
&= [I_{Nn} - b(k)\mathcal{D}_{\mathcal{G}(k)} \otimes I_n - a(k)\mathcal{H}^T(k)\mathcal{H}(k)]\\
&\times(x(k) - \mathbf{1}_N \otimes x_0 + \mathbf{1}_N \otimes x_0)\\
&+b(k)\sum_{q=0}^{d}\overline{A}(k,q)(x(k-q) - \mathbf{1}_N \otimes x_0 + \mathbf{1}_N \otimes x_0)\\
&+a(k)\mathcal{H}^T(k)z(k) - \mathbf{1}_N \otimes x_0\\
&= [I_{Nn} - b(k)\mathcal{D}_{\mathcal{G}(k)} \otimes I_n - a(k)\mathcal{H}^T(k)\mathcal{H}(k)]\\
&\times(e(k) + \mathbf{1}_N \otimes x_0) + a(k)\mathcal{H}^T(k)z(k) - \mathbf{1}_N \otimes x_0\\
&+b(k)\sum_{q=0}^{d}\overline{A}(k,q)(e(k-q) + \mathbf{1}_N \otimes x_0)\\
&= [I_{Nn} - b(k)\mathcal{D}_{\mathcal{G}(k)} \otimes I_n - a(k)\mathcal{H}^T(k)\mathcal{H}(k)]e(k)\\
&-(b(k)\mathcal{D}_{\mathcal{G}(k)} \otimes I_n + a(k)\mathcal{H}^T(k)\mathcal{H}(k))(\mathbf{1}_N \otimes x_0)\\
&+b(k)\sum_{q=0}^{d}\overline{A}(k,q)e(k-q) + a(k)\mathcal{H}^T(k)z(k)\\
&+b(k)(\mathcal{A}_{\mathcal{G}(k)} \otimes I_n)(\mathbf{1}_N \otimes x_0)\\
&= [I_{Nn} - b(k)\mathcal{D}_{\mathcal{G}(k)} \otimes I_n - a(k)\mathcal{H}^T(k)\mathcal{H}(k)]e(k)\\
&-a(k)\mathcal{H}^T(k)\mathcal{H}(k)(\mathbf{1}_N \otimes x_0)\\
&+b(k)\sum_{q=0}^{d}\overline{A}(k,q)e(k-q) + a(k)\mathcal{H}^T(k)z(k)\\
&-b(k)(\mathcal{L}_{\mathcal{G}(k)} \otimes I_n)(\mathbf{1}_N \otimes x_0)\\
&= [I_{Nn} - b(k)\mathcal{D}_{\mathcal{G}(k)} \otimes I_n - a(k)\mathcal{H}^T(k)\mathcal{H}(k)]e(k)\\
&-a(k)\mathcal{H}^T(k)\mathcal{H}(k)(\mathbf{1}_N \otimes x_0) + a(k)\mathcal{H}^T(k)v(k)\\
&+a(k)\mathcal{H}^T(k)H(k)x_0 + b(k)\sum_{q=0}^{d}\overline{A}(k,q)e(k-q),
\end{aligned}$$

which together with $\mathcal{H}(k)(\mathbf{1}_N \otimes x_0) = H(k)x_0$ gives the overall estimation error equation

$$\begin{aligned}
&e(k+1)\\
&= [I_{Nn} - b(k)\mathcal{D}_{\mathcal{G}(k)} \otimes I_n - a(k)\mathcal{H}^T(k)\mathcal{H}(k)]e(k)\\
&+b(k)\sum_{q=0}^{d}\overline{A}(k,q)e(k-q) + a(k)\mathcal{H}^T(k)v(k). \qquad (10)
\end{aligned}$$

For the delay-free case, $d = 0$. Then the algorithm (9) becomes

$$\begin{aligned}
&x(k+1)\\
&= [I_{Nn} - b(k)\mathcal{L}_{\mathcal{G}(k)} \otimes I_n - a(k)\mathcal{H}^T(k)\mathcal{H}(k)]x(k)\\
&+a(k)\mathcal{H}^T(k)z(k), \qquad (11)
\end{aligned}$$

and the estimation error equation (10) becomes

$$\begin{aligned}
&e(k+1)\\
&= [I_{Nn} - b(k)\mathcal{L}_{\mathcal{G}(k)} \otimes I_n - a(k)\mathcal{H}^T(k)\mathcal{H}(k)]e(k)\\
&+a(k)\mathcal{H}^T(k)v(k). \qquad (12)
\end{aligned}$$

Remark 4: In this paper, we use the concept of Laplacians of digraphs defined in [37], which is widely used in the literature on decentralized estimation ([11]–[24]). For the delay-free case, the consensus term $[\mathcal{L}_{\mathcal{G}(k)} \otimes I_n]x(k)$ naturally appears in the algorithm (11). Note that another concept of symmetric Laplacians of digraphs is proposed in [35]. This symmetric Laplacian involves the Perron vector of the weighted adjacency matrix. It has been pointed out in [35] that for a general digraph, there is no closed form solution for the Perron vector. Generally, the ith element of the Perron vector, which is not local information of the ith node, depends on the weights associated to all nodes. Therefore, though the Laplacian proposed in [35] is symmetric, it is generally incompatible with the decentralized nature of the estimation algorithm.

IV. The Delay-Free Case

In this section, we give the convergence conditions of the algorithm (7) for the delay-free case, i.e. $\lambda_{ji}(k) = 0$, a.s. $\forall\ j, i \in \mathcal{V}$, $\forall\ k \geq 0$. All proofs of results are put in Appendix B.

For any given positive integers h and m, denote

$$\Lambda_m^h = \lambda_{\min}\left[\sum_{k=mh}^{(m+1)h-1}\Big(\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}|\mathcal{F}(mh-1)] \otimes I_n + \mathbb{E}[\mathcal{H}^T(k)\mathcal{H}(k)|\mathcal{F}(mh-1)]\Big)\right],$$

$$\overline{\Lambda}_m^h = \lambda_{\min}\left[\sum_{k=mh}^{(m+1)h-1}\Big(b(k)\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}|\mathcal{F}(mh-1)] \otimes I_n + a(k)\mathbb{E}[\mathcal{H}^T(k)\mathcal{H}(k)|\mathcal{F}(mh-1)]\Big)\right].$$

We first give a result for the case with general processes of random graphs and observation matrices.

Theorem IV.1: Suppose that Assumptions **A1.a–A1.b** hold. If Condition **C1.a** holds, and there exists an integer $h > 0$, a constant $\rho_0 > 0$ and a positive real sequence $\{c(m), m \geq 0\}$ with

$$b^2(mh) = o(c(m)),\ \sum_{m=0}^{\infty} c(m) = \infty, \qquad (13)$$

such that (b.1) $\overline{\Lambda}_m^h \geq c(m)$ a.s., $m \geq 0$ and (b.2) $\sup_{k\geq 0} [\mathbb{E}[(\|\mathcal{L}_{\mathcal{G}(k)}\| + \|\mathcal{H}^T(k)\mathcal{H}(k)\|)^{2^{\max\{h,2\}}}|\mathcal{F}(k-1)]]^{\frac{1}{2^{\max\{h,2\}}}} \leq \rho_0$ a.s., then the algorithm (7) converges in mean square, that is, $\lim_{k\to\infty}\mathbb{E}\|x_i(k) - x_0\|^2 = 0, i \in \mathcal{V}$. In addition, if Assumption **A2.a** and Condition **C1.c** hold, then the algorithm (7) converges almost surely, i.e. $\lim_{k\to\infty} x_i(k) = x_0, i \in \mathcal{V}$ a.s.

Remark 5: Most existing literature on decentralized estimation suppose that the mean graphs are balanced ([22], [24]). Here, the condition (b.1) in Theorem IV.1 may still hold even if the mean graphs are unbalanced. For example, consider a fixed weighted graph $\mathcal{G} = \langle\mathcal{V} = \{1, 2\}, \mathcal{A}_{\mathcal{G}} = [a_{ij}]_{2\times 2}\rangle$ with $a_{12} = 1$ and $a_{21} = 0.3$. Obviously, $\mathcal{G}$ is unbalanced. Suppose $H_1 = 0, H_2 = 1$. Choose $a(k) = b(k) = \frac{1}{k+1}$. We have

$\lambda_{\min}(b(m)\widehat{\mathcal{L}}_{\mathcal{G}} + a(m)\mathcal{H}^T\mathcal{H}) = \frac{1}{m+1}\lambda_{\min}(\widehat{\mathcal{L}}_{\mathcal{G}} + \mathcal{H}^T\mathcal{H}) = \frac{0.5821}{m+1}$. Then, the condition (b.1) holds with $h = 1$ and $c(m) = \frac{0.5821}{m+1}$ satisfying (13). A more complex example with unbalanced mean graphs is given in Section VI.

Next, we give Theorem IV.2 for the case with conditionally balanced digraphs:

$$\Gamma_1 = \Big\{\{\mathcal{G}(k), k \geq 0\} | \mathbb{E}[\mathcal{A}_{\mathcal{G}(k)}|\mathcal{F}(k-1)] \text{ is nonnegative and its associated random graph is balanced a.s.}, k \geq 0\Big\}.$$

Theorem IV.2: Suppose that $\{\mathcal{G}(k), k \geq 0\} \in \Gamma_1$, Assumptions **A1.a–A1.b** hold. If Conditions **C1.a–C1.b** hold, and there exists an integer $h > 0$, positive constants θ and ρ_0 such that (c.1) $\inf_{m\geq 0}\Lambda_m^h \geq \theta > 0$ a.s. and (c.2) $\sup_{k\geq 0}[\mathbb{E}[(\|\mathcal{L}_{\mathcal{G}(k)}\| + \|\mathcal{H}^T(k)\mathcal{H}(k)\|)^{2^{\max\{h,2\}}}|\mathcal{F}(k-1)]]^{\frac{1}{2^{\max\{h,2\}}}} \leq \rho_0$ a.s., then the algorithm (7) converges in mean square. In addition, if Assumption **A2.a** and Condition **C1.c** hold, then the algorithm (7) converges almost surely.

Remark 6: The condition (b.1) in Theorem IV.1 and the condition (c.1) in Theorem IV.2 are the key convergence conditions. We call them the *stochastic spatio-temporal persistence of excitation* conditions. In detail, *spatio* emphasizes the reliance of the conditions on the communication graphs and observation matrices over all nodes rather than a single node, while *temporal* represents the summing matrices over a sequence of fixed-length time intervals rather than a single time step, and "persistence of excitation" represents that the minimum eigenvalues of matrices consisting of spatio-temporal observation matrices and Laplacian matrices are uniformly bounded away from zero with respect to the sample paths in some sense. Guo [38] considered centralized estimation algorithms with random observation matrices and proposed the "stochastic persistence of excitation" condition to ensure convergence. The condition (c.1) can be regarded as the generalization of "stochastic persistence of excitation" condition in [38] to that for decentralized algorithms. For a network with N isolated nodes, $\mathcal{L}_{\mathcal{G}(k)} \equiv \mathbf{0}_{N\times N}$ a.s., and the condition (c.1) degenerates to N independent "stochastic persistence of excitation" conditions.

In the most existing literature, it was also required that the sequence of observation matrices be i.i.d. and independent of the sequence of communication graphs, neither of which is necessary in Theorems IV.1 and IV.2. Subsequently, we give more intuitive convergence conditions for the case with Markovian switching communication graphs and observation matrices. We first make the following assumption.

A3 $\{\langle\mathcal{H}(k), \mathcal{A}_{\mathcal{G}(k)}\rangle, k \geq 0\} \subseteq \mathcal{S}$ is a homogeneous and uniform ergodic Markov chain with a unique stationary distribution π.

Here, the set $\mathcal{S} = \{\langle\mathcal{H}_l, \mathcal{A}_l\rangle, l = 1, 2, ...\}$ with $\mathcal{H}_l = diag(H_{1,l}, \cdots, H_{N,l})$, where $\{H_{i,l} \in \mathbb{R}^{n_i\times n}, l = 1, 2, ...\}$ is the state space of observation matrices of node i and $\{\mathcal{A}_l, l = 1, 2, ...\}$ being the state space of the weighted adjacency matrices, $\pi = [\pi_1, \pi_2, ...]^T$, $\pi_l \geq 0$, $l = 1, 2, ...$, and $\sum_{l=1}^{\infty}\pi_l = 1$ with π_l representing $\pi(\langle\mathcal{H}_l, \mathcal{A}_l\rangle)$.

Corollary IV.1: Suppose that Assumptions **A1.a–A1.b**, **A3** hold, and $\sup_{l\geq 1}\|\mathcal{A}_l\| < \infty$, $\sup_{l\geq 1}\|\mathcal{H}_l\| < \infty$. If Conditions **C1.a–C1.c** hold, and

(d.1) the stationary weighted adjacency matrix $\sum_{l=1}^{\infty}\pi_l\mathcal{A}_l$ is nonnegative and its associated graph is balanced with a spanning tree;

(d.2) the measurement model (1) is *spatio-temporally jointly observable*, i.e.

$$\lambda_{\min}\left(\sum_{i=1}^{N}\left(\sum_{l=1}^{\infty}\pi_l H_{i,l}^T H_{i,l}\right)\right) > 0, \tag{14}$$

then the algorithm (7) converges in mean square and almost surely.

Remark 7: Most of the existing decentralized estimation algorithms used the mathematical expectation of observation matrices which is restricted to be time-invariant and difficult to be obtained ([22], [24]). They required instantaneous global observability in the statistical sense for the measurement model, i.e., $\sum_{i=1}^{N}\overline{H}_i^T\overline{H}_i$ is positive definite, where $\overline{H}_i$ is a fixed matrix with $\mathbb{E}[H_i(k)] \equiv \overline{H}_i$, for all $k \geq 0$, $i = 1, 2, ..., N$. In contrast, we only use the sample paths of observation matrices in the algorithm (7). The mathematical expectations of observation matrices are allowed to be time-varying. We prove that for homogeneous and uniform ergodic Markovian switching observation matrices and communication graphs, the *stochastic spatio-temporal persistence of excitation* condition given in Theorem IV.2 holds if the stationary graph is balanced with a spanning tree and the measurement model is spatio-temporally jointly observable, that is, (14) holds, implying that neither local observability of each node, i.e. $\lambda_{\min}(\sum_{l=1}^{\infty}\pi_l H_{i,l}^T H_{i,l}) > 0$, $i \in \mathcal{V}$, nor instantaneous global observability of the entire measurement model, i.e. $\lambda_{\min}(\sum_{i=1}^{N}H_{i,l}^T H_{i,l}) > 0$, $l = 1, 2, ...$, is needed.

V. The Case With Random Time-Varying Communication Delays

In this section, we analyze the convergence of the algorithm (7) with random observation matrices, communication graphs and time delays simultaneously. All proofs of results are put in Appendix C.

In the presence of random time-varying communication delays, the mean square convergence analysis of the algorithm becomes very difficult. To address this, we transform the estimation error equation (10) into the following equivalent system ([32], [33]).

$$\begin{aligned} r(k+1) &= F(k)r(k) + g(k), \\ g(k) &= \sum_{q=1}^{d} C_q(k)g(k-q) + a(k)\mathcal{H}^T(k)v(k), \\ & \quad k \geq 0, \end{aligned} \tag{15}$$

where $F(k)$, $C_q(k)$, $1 \leq q \leq d$, $k \geq 0$ satisfy

$$\begin{aligned} & F(k) + C_1(k) = I_{Nn} - b(k)\mathcal{D}_{\mathcal{G}(k)} \otimes I_n \\ & -a(k)\mathcal{H}^T(k)\mathcal{H}(k) + b(k)\overline{A}(k,0), \\ & C_1(k)F(k-1) - C_2(k) = -b(k)\overline{A}(k,1), \\ & C_2(k)F(k-2) - C_3(k) = -b(k)\overline{A}(k,2), \end{aligned}$$

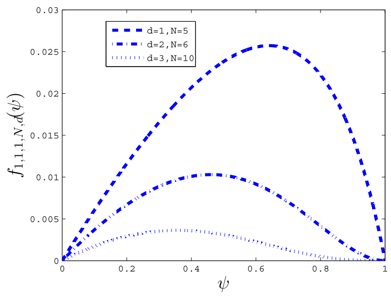

Fig. 1. The curves of $f_{1,1,1,N,d}(\cdot)$.

$$\vdots$$
$$C_{d-1}(k)F(k-d+1)-C_d(k)=-b(k)\overline{A}(k,d-1),$$
$$C_d(k)F(k-d)=-b(k)\overline{A}(k,d). \quad (16)$$

Here, $F(k)=I_{Nn}$, $-d\le k\le -1$. It can be verified that if $r(k)=e(k)$, $-d\le k\le -1$, then $r(k)=e(k)$, $\forall\ k\ge 0$, i.e. the system (10) and the system (15)-(16) are equivalent.

We need the following condition on the consensus gain.

C1.d The initial gain $b(0)\le \max_{0<\psi<1} f_{C_1,\beta_a,\beta_H,N,d}(\psi)$, where $f_{C_1,\beta_a,\beta_H,N,d}(\psi)\triangleq\frac{\psi}{N\beta_a+C_1\beta_H^2+\frac{N\beta_a[(1-\psi)^{-(d+1)}-1]}{(1-\psi)^{-1}-1}}$, $d\ge 1,\psi\in(0,1)$, with $C_1\triangleq\sup_{k\ge 0}\frac{a(k)}{b(k)}$.

It can be verified that given Assumption **A2.b** and Condition **C1.a**, $\max_{0<\psi<1} f_{C_1,\beta_a,\beta_H,N,d}(\psi)$ is well-defined. Examples of $f_{1,1,1,N,d}(\cdot)$ with different d and N are shown in Figure 1.

We first establish a lemma as the basis of convergence analysis.

Lemma V.1: If Assumption **A2.b**, Conditions **C1.a** and **C1.d** hold, then $F(k)$ is invertible and $\|F^{-1}(k)\|\le(1-\psi_1)^{-1}$ a.s., $\forall k\ge 0$, where $\psi_1=\min\{\psi\in(0,1)|f_{C_1,\beta_a,\beta_H,N,d}(\psi)\ge b(0)\}$.

Note that, by the continuity of the function $f_{C_1,\beta_a,\beta_H,N,d}(\cdot)$ and Conditon **C1.d**, it is known that the set $\{\psi\in(0,1)|f_{C_1,\beta_a,\beta_H,N,d}(\psi)\ge b(0)\}$ is a nonempty and bounded closed set. Then, ψ_1 is well-defined.

If the conditions of Lemma V.1 hold, then $F(k)$ is invertible a.s. Thus, by (16), we have

$$\begin{aligned}
&F(k)\\
&=I_{Nn}-b(k)\mathcal{D}_{\mathcal{G}(k)}\otimes I_n-a(k)\mathcal{H}^T(k)\mathcal{H}(k)\\
&+b(k)\overline{A}(k,0)-C_1(k)\\
&=I_{Nn}-b(k)\mathcal{D}_{\mathcal{G}(k)}\otimes I_n-a(k)\mathcal{H}^T(k)\mathcal{H}(k)\\
&+b(k)\overline{A}(k,0)-(C_2(k)-b(k)\overline{A}(k,1))F^{-1}(k-1)\\
&\vdots\\
&=I_{Nn}-G(k),k\ge 0, \quad (17)
\end{aligned}$$

where

$$\begin{aligned}
G(k)\ \triangleq\ &b(k)\mathcal{D}_{\mathcal{G}(k)}\otimes I_n+a(k)\mathcal{H}^T(k)\mathcal{H}(k)\\
&-b(k)\sum_{q=0}^{d}\overline{A}(k,q)\big[\Phi_F(k-1,k-q)\big]^{-1}. \quad (18)
\end{aligned}$$

For any given positive integers h and m, denote

$$\begin{aligned}
&\widetilde{\Lambda}_m^h\\
&=\lambda_{\min}\Bigg[\sum_{k=mh}^{(m+1)h-1}\bigg(b(k)\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}|\mathcal{F}(mh-1)]\otimes I_n\\
&+a(k)\mathbb{E}[\mathcal{H}^T(k)\mathcal{H}(k)|\mathcal{F}(mh-1)]\\
&-\frac{b(k)}{2}\sum_{q=0}^{d}\mathbb{E}\Big[\overline{A}(k,q)[[\Phi_F(k-1,k-q)]^{-1}-I_{Nn}]\\
&+[[\Phi_F(k-1,k-q)]^{-1}-I_{Nn}]^T\\
&\times\overline{A}^T(k,q)\Big|\mathcal{F}(mh-1)\Big]\bigg)\Bigg]. \quad (19)
\end{aligned}$$

Theorem V.1: Suppose that Assumptions **A1.a–A1.b**, **A2.b** hold. If Conditions **C1.a**, **C1.d** hold, and there exists an integer $h>0$ and a positive real sequence $\{c(m),m\ge 0\}$ with $b^2(mh)=o(c(m))$, $\sum_{m=0}^{\infty}c(m)=\infty$, such that

$$\widetilde{\Lambda}_m^h\ge c(m)\ \text{a.s.},\ \ m\ge 0, \quad (20)$$

then the algorithm (7) converges in mean square.

If $\{\langle\mathcal{H}(k),\mathcal{A}_{\mathcal{G}(k)},\lambda_{ji}(k),j,i\in\mathcal{V}\rangle,k\ge 0\}$ is an independent random process, then Corollary V.1 below gives a sufficient condition for the condition (20) in Theorem V.1 to hold, which is more intuitive and computable.

Corollary V.1: Suppose that Assumptions **A1.a–A1.b**, **A2.b** hold, $\{\langle\mathcal{H}(k),\mathcal{A}_{\mathcal{G}(k)},\lambda_{ji}(k),j,i\in\mathcal{V}\rangle,k\ge 0\}$ is an independent process. If Condition **C1.a** holds, $b(0)\le f_{C_1,\beta_a,\beta_H,N,d}(\psi_2)$ with $\psi_2\in(0,2^{\frac{1}{d}}-1)$, and there exists an integer $h>0$ and a positive real sequence $\{c(m),m\ge 0\}$ with $b^2(mh)=o(c(m))$ and $\sum_{m=0}^{\infty}c(m)=\infty$, such that

$$\begin{aligned}
&\overline{\Lambda}_m^h-\sum_{k=mh}^{(m+1)h-1}\left[b(k)\sum_{q=0}^{d}\frac{\|\mathbb{E}[\overline{A}(k,q)]\|[(1+\psi_2)^q-1]}{2-(1+\psi_2)^q}\right]\\
&\ge c(m),\ m\ge 0, \quad (21)
\end{aligned}$$

then the algorithm (7) converges in mean square.

Next, for the case with conditionally balanced digraphs, the following corollary presents a more intuitive convergence condition.

Corollary V.2: Suppose that Assumptions **A1.a–A1.b**, **A2.b** hold and $\{\mathcal{G}(k),k\ge 0\}\in\Gamma_1$. If Conditions **C1.a–C1.b**, **C1.d** hold, $b(k)=O(a(k))$, and there exists an integer $h>0$, a constant $\theta>0$ such that

$$\inf_{m\ge 0}(\Lambda_m^h-\Sigma_m^h)\ge\theta\ \text{a.s.} \quad (22)$$

where

$$\begin{aligned}
\Sigma_m^h\ =\ &C_2(C_3)^h\max\{1,C_1\}\sum_{k=mh}^{(m+1)h-1}\sum_{q=0}^{d}\|\mathbb{E}[\overline{A}(k,q)\\
&\times([\Phi_F(k-1,k-q)]^{-1}-I_{Nn})|\mathcal{F}(mh-1)]\|
\end{aligned}$$

with $C_2\triangleq\sup_{k\ge 0}\frac{b(k)}{a(k)}$ and $C_3\triangleq\sup_{k\ge 0}\frac{a(k)}{a(k+1)}$, then, the algorithm (7) converges in mean square. Furthermore, if $\{\langle\mathcal{H}(k),\mathcal{A}_{\mathcal{G}(k)},\lambda_{ji}(k),j,i\in\mathcal{V}\rangle,k\ge 0\}$ is an independent process, then (22) holds if there exist an integer $h > 0$ such that

$$\inf_{m\geq 0} \Lambda_m^h > C_2(C_3)^h \max\{1, C_1\} \sup_{m\geq 0} \left[\sum_{k=mh}^{(m+1)h-1} \sum_{q=0}^{d} \left(\|\mathbb{E}[\overline{A}(k,q)]\| \frac{[(1+\psi_2)^q - 1]}{2-(1+\psi_2)^q} \right) \right], \quad (23)$$

and $b(0) \leq f_{C_1,\beta_a,\beta_H,N,d}(\psi_2)$, with $\psi_2 \in (0, 2^{\frac{1}{d}} - 1)$, where C_1 is defined in Condition **C1.d**.

Remark 8: Theorem V.1, Corollaries V.1-V.2 give explicit convergence conditions under which all nodes' estimates converge to the true parameter in mean square. Existing literature used the Lyapunov-Krasovskii functional method to deal with time delays and obtained the non-explicit LMI type convergence condition ([30]). In contrast, here, we transform the system with random time-varying communication delays into an equivalent delay-free system by introducing an auxiliary system and then adopt the method of binomial expansion of random matrix products to transform the mean square convergence analysis of the delay-free system into that of the mathematical expectation of random matrix products, and obtain the key convergence conditions (20)-(22) which explicitly rely on the conditional expectations of delay matrices, observation matrices and weighted adjacency matrices of communication graphs over a sequence of fixed-length time intervals. In the absence of time delays, the condition (20) degenerates to the condition (b.1) in Theorem IV.1.

Remark 9: The conditions (21) and (23) can be further simplified for special delay processes. If the delays are independent of the graphs, then $\mathbb{E}[\overline{A}(k,q)] = \mathbb{E}[\mathcal{A}_{\mathcal{G}(k)}] \circ \mathbb{E}[\mathcal{I}(k,q)]$. Here, the element in the ith row and the jth column of $\mathbb{E}[\mathcal{I}(k,q)]$, $\mathbb{E}[\mathcal{I}(k,q)]_{ij} = \mathbb{P}\{\lambda_{ji}(k) = q\} = p_{ji,q}(k)$. In addition,

- if $\lambda_{ji}(k)$ are identically distributed w.r.t. k, then $\mathbb{E}[\mathcal{I}(k,q)]_{ij} = p_{ji,q}(0), \forall k \geq 0$;
- if $\lambda_{ji}(k)$ are identically distributed w.r.t. both k and (j,i), then $\mathbb{E}[\mathcal{I}(k,q)]_{ij} = p_q, i \neq j$ where p_q denotes the probability that the packet is delayed by q steps for all k and $(j,i), j \neq i$. Therefore, $\|\mathbb{E}[\overline{A}(k,q)]\| = p_q \|\mathbb{E}[\mathcal{A}_{\mathcal{G}(k)}]\|$. Furthermore, if the graph sequence is an i.i.d. process, then the condition (23) becomes

$$\inf_{m\geq 0} \Lambda_m^h > C_2(C_3)^h \max\{1, C_1\} h \|\mathbb{E}[\mathcal{A}_{\mathcal{G}(0)}]\| \times \sum_{q=0}^{d} \frac{p_q[(1+\psi_2)^q - 1]}{2-(1+\psi_2)^q}.$$

Corollaries V.1-V.2 show that for given algorithm gains $\{a(k), k \geq 0\}$ and $\{b(k), k \geq 0\}$, if the communication graphs and observation matrices are persistently excited with enough intensity, then the additional effects of time delays can be mitigated. The maximum delay bound d that can be allowed is related to the weighted adjacency matrix of mean graphs $\mathbb{E}[\mathcal{A}_{\mathcal{G}(k)}]$, the probability distribution of time delays $\mathbb{E}[\mathcal{I}(k,q)]$ and the algorithm gains. In the absence of time delays, (22) degenerates to the condition (c.1) in Theorem IV.2. The following corollary shows that for the case with conditionally balanced graphs, if the *stochastic spatio-temporal persistence of excitation* condition $\inf_{m\geq 0} \Lambda_m^h \geq \theta$ a.s. holds, then for any given bounded delays, mean square convergence of the algorithm can be guaranteed if the algorithms gains are properly designed and sufficiently small.

Corollary V.3: Suppose that Assumptions **A1.a–A1.b**, **A2.b** hold, $\{\mathcal{G}(k), k \geq 0\} \in \Gamma_1$ and there exists an integer $h > 0$, a constant $\theta > 0$ such that $\inf_{m\geq 0} \Lambda_m^h \geq \theta$ a.s. If Conditions **C1.a–C1.b** hold, $b(k) = O(a(k))$, and $b(0) \leq f_{C_1,\beta_a,\beta_H,N,d}(\psi_3)$ with $\psi_3 \in (0, (1 + \theta/[\theta + NC_2(C_3)^h \max\{1, C_1\}\beta_a dh])^{\frac{1}{d}} - 1)$, then the algorithm (7) converges in mean square.

VI. Numerical Example

We apply our algorithm to decentralized multi-area online state estimation in power systems to illustrate the effectiveness of the obtained theoretical results. An IEEE 14-bus system is used for the test, which has 14 buses and is partitioned into 4 areas A_1, A_2, A_3, A_4, shown in Figure 2. After a DC power flow approximation ([34]), the grid state to be estimated degenerates into a vector of voltage phase angles at all buses. Let bus 1's voltage phase angle be zero, as the reference bus. The grid state to be estimated is given by

$$x_0 = [-4.98, -12.72, -11.33, -8.78, -14.22, -13.37, -13.36, -14.94, -15.10, -14.79, -15.05, -15.12, -16.03]^T.$$

The measurements $z_i(k)$ are linearly related to x_0, given by $z_i(k) = s_i(k)H_i'x_0 + v_i(k),\ i = 1,2,3,4$, where the noise $\{v_i(k), k \geq 0\}$ is assumed to be an i.i.d. process with the standard normal distribution, $\{s_i(k), k \geq 0\}$ is an i.i.d. sequence, modelling the sensing failures with $\mathbb{P}\{s_i(k) = 1\} = \mathbb{P}\{s_i(k) = 0\} = 0.5$, and H_i', $i = 1, ..., 4$ are the observation matrices, which are deterministic and given in Appendix D. There are 4 random communication links with $0-1$ weights, represented by the red dotted lines in Figure 2. At odd time instants, the link from A_2 to A_3 awakes with the probability 0.5 and the others sleep; at even time instants, the link from A_2 to A_3 sleeps and the others awake with the probability 0.5. Both $\{\mathcal{G}(k), k \geq 0\}$ and $\{H(k), k \geq 0\}$ are independent processes. We use the averaged relative error, $\frac{\sum_{i=1}^4 \|x_i(k)-x_0\|}{4\|x_0\|}$, to evaluate the performance of the algorithm.

For the delay-free case, set $a(k) = b(k) = \frac{0.5}{(k+1)^{0.52}}$. Let $c(m) = \frac{0.0112}{(2m+2)^{0.52}}$. When $h = 2$, we plot the curves of $\overline{\Lambda}_m^2$ and $c(m)$ w.r.t. m in Figure 3, which shows that $\overline{\Lambda}_m^2 \geq c(m), m \geq 0$. The conditions of Theorem IV.1 hold. Figure 4 is depicted with the curves of the averaged relative errors, where the red line represents the error curve of the algorithm without random link failures and sensing failures, as the base case. It shows that in spite of the unbalance of the mean graphs and the sensing failures, the four areas' estimates converge to x_0. For the case with time delays, assume that the delays are independent of the communication graphs, observation matrices and measurement noises, and subjected

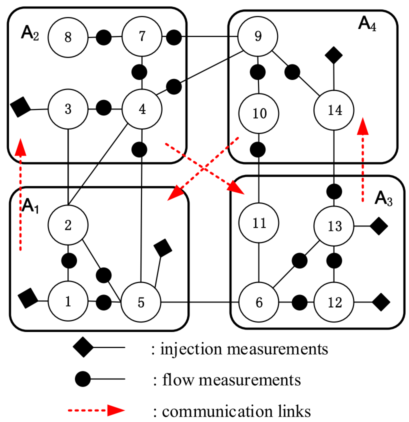

Fig. 2. IEEE-14 multi-area buses and the communication graphs.

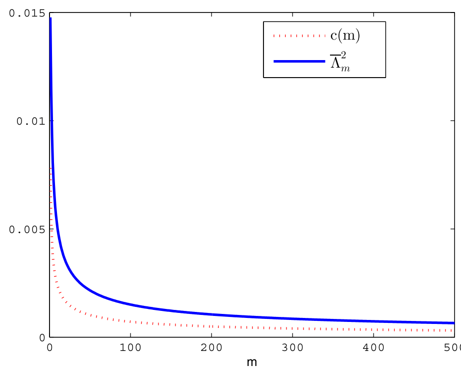

Fig. 3. Curves of $\overline{\Lambda}_m^2$ and $c(m)=\frac{0.0112}{(2m+2)^{0.52}}$ w.r.t. m.

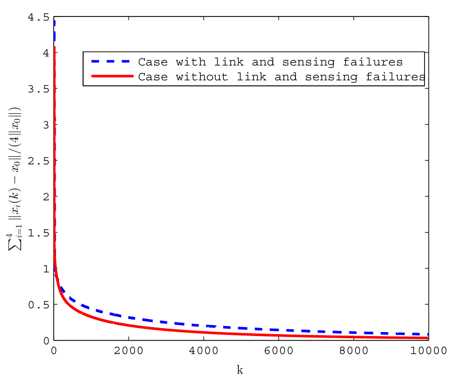

Fig. 4. Curves of the averaged relative error $\frac{\sum_{i=1}^4\|x_i(k)-x_0\|}{4\|x_0\|}$ for the delay-free case.

to the Bernoulli distribution, i.e. $\lambda_{ji}(k)\sim B(d,p)$ for all k and (j,i). Then

$$\mathbb{P}\{\lambda_{ji}(k)=q\}=\mathbb{C}_d^q p^q(1-p)^{d-q}, q=0,\cdots,d. \tag{24}$$

Set $d=4, p=0.4$. We now verify the convergence conditions in Corollary V.1. Let $a(k)=b(k)$. Then $C_1=1$. By the above settings of communication graphs and observation matrices, we know that $\beta_a=1$,$\beta_H=4.07$. Let $\psi_2=0.01$. Then we have $f_{C_1,\beta_a,\beta_H,N,d}(\psi_2)=0.0005$. Then, let $b(k)=\frac{0.0005}{(k+1)^{0.1}}$. By the definition of p_q in Remark 9, it follows from (24) that $p_q=\mathbb{C}_4^q 0.4^q 0.6^{4-q}, q=0,...,4$. Note that $\|\mathbb{E}[\mathcal{A}_{\mathcal{G}(k)}]\|\equiv 0.5$. As is discussed in Remark 9, $\|\mathbb{E}[\overline{A}(k,q)]\|=p_q\|\mathbb{E}[\mathcal{A}_{\mathcal{G}(k)}]\|$. Hence, it can be calculated that

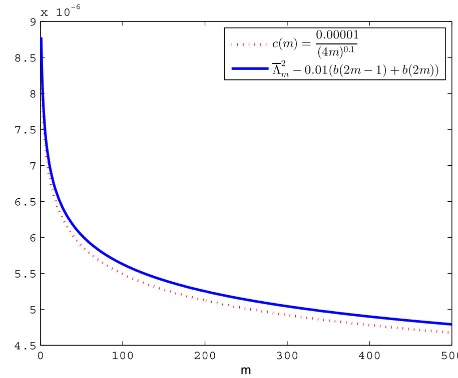

Fig. 5. Curves of $\overline{\Lambda}_m^2-0.01\sum_{k=2m}^{2m+1}b(k)$ and $c(m)$ w.r.t. m.

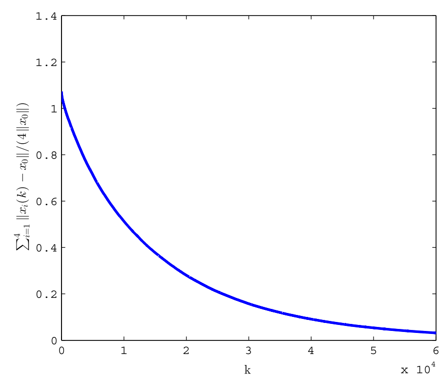

Fig. 6. Curve of the averaged relative error for the case with random time delays.

$\sum_{k=mh}^{(m+1)h-1}b(k)\sum_{q=0}^d\|\mathbb{E}[\overline{A}(k,q)]\|\frac{[(1+\psi_2)^q-1]}{2-(1+\psi_2)^q}=\sum_{k=2m}^{2m+1}b(k)\sum_{q=0}^4\frac{0.5p_q[(1+\psi_2)^q-1]}{2-(1+\psi_2)^q}=0.01\sum_{k=2m}^{2m+1}b(k)$. Note that $\overline{\Lambda}_m^2=\lambda_{\min}[\sum_{k=2m}^{2m+1}b(k)(\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}]\otimes I_{13}+\mathbb{E}[\mathcal{H}^T(k)\mathcal{H}(k)])]$. Let $c(m)=\frac{0.00001}{(4m+4)^{0.1}}$. We plot the curves of $\overline{\Lambda}_m^2-0.01\sum_{k=2m}^{2m+1}b(k)$ and $c(m)$ w.r.t. m in Figure 5, showing that the condition (21) in Corollary V.1 holds. Figure 6 is depicted with curve of the averaged relative error, which confirms Corollary V.1.

VII. CONCLUSION

In this paper, we analyzed the convergence of the decentralized cooperative online parameter estimation algorithm in an uncertain communication environment. Each node has a partial linear observation of the unknown parameter with random time-varying observation matrices. The underlying communication network is modeled by a sequence of random digraphs and is subjected to nonuniform random time-varying delays in channels.

For the delay-free case, we proved that if the observation matrices and the graph sequence satisfy the *stochastic spatio-temporal persistence of excitation* condition, then the algorithm gains can be designed properly such that all nodes' estimates converge to the true parameter in mean square and almost surely. Specially, for Markovian switching communication graphs and observation matrices, this condition holds if the stationary graph is balanced with a spanning tree and the measurement model is spatio-temporally jointly observable. For the case with communication delays, we introduced delay matrices to model the random time-varying communication delays, adopted the method of binomial expansion of random matrix products to transform the mean square convergence analysis of the algorithm into that of the mathematical expectation of random matrix products, and obtained mean square convergence conditions explicitly relying on the conditional expectations of delay matrices, observation matrices and weighted adjacency matrices of communication graphs over a sequence of fixed-length intervals. In the absence of time delays, these mean square convergence conditions degenerate to the *stochastic spatio-temporal persistence of excitation* conditions. Especially, given that the digraphs are conditionally balanced, we show that if the *stochastic spatio-temporal persistence of excitation* condition holds, then for any given bounded delay, proper algorithm gains can be designed to guarantee mean square convergence of the algorithm.

There are many interesting open issues for future research. Theorem V.1 is established for a very general type of delays, namely random and unordered. This means that in the practical implementation, the packets of information exchanged by pairs of nodes are being placed in a processing queue without any regard to their transmit time stamp. In some cases, all received packets are ordered by the time stamp of their transmission, and the communication delays would be random and monotone ([32], [44], [45]). How to explore monotonicity constraints in the random delay process to relax the conditions or strengthen the results of Theorem V.1 would be an interesting and challenging issue. The main obstacle is how to deal with the delay-induced products of the inverse of matrices, which is difficult and may need more advanced techniques. Another important issue is the convergence rate of the algorithm. Especially, Corollary V.3 shows that for the case with conditionally balanced graphs, if the *stochastic spatio-temporal persistence of excitation* condition holds, then for any given bounded delays, mean square convergence of the algorithm can be guaranteed if we choose sufficiently small algorithms gains. However, smaller algorithm gains generally lead to a slower convergence. Thus, how to choose the algorithms gains for optimizing the convergence rate is an interesting topic for future investigation.

Appendix A
Several Useful Lemmas

Definition A.1 ([39]): A Markov chain on a countable state space $\mathcal{S}$ with a stationary distribution π and transition function $\mathbb{P}(x,\cdot)$ is called uniform ergodic, if there exist positive constants $r>1$ and R such that for all $x\in\mathcal{S}$, $\|\mathbb{P}^n(x,\cdot)-\pi\|\le Rr^{-n}$. Here, $\|\mathbb{P}^n(x,\cdot)-\pi\|=\sum_y|\mathbb{P}^n(x,y)-\pi_y|$.

Lemma A.1 ([40]): For any given matrix P, denote $W=I-P$. If there exists a constant $\psi\in(0,1)$ such that $\|P\|\le\psi$, then W is invertible and $\|W^{-1}\|\le(1-\|P\|)^{-1}\le(1-\psi)^{-1}$.

Lemma A.2 ([41]): Assume that $\{s_1(k),k\ge 0\}$ and $\{s_2(k),k\ge 0\}$ are real sequences satisfying $0\le s_2(k)<1$, $\sum_{k=0}^{\infty}s_2(k)=\infty$ and $\lim_{k\to\infty}\frac{s_1(k)}{s_2(k)}$ exists. Then

$$\lim_{k\to\infty}\sum_{i=1}^{k}s_1(i)\prod_{l=i+1}^{k}(1-s_2(l))=\lim_{k\to\infty}\frac{s_1(k)}{s_2(k)}.$$

Lemma A.3 ([42]): Assume that $\{x(k),\mathcal{F}(k)\}$, $\{\alpha(k),\mathcal{F}(k)\}$, $\{\beta(k),\mathcal{F}(k)\}$ and $\{\gamma(k),\mathcal{F}(k)\}$ are all nonnegative adaptive sequences, satisfying

$$\mathbb{E}[x(k+1)|\mathcal{F}(k)]\le(1+\alpha(k))x(k)-\beta(k)+\gamma(k),k\ge 0 \text{ a.s.}$$

If $\sum_{k=0}^{\infty}(\alpha(k)+\gamma(k))<\infty$ a.s., then $x(k)$ converges to a finite random variable a.s. and $\sum_{k=0}^{\infty}\beta(k)<\infty$ a.s.

For the subsequent Lemmas A.4 and A.5, the readers may be referred to Theorem 6.4 and its next paragraph in Ch. 6 of [43].

Lemma A.4 (Conditional Lyapunov Inequality): Denote the probability space by $(\Omega,\mathcal{F},P)$. Let $\mathcal{F}_1$ be a sub $\sigma-$algebra of $\mathcal{F}$ and ξ be a random variable on $(\Omega,\mathcal{F},P)$. Then $(\mathbb{E}[|\xi|^s|\mathcal{F}_1])^{\frac{1}{s}}\le(\mathbb{E}[|\xi|^t|\mathcal{F}_1])^{\frac{1}{t}}$ a.s., $0<s<t$.

Lemma A.5 (Conditional Hölder Inequality): Denote the probability space $(\Omega,\mathcal{F},P)$. Let $\mathcal{F}_1$ be a sub $\sigma-$algebra of $\mathcal{F}$. Let ξ and η be two random variables on $(\Omega,\mathcal{F},P)$. Let constants $p\in(1,\infty)$, $q\in(1,\infty)$ and $1/p+1/q=1$. If $\mathbb{E}[|\xi|^p]<\infty$ and$\mathbb{E}[|\eta|^q]<\infty$, then $\mathbb{E}[|\xi\eta||\mathcal{F}_1]\le(\mathbb{E}[|\xi|^p|\mathcal{F}_1])^{\frac{1}{p}}(\mathbb{E}[|\eta|^q|\mathcal{F}_1])^{\frac{1}{q}}$ a.s.

Lemma A.6: For any random matrix $A\in\mathbb{R}^{m\times n}$, $\|\mathbb{E}[AA^T]\|\le n\|\mathbb{E}[A^TA]\|$.

Proof: By the properties of matrix trace, it follows that $\|\mathbb{E}[AA^T]\|=\lambda_{\max}(\mathbb{E}[AA^T])\le\mathrm{Tr}(\mathbb{E}[AA^T])=\mathrm{Tr}(\mathbb{E}[A^TA])\le n\lambda_{\max}(\mathbb{E}[A^TA])=n\|\mathbb{E}[A^TA]\|$. □

Lemma A.7: Let $\mathcal{A}=[a_{ij}]_{N\times N}$ be a weighted adjacency matrix of an undirected graph with N nodes and $\mathcal{L}$ be the associated Laplacian matrix. Let $x=[x_1^T,...,x_N^T]^T\in\mathbb{R}^{Nn}$ be any given nonzero Nn-dimensional vector where $x_i\in\mathbb{R}^n$, $i=1,2,...,N$ and there exists $i\ne j$, such that $x_i\ne x_j$. If $a_{ij}\ge 0$, $i,j=1,2,...,N$ and the graph is connected, then $x^T(\mathcal{L}\otimes I_n)x>0$.

Proof: By the definition of Laplacian matrix, we have $x^T(\mathcal{L}\otimes I_n)x=\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}a_{ij}\|x_i-x_j\|^2$. Noting that there exists $i\ne j$, such that $x_i\ne x_j$ and the graph is connected, by $a_{ij}\ge 0$, $i,j=1,2,...,N$, we get $x^T(\mathcal{L}\otimes I_n)x>0$. □

Appendix B
Proofs in Section IV

Let

$$P(k)=I_{Nn}-D(k), \tag{25}$$

where

$$D(k)=b(k)\mathcal{L}_{\mathcal{G}(k)}\otimes I_n+a(k)\mathcal{H}^T(k)\mathcal{H}(k). \tag{26}$$

The proof of Theorem IV.1 needs the following lemma.

Lemma B.1: For the algorithm (7), if Condition **C1.a**, the conditions (b.1) and (b.2) in Theorem IV.1 hold, then

$$\lim_{k\to\infty}\|\mathbb{E}[\Phi_P(k,0)\Phi_P^T(k,0)]\|=0. \tag{27}$$

Proof: By (25), we have

$$\begin{aligned}&\Phi_P^T((m+1)h-1,mh)\Phi_P((m+1)h-1,mh)\\&=(I_{Nn}-D^T(mh))\cdots(I_{Nn}-D^T((m+1)h-1))\\&\times(I_{Nn}-D((m+1)h-1))\cdots(I_{Nn}-D(mh)).\end{aligned}\tag{28}$$

Taking conditional expectation w.r.t. $\mathcal{F}(mh-1)$ on both sides of the above, by the binomial expansion, we have

$$\begin{aligned}&\|\mathbb{E}[\Phi_P^T((m+1)h-1,mh)\\&\times\Phi_P((m+1)h-1,mh)|\mathcal{F}(mh-1)]\|\\&=\|\mathbb{E}[(I_{Nn}-D^T(mh))\cdots(I_{Nn}-D^T((m+1)h-1))\\&\times(I_{Nn}-D((m+1)h-1))\cdots\\&\times(I_{Nn}-D(mh))|\mathcal{F}(mh-1)]\|\\&=\Bigg\|I_{Nn}-\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[D^T(k)+D(k)|\mathcal{F}(mh-1)]\\&+\mathbb{E}[M_2(m)+\cdots+M_{2h}(m)|\mathcal{F}(mh-1)]\Bigg\|\\&\leq\Bigg\|I_{Nn}-\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[D^T(k)+D(k)|\mathcal{F}(mh-1)]\Bigg\|\\&+\|\mathbb{E}[M_2(m)+\cdots+M_{2h}(m)|\mathcal{F}(mh-1)]\|.\end{aligned}\tag{29}$$

Here, $M_i(m), i=2,\cdots,2h$ represent the i-th order terms in the binomial expansion of $\Phi_P((m+1)h-1,mh)\Phi_P^T((m+1)h-1,mh)$.

Since the 2-norm of a symmetric matrix is equal to its spectral radius, by the definition of spectral radius, we have

$$\begin{aligned}&\Bigg\|I_{Nn}-\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[D(k)+D^T(k)|\mathcal{F}(mh-1)]\Bigg\|\\&=\rho\left(I_{Nn}-\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[D(k)+D^T(k)|\mathcal{F}(mh-1)]\right)\\&=\max_{1\leq i\leq Nn}\left|\lambda_i\left(I_{Nn}-\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[D(k)\right.\right.\\&\left.\left.+D^T(k)|\mathcal{F}(mh-1)]\right)\right|\\&=\max_{1\leq i\leq Nn}\left|1-\lambda_i\left(\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[D(k)\right.\right.\\&\left.\left.+D^T(k)|\mathcal{F}(mh-1)]\right)\right|.\end{aligned}\tag{30}$$

Since both $a(k)$ and $b(k)$ tend to zero, by the condition (b.2), we know that there exists a positive integer m_1, which is independent of the sample paths, such that

$$\begin{aligned}&\lambda_i\left(\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[D(k)+D^T(k)|\mathcal{F}(mh-1)]\right)\leq 1,\\&i=1,\cdots,Nn,\forall m\geq m_1\text{ a.s.}\end{aligned}$$

This together with (29) and (30) leads to

$$\begin{aligned}&\|\mathbb{E}[\Phi_P^T((m+1)h-1,mh)\\&\times\Phi_P((m+1)h-1,mh)|\mathcal{F}(mh-1)]\|\\&\leq 1-\lambda_{\min}\left(\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[D(k)+D^T(k)|\mathcal{F}(mh-1)]\right)\\&+\|\mathbb{E}[M_2(m)+\cdots+M_{2h}(m)|\mathcal{F}(mh-1)]\|,\\&\forall m\geq m_1\text{ a.s.}\end{aligned}\tag{31}$$

We next bound the two terms on the right side of the above. For the first term, by the definitions of $D(k)$ and $\overline{\Lambda}_m^h$ and the condition (b.1), we have

$$\begin{aligned}&1-\lambda_{\min}\left(\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[D(k)+D^T(k)|\mathcal{F}(mh-1)]\right)\\&=1-\lambda_{\min}\left(\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[2b(k)\widehat{\mathcal{L}}_{\mathcal{G}(k)}\otimes I_n\right.\\&\left.+2a(k)\mathcal{H}^T(k)\mathcal{H}(k)|\mathcal{F}(mh-1)]\right)\\&=1-2\overline{\Lambda}_m^h\leq 1-c(m),\forall m\geq m_1\text{ a.s.}\end{aligned}\tag{32}$$

By Lemma A.4 and the condition (b.2), it follows that

$$\begin{aligned}&\sup_{k\geq 0}\mathbb{E}[\|\widetilde{D}(k)\|^i|\mathcal{F}(k-1)]\\&\leq\sup_{k\geq 0}[\mathbb{E}[\|\widetilde{D}(k)\|^{2^h}|\mathcal{F}(k-1)]]^{\frac{i}{2^h}}\leq\rho_0^i\text{ a.s.},2\leq i\leq 2^h,\end{aligned}$$

where $\widetilde{D}(k)=\mathcal{L}_{\mathcal{G}(k)}\otimes I_n+\mathcal{H}^T(k)\mathcal{H}(k)$. Note that for any given random variable ξ and σ-algebra $\mathcal{F}_1\subseteq\mathcal{F}_2$, it is true that

$$\mathbb{E}[\xi|\mathcal{F}_1]=\mathbb{E}[\mathbb{E}[\xi|\mathcal{F}_2]|\mathcal{F}_1].\tag{33}$$

We then have

$$\begin{aligned}&\mathbb{E}[\|\widetilde{D}(k)\|^l|\mathcal{F}(mh-1)]\\&=\mathbb{E}[\mathbb{E}[\|\widetilde{D}(k)\|^l|\mathcal{F}(k-1)]|\mathcal{F}(mh-1)],\\&2\leq l\leq 2^h,k\geq mh.\end{aligned}$$

From the definitions of $M_i(m), i=2,\cdots,2h$ and the above, by termwise multiplication and using Lemma A.5 repeatedly, for the second term on the right side of (31), we have

$$\begin{aligned}&\|\mathbb{E}[M_2(m)+\cdots+M_{2h}(m)|\mathcal{F}(mh-1)]\|\\&\leq b^2(mh)\left(\sum_{i=2}^{2h}\mathbb{C}_{2h}^i(\max\{1,\phi\}\rho_0)^i\right)\\&=b^2(mh)\alpha,\end{aligned}\tag{34}$$

where ϕ satisfies $a(k)\leq\phi b(k)$, $\alpha=(1+\max\{1,\phi\}\rho_0)^{2h}-1-2h\max\{1,\phi\}\rho_0$ and $\mathbb{C}_m^p$ denotes the combinatorial number of choosing p elements from m elements. By (31)-(34), we have

$$\begin{aligned}&\|\mathbb{E}[\Phi_P^T((m+1)h-1,mh)\\&\times\Phi_P((m+1)h-1,mh)|\mathcal{F}(mh-1)]\|\\&\leq 1-c(m)+b^2(mh)\alpha,m\geq m_1\text{ a.s.}\end{aligned}\tag{35}$$

Denote $m_k = \lfloor \frac{k}{h} \rfloor$. By the properties of the conditional expectation, Lemma A.6 and (35), we have

$$\begin{aligned}
&\|\mathbb{E}[\Phi_P(k,0)\Phi_P^T(k,0)]\| \\
&\leq Nn\|\mathbb{E}[\Phi_P^T(k,0)\Phi_P(k,0)]\| \\
&= Nn\|\mathbb{E}[\Phi_P^T(m_kh-1,0)\Phi_P^T(k,m_kh) \\
&\times\Phi_P(k,m_kh)\Phi_P(m_kh-1,0)]\| \\
&\leq Nn\|\mathbb{E}[\Phi_P^T(m_kh-1,0)\|\Phi_P^T(k,m_kh)\Phi_P(k,m_kh)\| \\
&\times\Phi_P(m_kh-1,0)]\| \\
&= Nn\|\mathbb{E}[\mathbb{E}[\Phi_P^T(m_kh-1,0) \\
&\times\|\Phi_P^T(k,m_kh)\Phi_P(k,m_kh)\| \\
&\times\Phi_P(m_kh-1,0)|\mathcal{F}(m_kh-1)]]\| \\
&= Nn\|\mathbb{E}[\Phi_P^T(m_kh-1,0)\mathbb{E}[\|\Phi_P^T(k,m_kh) \\
&\times\Phi_P(k,m_kh)\||\mathcal{F}(m_kh-1)]\Phi_P(m_kh-1,0)]\|. \quad (36)
\end{aligned}$$

For any positive integers m, n satisfying $0 \leq m-n \leq h-1$, it follows from the condition (b.2) that there exists a constant $\rho_h^* > 0$ such that

$$\|\mathbb{E}[\Phi_P^T(m,n)\Phi_P(m,n)|\mathcal{F}(n-1)]\| < \rho_h^* \text{ a.s.} \quad (37)$$

By the above and (36), noting that $k - m_kh \leq h-1$, we have

$$\begin{aligned}
&\|\mathbb{E}[\Phi_P(k,0)\Phi_P^T(k,0)]\| \\
&\leq \rho_h^*Nn\|\mathbb{E}[\Phi_P^T(m_kh-1,0)\Phi_P(m_kh-1,0)]\| \\
&= \rho_h^*Nn\|\mathbb{E}[\Phi_P^T(m_1h-1,0)\Phi_P^T(m_kh-1,m_1h) \\
&\times\Phi_P(m_kh-1,m_1h)\Phi_P(m_1h-1,0)]\| \\
&= \rho_h^*Nn\|\mathbb{E}[\mathbb{E}(\Phi_P^T(m_1h-1,0)\Phi_P^T(m_kh-1,m_1h) \\
&\times\Phi_P(m_kh-1,m_1h) \\
&\times\Phi_P(m_1h-1,0)|\mathcal{F}(m_1h-1))]\| \\
&\leq \rho_h^*Nn\|\mathbb{E}[\Phi_P^T(m_1h-1,0)\|\mathbb{E}[\Phi_P^T(m_kh-1,m_1h) \\
&\times\Phi_P(m_kh-1,m_1h)|\mathcal{F}(m_1h-1)]\| \\
&\times\Phi_P(m_1h-1,0)]\|. \quad (38)
\end{aligned}$$

By (33) and (35), we have

$$\begin{aligned}
&\|\mathbb{E}[\Phi_P^T(m_kh-1,m_1h) \\
&\times\Phi_P(m_kh-1,m_1h)|\mathcal{F}(m_1h-1)]\| \\
&= \|\mathbb{E}[\Phi_P^T((m_k-1)h-1,m_1h) \\
&\times\Phi_P^T(m_kh-1,(m_k-1)h) \\
&\times\Phi_P(m_kh-1,(m_k-1)h) \\
&\times\Phi_P((m_k-1)h-1,m_1h)|\mathcal{F}(m_1h-1)]\| \\
&= \|\mathbb{E}[\mathbb{E}[\Phi_P^T((m_k-1)h-1,m_1h) \\
&\times\Phi_P^T(m_kh-1,(m_k-1)h) \\
&\times\Phi_P(m_kh-1,(m_k-1)h) \\
&\times\Phi_P((m_k-1)h-1,m_1h)|\mathcal{F}((m_k-1)h-1)] \\
&|\mathcal{F}(m_1h-1)]\| \\
&= \|\mathbb{E}[\Phi_P^T((m_k-1)h-1,m_1h) \\
&\times\mathbb{E}[\Phi_P^T(m_kh-1,(m_k-1)h) \\
&\times\Phi_P(m_kh-1,(m_k-1)h)|\mathcal{F}((m_k-1)h-1)] \\
&\times\Phi_P((m_k-1)h-1,m_1h)|\mathcal{F}(m_1h-1)]\| \\
&\leq \|\mathbb{E}[\Phi_P^T((m_k-1)h-1,m_1h) \\
&\times\|\mathbb{E}[\Phi_P^T(m_kh-1,(m_k-1)h) \\
&\times\Phi_P(m_kh-1,(m_k-1)h)|\mathcal{F}((m_k-1)h-1)]\| \\
&\times\Phi_P((m_k-1)h-1,m_1h)|\mathcal{F}(m_1h-1)]\| \\
&\leq [1-c(m_k-1)+b^2((m_k-1)h)\alpha] \\
&\times\|\mathbb{E}[\Phi_P^T((m_k-1)h-1,m_1h) \\
&\times\Phi_P((m_k-1)h-1,m_1h)|\mathcal{F}(m_1h-1)]\| \\
&\leq \prod_{s=m_1}^{m_k-1}[1-c(s)+b^2(sh)\alpha] \text{ a.s.}, \quad (39)
\end{aligned}$$

which together with (38) leads to

$$\begin{aligned}
&\|\mathbb{E}[\Phi_P(k,0)\Phi_P^T(k,0)]\| \\
&\leq \rho_h^*Nn\|\mathbb{E}[\Phi_P^T(m_1h-1,0)\Phi_P(m_1h-1,0)]\| \\
&\times\prod_{s=m_1}^{m_k-1}[1-c(s)+b^2(sh)\alpha]. \quad (40)
\end{aligned}$$

By (13), we know that there exists a positive integer m_2 such that

$$b^2(mh)\alpha \leq \frac{1}{2}c(m), \forall m \geq m_2, \quad (41)$$

Let $m_3 = \max\{m_2, m_1\}$ and $r_1 = \prod_{s=m_1}^{m_3-1}[1-c(s)+b^2(sh)\alpha]$. By (13) and (41), we have

$$\begin{aligned}
&\lim_{k\to\infty}\prod_{s=m_1}^{m_k-1}[1-c(s)+b^2(sh)\alpha] \\
&\leq \lim_{k\to\infty} r_1\prod_{s=m_3}^{m_k-1}[1-\frac{1}{2}c(s)] \\
&\leq \lim_{k\to\infty} r_1\exp\Big(-\frac{1}{2}\sum_{s=m_3}^{m_k-1}c(s)\Big) \\
&= r_1\exp\Big(-\frac{1}{2}\sum_{s=m_3}^{\infty}c(s)\Big) = 0. \quad (42)
\end{aligned}$$

Since $\|\mathbb{E}[\Phi_P^T(m_1h-1,0)\Phi_P(m_1h-1,0)]\| < \infty$ by the condition (b.2), (40) and (42), we have (27). The lemma is proved. □

Proof of Theorem IV.1: If $\lambda_{ji}(k) = 0$ a.s., $\forall\ j,i \in \mathcal{V}$, $\forall\ k \geq 0$, then by (12), we have

$$\begin{aligned}
&e(k+1) \\
&= P(k)e(k) + a(k)\mathcal{H}^T(k)v(k) \\
&= \Phi_P(k,0)e(0) \\
&+ \sum_{i=0}^{k}a(i)\Phi_P(k,i+1)\mathcal{H}^T(i)v(i), k \geq 0. \quad (43)
\end{aligned}$$

By the above, we have

$$\begin{aligned}
&\mathbb{E}[e(k+1)e^T(k+1)] \\
&= \mathbb{E}[\Phi_P(k,0)e(0)e^T(0)\Phi_P^T(k,0)] \\
&+\mathbb{E}\Big[\Phi_P(k,0)e(0)\sum_{i=0}^{k}a(i)[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)]^T\Big] \\
&+\mathbb{E}\Big[\sum_{i=0}^{k}a(i)\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)[\Phi_P(k,0)e(0)]^T\Big]
\end{aligned}$$

$$+\mathbb{E}\left[\Big(\sum_{i=0}^{k}a(i)\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\Big)\times\Big(\sum_{i=0}^{k}a(i)\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\Big)^T\right]. \quad (44)$$

By Assumptions **A1.a** and **A1.b**, we know that the second and third terms on the right side of (44) are both equal to zero. Moreover, from

$$\begin{aligned}\mathbb{E}[v(i)v^T(j)] &= \mathbb{E}[\mathbb{E}[v(i)v^T(j)|\mathcal{F}(i-1)]]\\ &= \mathbb{E}[\mathbb{E}[v(i)|\mathcal{F}(i-1)]v^T(j)]\\ &= 0, \forall i>j, \end{aligned}\quad (45)$$

we have

$$\begin{aligned}&\mathbb{E}\left[\left(\sum_{i=0}^{k}a(i)\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\right)\times\left(\sum_{i=0}^{k}a(i)\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\right)^T\right]\\ &=\mathbb{E}\left[\sum_{i=0}^{k}a^2(i)\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)v^T(i)\times\mathcal{H}(i)\Phi_P(k,i+1)\right].\end{aligned}$$

Substituting the above into (44) and taking the 2-norm leads to

$$\begin{aligned}&\|\mathbb{E}[e(k+1)e^T(k+1)]\|\\ &\le \|\mathbb{E}[\Phi_P(k,0)\Phi_P^T(k,0)]\|\|e(0)\|^2\\ &+\sum_{i=0}^{k}a^2(i)\|\mathbb{E}[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\\ &\times v^T(i)\mathcal{H}(i)\Phi_P^T(k,i+1)]\|\\ &= \|\mathbb{E}[\Phi_P(k,0)\Phi_P^T(k,0)]\|\|e(0)\|^2\\ &+\sum_{i=k-3h}^{k}a^2(i)\|\mathbb{E}[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\\ &\times v^T(i)\mathcal{H}(i)\Phi_P^T(k,i+1)]\|\\ &+\sum_{i=0}^{k-3h-1}a^2(i)\|\mathbb{E}[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\\ &\times v^T(i)\mathcal{H}(i)\Phi_P^T(k,i+1)]\|.\end{aligned}\quad (46)$$

By Lemma B.1, we know that the first term in the above converges to zero. For the second term in the above, when $k-h\le i<k$, we have by (37) that $\|\mathbb{E}[\Phi_P^T(k,i+1)\Phi_P(k,i+1)|\mathcal{F}(i)]\|\le\rho_h^*$ a.s.; when $k-2h\le i<k-h$, it follows from Lemma A.6 and (37) that

$$\begin{aligned}&\|\mathbb{E}[\Phi_P(k,i+1)\Phi_P^T(k,i+1)|\mathcal{F}(i)]\|\\ &\le Nn\|\mathbb{E}[\Phi_P^T(k,i+1)\Phi_P(k,i+1)|\mathcal{F}(i)]\|\\ &= Nn\|\mathbb{E}[\Phi_P^T(k-h,i+1)\Phi_P^T(k,k-h+1)\\ &\times\Phi_P(k,k-h+1)\Phi_P(k-h,i+1)|\mathcal{F}(i)]\|\\ &= Nn\|\mathbb{E}[\mathbb{E}[\Phi_P^T(k-h,i+1)\Phi_P^T(k,k-h+1)\\ &\times\Phi_P(k,k-h+1)\\ &\times\Phi_P(k-h,i+1)|\mathcal{F}(k-h)]|\mathcal{F}(i)]\|\\ &= Nn\|\mathbb{E}[\Phi_P^T(k-h,i+1)\mathbb{E}[\Phi_P^T(k,k-h+1)\\ &\times\Phi_P(k,k-h+1)|\mathcal{F}(k-h)]\\ &\times\Phi_P(k-h,i+1)|\mathcal{F}(i)]\|\\ &\le Nn\|\mathbb{E}[\Phi_P^T(k-h,i+1)\\ &\times\|\mathbb{E}[\Phi_P^T(k,k-h+1)\Phi_P(k,k-h+1)|\mathcal{F}(k-h)]\|\\ &\times\Phi_P(k-h,i+1)|\mathcal{F}(i)]\|\\ &\le Nn\rho_h^*\|\mathbb{E}[\Phi_P^T(k-h,i+1)\Phi_P(k-h,i+1)|\mathcal{F}(i)]\|\\ &\le Nn(\rho_h^*)^2 \text{ a.s.;}\end{aligned}$$

when $k-3h\le i<k-2h$, similar to the above, we have $\|\mathbb{E}[\Phi_P(k,i+1)\Phi_P^T(k,i+1)|\mathcal{F}(i)]\|\le Nn(\rho_h^*)^3$ a.s. Hence, by Assumptions **A1.a** and **A1.b**, we have

$$\begin{aligned}&\sup_{k\ge0}\|\mathbb{E}[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)v^T(i)\mathcal{H}(i)\Phi_P^T(k,i+1)]\|\\ &<\infty, k-3h\le i\le k \text{ a.s.}\end{aligned}$$

Then, noting that $a(k)$ decays to zero, the second term on the right side of (46) tends to zero.

We next prove that the third term on the right side of (46) tends to zero. Let $\widetilde{m}_i=\lceil\frac{i}{h}\rceil$. We have

$$\begin{aligned}&\|\mathbb{E}[\Phi_P(k,i+1)\Phi_P^T(k,i+1)|\mathcal{F}(i)]\|\\ &\le Nn\|\mathbb{E}[\Phi_P^T(k,i+1)\Phi_P(k,i+1)|\mathcal{F}(i)]\|\\ &= Nn\|\mathbb{E}[\Phi_P^T(\widetilde{m}_{i+1}h-1,i+1)\Phi_P^T(m_kh-1,\widetilde{m}_{i+1}h)\\ &\times\Phi_P^T(k,m_kh)\Phi_P(k,m_kh)\Phi_P(m_kh-1,\widetilde{m}_{i+1}h)\\ &\times\Phi_P(\widetilde{m}_{i+1}h-1,i+1)|\mathcal{F}(i)]\|\\ &= Nn\|\mathbb{E}[\mathbb{E}[\Phi_P^T(\widetilde{m}_{i+1}h-1,i+1)\Phi_P^T(m_kh-1,\widetilde{m}_{i+1}h)\\ &\times\Phi_P^T(k,m_kh)\Phi_P(k,m_kh)\Phi_P(m_kh-1,\widetilde{m}_{i+1}h)\\ &\times\Phi_P(\widetilde{m}_{i+1}h-1,i+1)|\mathcal{F}(m_kh-1)]|\mathcal{F}(i)]\|\\ &= Nn\|\mathbb{E}[\Phi_P^T(\widetilde{m}_{i+1}h-1,i+1)\Phi_P^T(m_kh-1,\widetilde{m}_{i+1}h)\\ &\times\mathbb{E}[\Phi_P^T(k,m_kh)\Phi_P(k,m_kh)|\mathcal{F}(m_kh-1)]\\ &\times\Phi_P(m_kh-1,\widetilde{m}_{i+1}h)\\ &\times\Phi_P(\widetilde{m}_{i+1}h-1,i+1)|\mathcal{F}(i)]\|\\ &\le Nn\rho_h^*\|\mathbb{E}[\Phi_P^T(\widetilde{m}_{i+1}h-1,i+1)\\ &\times\Phi_P^T(m_kh-1,\widetilde{m}_{i+1}h)\Phi_P(m_kh-1,\widetilde{m}_{i+1}h)\\ &\times\Phi_P(\widetilde{m}_{i+1}h-1,i+1)|\mathcal{F}(i)]\| \text{ a.s.,}\end{aligned}\quad (47)$$

where the first inequality follows by Lemma A.6, the second equality follows by (33) and the last inequality follows by (37). Similarly to (39) in the proof of Lemma B.1, we have

$$\begin{aligned}&\|\mathbb{E}[\Phi_P^T(m_kh-1,\widetilde{m}_{i+1}h)\\ &\times\Phi_P(m_kh-1,\widetilde{m}_{i+1}h)|\mathcal{F}(\widetilde{m}_{i+1}h-1)]\|\\ &\le\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-c(s)+b^2(sh)\alpha],\end{aligned}$$

From the above, (37) and (47), we have

$$\begin{aligned}&\|\mathbb{E}[\Phi_P(k,i+1)\Phi_P^T(k,i+1)|\mathcal{F}(i)]\|\\ &\le Nn\rho_h^*\|\mathbb{E}[\Phi_P^T(\widetilde{m}_{i+1}h-1,i+1)\\ &\times\Phi_P^T(m_kh-1,\widetilde{m}_{i+1}h)\Phi_P(m_kh-1,\widetilde{m}_{i+1}h)\\ &\times\Phi_P(\widetilde{m}_{i+1}h-1,i+1)|\mathcal{F}(i)]\|\end{aligned}$$

$$
\begin{aligned}
&= Nn\rho_h^*\|\mathbb{E}[\mathbb{E}[\Phi_P^T(\widetilde{m}_{i+1}h-1,i+1)\\
&\times\Phi_P^T(m_kh-1,\widetilde{m}_{i+1}h)\Phi_P(m_kh-1,\widetilde{m}_{i+1}h)\\
&\times\Phi_P(\widetilde{m}_{i+1}h-1,i+1)|\mathcal{F}(\widetilde{m}_{i+1}h-1)]|\mathcal{F}(i)]\|\\
&= Nn\rho_h^*\|\mathbb{E}[\Phi_P^T(\widetilde{m}_{i+1}h-1,i+1)\\
&\times\mathbb{E}[\Phi_P^T(m_kh-1,\widetilde{m}_{i+1}h)\\
&\times\Phi_P(m_kh-1,\widetilde{m}_{i+1}h)|\mathcal{F}(\widetilde{m}_{i+1}h-1)]\\
&\times\Phi_P(\widetilde{m}_{i+1}h-1,i+1)|\mathcal{F}(i)]\|\\
&\leq Nn\rho_h^*\|\mathbb{E}[\Phi_P^T(\widetilde{m}_{i+1}h-1,i+1)\\
&\times\|\mathbb{E}[\Phi_P^T(m_kh-1,\widetilde{m}_{i+1}h)\\
&\times\Phi_P(m_kh-1,\widetilde{m}_{i+1}h)|\mathcal{F}(\widetilde{m}_{i+1}h-1)]\|\\
&\times\Phi_P(\widetilde{m}_{i+1}h-1,i+1)|\mathcal{F}(i)]\|\\
&\leq Nn\rho_h^*\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-c(s)+b^2(sh)\alpha]\\
&\times\|\mathbb{E}[\Phi_P^T(\widetilde{m}_{i+1}h-1,i+1)\\
&\times\Phi_P(\widetilde{m}_{i+1}h-1,i+1)|\mathcal{F}(i)]\|\\
&\leq Nn(\rho_h^*)^2\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-c(s)+b^2(sh)\alpha],\\
&0\leq i\leq k-3h-1 \text{ a.s.,}
\end{aligned}
\tag{48}
$$

By (48), the condition (b.2), Assumptions **A1.a** and **A1.b**, it follows that

$$
\begin{aligned}
&\|\mathbb{E}[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)v^T(i)\mathcal{H}(i)\Phi_P^T(k,i+1)]\|\\
&= \|\mathbb{E}[\mathbb{E}[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\\
&\times v^T(i)\mathcal{H}(i)\Phi_P^T(k,i+1)|\mathcal{F}(i)]]\|\\
&\leq \|\mathbb{E}[\|\mathcal{H}^T(i)v(i)v^T(i)\mathcal{H}(i)\|\\
&\times\mathbb{E}[\Phi_P(k,i+1)\Phi_P^T(k,i+1)|\mathcal{F}(i)]]\|\\
&\leq \mathbb{E}[\|\mathcal{H}^T(i)v(i)v^T(i)\mathcal{H}(i)\|\\
&\times\|\mathbb{E}[\Phi_P(k,i+1)\Phi_P^T(k,i+1)|\mathcal{F}(i)]\|]\\
&\leq Nn(\rho_h^*)^2\mathbb{E}[\|\mathcal{H}^T(i)v(i)v^T(i)\mathcal{H}(i)\|]\\
&\times\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-c(s)+b^2(sh)\alpha]\\
&\leq Nn\beta_v\rho_0(\rho_h^*)^2\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-c(s)+b^2(sh)\alpha]\\
&\leq Nn\beta_v\rho_0(\rho_h^*)^2\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)],\\
&m_3h-1\leq i\leq k-3h-1.
\end{aligned}
$$

By the above, we have

$$
\begin{aligned}
&\sum_{i=0}^{k-3h-1}a^2(i)\|\mathbb{E}[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\\
&\times v^T(i)\mathcal{H}(i)\Phi_P^T(k,i+1)]\|\\
&= \sum_{i=0}^{m_3h-2}a^2(i)\|\mathbb{E}[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\\
&\times v^T(i)\mathcal{H}(i)\Phi_P^T(k,i+1)]\|\\
&+\sum_{i=m_3h-1}^{k-3h-1}a^2(i)\|\mathbb{E}[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i)\\
&\times v^T(i)\mathcal{H}(i)\Phi_P^T(k,i+1)]\|\\
&\leq \sum_{i=0}^{m_3h-2}a^2(i)\|\mathbb{E}[\Phi_P(k,i+1)\Phi_P^T(k,i+1)\\
&\times\mathbb{E}[\|\mathcal{H}(i)\|^2\|v(i)\|^2\||\mathcal{F}(i)]]\|\\
&+Nn\beta_v\rho_0(\rho_h^*)^2\sum_{i=m_3h-1}^{k-3h-1}a^2(i)\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)]\\
&\leq \beta_v\rho_0\sum_{i=0}^{m_3h-2}a^2(i)\|\mathbb{E}[\Phi_P(k,i+1)\Phi_P^T(k,i+1)]\|\\
&+Nn\beta_v\rho_0(\rho_h^*)^2\\
&\times\sum_{i=m_3h-1}^{k-3h-1}a^2(i)\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)].
\end{aligned}
\tag{49}
$$

By Lemma B.1, we know that $\lim_{k\to\infty}\|\mathbb{E}[\Phi_P(k,i+1)\Phi_P^T(k,i+1)]\|=0, 0\leq i\leq m_3h-2$. Then,

$$
\begin{aligned}
&\lim_{k\to\infty}\beta_v\rho_0\sum_{i=0}^{m_3h-2}a^2(i)\|\mathbb{E}[\Phi_P(k,i+1)\Phi_P^T(k,i+1)]\|\\
&=0.
\end{aligned}
\tag{50}
$$

By direct calculations, it follows that

$$
\begin{aligned}
&\sum_{i=m_3h-1}^{k-3h-1}a^2(i)\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)]\\
&\leq \sum_{i=0}^{k}a^2(i)\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)]\\
&= \sum_{i=0}^{m_kh-1}a^2(i)\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)]\\
&+ \sum_{i=m_kh}^{k}a^2(i)\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)]\\
&= \sum_{i=0}^{m_k-1}\left[\sum_{j=ih}^{(i+1)h-1}a^2(j)\right]\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)]\\
&+ \sum_{i=m_kh}^{k}a^2(i)\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)].
\end{aligned}
\tag{51}
$$

Since $a(k)$ decays to zero, it follows that

$$
\lim_{k\to\infty}\sum_{i=m_kh}^{k}a^2(i)\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)]=0.
\tag{52}
$$

By (13) and Condition **C1.a**, we have

$$
\frac{\sum_{j=(m_k-1)h}^{m_kh-1}a^2(j)}{c(m_k-1)}\leq\frac{ha^2((m_k-1)h)}{c(m_k-1)}
$$

and

$$
\begin{aligned}
&\lim_{k\to\infty}\frac{ha^2((m_k-1)h)}{c(m_k-1)}\\
&=\lim_{k\to\infty}\frac{ha^2((m_k-1)h)}{b^2((m_k-1)h)}\frac{b^2((m_k-1)h)}{c(m_k-1)}=0.
\end{aligned}
$$

Then, from (13) and Lemma A.2, we have

$$\lim_{k\to\infty}\sum_{i=0}^{m_k-1}\left[\sum_{j=ih}^{(i+1)h-1}a^2(j)\right]\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s+1)] = \lim_{k\to\infty}\frac{2\sum_{j=(m_k-1)h}^{m_kh-1}a^2(j)}{c(m_k-1)}=0.$$

By the above, (51) and (52), it follows that

$$\lim_{k\to\infty}\sum_{i=m_3h-1}^{k-3h-1}a^2(i)\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-\frac{1}{2}c(s)]=0. \tag{53}$$

Then, by (49), (50) and the above, we have

$$\lim_{k\to\infty}\sum_{i=0}^{k-3h-1}a^2(i)\|\mathbb{E}[\Phi_P(k,i+1)\mathcal{H}^T(i)v(i) \times v^T(i)\mathcal{H}(i)\Phi_P^T(k,i+1)]\|=0.$$

Thus, the third term on the right side of (46) tends to zero. We have $\lim_{k\to\infty}\|\mathbb{E}[e(k)e^T(k)]\|=0$. Since $\mathbb{E}\|e(k)\|^2\le Nn\|\mathbb{E}[e(k)e^T(k)]\|$, it follows that $\lim_{k\to\infty}\mathbb{E}\|e(k)\|^2=0$. The algorithm (7) converges in mean square.

We next prove that the algorithm (7) converges almost surely. By (43), it follows that

$$e((m+1)h) = \Phi_P((m+1)h-1,mh)e(mh)+\sum_{k=mh}^{(m+1)h-1}a(k) \times\Phi_P((m+1)h-1,k+1)\mathcal{H}^T(k)v(k), m\ge 0.$$

Taking the 2-norm and then conditional expectation w.r.t. $\mathcal{F}(mh-1)$ on both sides of the above, we have

$$\begin{aligned}&\mathbb{E}[\|e((m+1)h)\|^2|\mathcal{F}(mh-1)]\\&=e^T(mh)\mathbb{E}[\Phi_P^T((m+1)h-1,mh)\\&\times\Phi_P((m+1)h-1,mh)|\mathcal{F}(mh-1)]e(mh)\\&+\mathbb{E}\Bigg[\Bigg(\sum_{k=mh}^{(m+1)h-1}a(k)\Phi_P((m+1)h-1,k+1)\\&\times\mathcal{H}^T(k)v(k)\Bigg)^T\\&\times\Bigg(\sum_{k=mh}^{(m+1)h-1}a(k)\Phi_P((m+1)h-1,k+1)\\&\times\mathcal{H}^T(k)v(k)\Bigg)\Bigg|\mathcal{F}(mh-1)\Bigg]\\&+2e^T(mh)\mathbb{E}\Bigg[\Phi_P^T((m+1)h-1,mh)\\&\times\Bigg(\sum_{k=mh}^{(m+1)h-1}a(k)\Phi_P((m+1)h-1,k+1)\\&\times\mathcal{H}^T(k)v(k)\Bigg)\Bigg|\mathcal{F}(mh-1)\Bigg].\end{aligned}$$

By Lemma A.1 in [36] and Assumptions **A1.a** and **A1.b**, the above can be written as

$$\begin{aligned}&\mathbb{E}[\|e((m+1)h)\|^2|\mathcal{F}(mh-1)]\\&=e^T(mh)\mathbb{E}[\Phi_P^T((m+1)h-1,mh)\\&\times\Phi_P((m+1)h-1,mh)|\mathcal{F}(mh-1)]e(mh)\\&+\sum_{k=mh}^{(m+1)h-1}a^2(k)\mathbb{E}[\|\Phi_P((m+1)h-1,k+1)\\&\times\mathcal{H}^T(k)v(k)\|^2|\mathcal{F}(mh-1)].\end{aligned} \tag{54}$$

In the light of the condition (b.2), Assumptions **A1.a** and **A1.b**, we know that there exists a constant ρ_4 such that

$$\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[\|\Phi_P((m+1)h-1,k+1) \times\mathcal{H}^T(k)v(k)\|^2|\mathcal{F}(mh-1)]\le\rho_4 \text{ a.s.}, \forall m\ge 0,$$

which together with (35) and (54) gives

$$\begin{aligned}&\mathbb{E}[\|e((m+1)h)\|^2|\mathcal{F}(mh-1)]\\&\le\|\mathbb{E}[\Phi_P^T((m+1)h-1,mh)\\&\times\Phi_P((m+1)h-1,mh)|\mathcal{F}(mh-1)]\|\|e(mh)\|^2\\&+a^2(mh)\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[\|\Phi_P((m+1)h-1,k+1)\\&\times\mathcal{H}^T(k)v(k)\|^2|\mathcal{F}(mh-1)]\\&\le(1+b^2(mh)\alpha)\|e(mh)\|^2+a^2(mh)\rho_4 \text{ a.s.}\end{aligned}$$

By Lemma A.3 and Condition **C1.c**, we know that $\{e(mh),m\ge 0\}$ converges almost surely, which, along with $\lim_{m\to 0}\mathbb{E}\|e(mh)\|^2=0$ by Theorem IV.1, gives

$$\lim_{m\to 0}e(mh)=\mathbf{0}_{Nn\times 1} \text{ a.s.} \tag{55}$$

For arbitrarily small $\epsilon>0$, by Markov inequality, we have

$$\mathbb{P}\{a(k)\|v(k)\|\ge\epsilon\}\le\frac{a^2(k)\mathbb{E}\|v(k)\|^2}{\epsilon^2}, k\ge 0,$$

which together with Assumption **A1.b**, Conditions **C1.a** and **C1.c** gives

$$\sum_{k=0}^{\infty}\mathbb{P}\{a(k)\|v(k)\|\ge\epsilon\}\le\frac{\sum_{k=0}^{\infty}a^2(k)\mathbb{E}\|v(k)\|^2}{\epsilon^2}\le\frac{\beta_v\sum_{k=0}^{\infty}a^2(k)}{\epsilon^2}<\infty.$$

Then by the Borel-Cantelli lemma, we have $\mathbb{P}\{a(k)\|v(k)\|\ge\epsilon \text{ i.o.}\}=0$, which means

$$a(k)\|v(k)\|\to 0, k\to\infty \text{ a.s.} \tag{56}$$

By (43), we have

$$\begin{aligned}&\|e(k)\|\\&\le\|\Phi_P(k-1,m_kh)\|\|e(m_kh)\|\\&+\sum_{i=m_kh}^{k-1}a(i)\|v(i)\|\|\Phi_P(k-1,i+1)\|\|\mathcal{H}^T(i)\|.\end{aligned} \tag{57}$$

By Assumption **A2.a** and noting $0\le k-m_kh<h$, we know that $\sup_{k\ge 0}\|\Phi_P(k-1,m_kh)\|<\infty$ a.s. and $\sup_{k\ge 0}$

$\|\Phi_P(k-1,i+1)\|\|\mathcal{H}^T(i)\| < \infty$ a.s., $m_k h \le i \le k-1$. Then by (55)-(57), we have $\lim_{k\to\infty} e(k) = \mathbf{0}_{Nn\times 1}$ a.s. The proof is completed. □

Proof of Theorem IV.2: Since $\{\mathcal{G}(k), k \ge 0\} \in \Gamma_1$, $\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}|\mathcal{F}(k-1)]$ is positive semi-definite, which together with $\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}|\mathcal{F}(mh-1)] = \mathbb{E}[\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}|\mathcal{F}(k-1)]|\mathcal{F}(mh-1)]$ leads to that $\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}|\mathcal{F}(mh-1)]$ is positive semi-definite, $k \ge mh$. Let $c(m) = \min\{a((m+1)h), b((m+1)h)\}$. Then, by Condition **C1.a** and the condition (c.1), we have

$$\begin{aligned}
&\overline{\Lambda}_m^h \\
&= \lambda_{\min}\Bigg[\sum_{k=mh}^{(m+1)h-1}\Big(b(k)\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}|\mathcal{F}(mh-1)]\otimes I_n \\
&\quad + a(k)\mathbb{E}[\mathcal{H}^T(k)\mathcal{H}(k)|\mathcal{F}(mh-1)]\Big)\Bigg] \\
&\ge \lambda_{\min}\Bigg[\sum_{k=mh}^{(m+1)h-1}\Big(b((m+1)h) \\
&\quad \times\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}|\mathcal{F}(mh-1)]\otimes I_n \\
&\quad + a((m+1)h)\mathbb{E}[\mathcal{H}^T(k)\mathcal{H}(k)|\mathcal{F}(mh-1)]\Big)\Bigg] \\
&\ge c(m)\Lambda_m^h \ge c(m)\theta.
\end{aligned}$$

Note that

$$\sum_{m=0}^{\infty} a((m+1)h) \ge \frac{1}{h}\sum_{s=0}^{\infty}\sum_{i=(m+1)h}^{(m+2)h-1} a(i) = \frac{1}{h}\sum_{k=h}^{\infty} a(k).$$

This together with Conditions **C1.a** and **C1.b**, and $c(m) \ge \min\{a((m+1)h), a((m+1)h)/C_1\} = \min\{1, 1/C_1\}a((m+1)h)$ where $C_1 \triangleq \sup_{k\ge 0}\frac{a(k)}{b(k)}$, gives

$$\begin{aligned}
\sum_{m=0}^{\infty} c(m) &\ge \min\{1,1/C_1\}\sum_{m=0}^{\infty} a((m+1)h) \\
&\ge \frac{\min\{1,1/C_1\}}{h}\sum_{k=h}^{\infty} a(k) = \infty. \qquad (58)
\end{aligned}$$

By Conditions **C1.a** and **C1.b**, we get

$$\begin{aligned}
&\sup_{m\ge 0}\frac{a(mh)}{c(m)} \\
&= \sup_{m\ge 0}\frac{a(mh)}{a(mh+h)}\frac{a(mh+h)}{c(m)} \\
&\le \sup_{m\ge 0}\frac{a(mh)}{a(mh+h)}\frac{a(mh+h)}{\min\{a(mh+h), \frac{1}{C_1}a(mh+h)\}} \\
&< \infty,
\end{aligned}$$

which together with Condition **C1.b** gives

$$\lim_{m\to\infty}\frac{b^2(mh)}{c(m)} = \lim_{m\to\infty}\frac{b^2(mh)}{a(mh)}\frac{a(mh)}{c(m)} = 0. \qquad (59)$$

Then, $c(m)$ satisfies $b^2(mh) = o(c(m))$ and $\sum_{m=0}^{\infty} c(m) = \infty$. The proof is completed by Theorem IV.1. □

Proof of Corollary IV.1: By Assumption **A3** and the one-to-one correspondence among $\mathcal{A}_{\mathcal{G}(k)}$ and $\mathcal{L}_{\mathcal{G}(k)}$, we know that $\mathcal{L}_{\mathcal{G}(k)}$ is a homogeneous and uniform ergodic Markov chain (See Definition A.1) with the unique stationary distribution π. Denote the associated Laplacian matrix of $\mathcal{A}_l$ by $\mathcal{L}_l$ and $\widehat{\mathcal{L}}_l = \frac{\mathcal{L}_l + \mathcal{L}_l^T}{2}$, $l = 1, 2, ...$ By the definition of Λ_m^h, we have

$$\begin{aligned}
\Lambda_m^h &= \lambda_{\min}\Bigg[\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}\otimes I_n \\
&\qquad + \mathcal{H}^T(k)\mathcal{H}(k)|\mathcal{F}(mh-1)]\Bigg] \\
&= \lambda_{\min}\Bigg[\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}\otimes I_n \\
&\qquad + \mathcal{H}^T(k)\mathcal{H}(k)|\langle\widehat{\mathcal{L}}_{\mathcal{G}(mh-1)}, \mathcal{H}(mh-1)\rangle = S_0]\Bigg] \\
&= \lambda_{\min}\Bigg[\sum_{k=1}^{h}\sum_{l=1}^{\infty}(\widehat{\mathcal{L}}_l\otimes I_n + \mathcal{H}_l^T\mathcal{H}_l)\mathbb{P}^k(S_0, \langle\widehat{\mathcal{L}}_l, \mathcal{H}_l\rangle)\Bigg] \\
&\quad \forall\ S_0 \in \mathcal{S},\ \forall\ m \ge 0, h \ge 1. \qquad (60)
\end{aligned}$$

Noting the uniform ergodicity of $\{\widehat{\mathcal{L}}_{\mathcal{G}(k)}, k \ge 0\}$ and $\{\mathcal{H}(k), k \ge 0\}$ and the uniqueness of the stationary distribution π, since $\sup_{l\ge 1}\|\mathcal{L}_l\| < \infty$ and $\sup_{l\ge 1}\|\mathcal{H}_l\| < \infty$, we have

$$\begin{aligned}
&\Bigg\|\frac{\sum_{k=1}^{h}\sum_{l=1}^{\infty}(\widehat{\mathcal{L}}_l\otimes I_n + \mathcal{H}_l^T\mathcal{H}_l)\mathbb{P}^k(S_0, \langle\widehat{\mathcal{L}}_l, \mathcal{H}_l\rangle)}{h} \\
&\quad - \sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n + \mathcal{H}_l^T\mathcal{H}_l)\Bigg\| \\
&= \Bigg\|\frac{\sum_{k=1}^{h}\sum_{l=1}^{\infty}[(\widehat{\mathcal{L}}_l\otimes I_n + \mathcal{H}_l^T\mathcal{H}_l)\mathbb{P}^k(S_0, \langle\widehat{\mathcal{L}}_l, \mathcal{H}_l\rangle)]}{h} \\
&\quad - \frac{\sum_{k=1}^{h}\sum_{l=1}^{\infty}[\pi_l(\widehat{\mathcal{L}}_l\otimes I_n + \mathcal{H}_l^T\mathcal{H}_l)]}{h}\Bigg\| \\
&= \Bigg\|\frac{1}{h}\sum_{k=1}^{h}\sum_{l=1}^{\infty}[(\widehat{\mathcal{L}}_l\otimes I_n + \mathcal{H}_l^T\mathcal{H}_l) \\
&\quad \times(\mathbb{P}^k(S_0, \langle\widehat{\mathcal{L}}_l, \mathcal{H}_l\rangle) - \pi_l)]\Bigg\| \\
&\le \sup_{l\ge 1}\|\widehat{\mathcal{L}}_l\otimes I_n + \mathcal{H}_l^T\mathcal{H}_l\|\frac{\sum_{k=1}^{h} Rr^{-k}}{h} \to 0,\ h\to\infty,
\end{aligned}$$

where constants R and r are positive with $r > 1$. By the definition of uniform convergence, we know that $\frac{1}{h}\Big[\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}\otimes I_n + \mathcal{H}^T(k)\mathcal{H}(k)|\mathcal{F}(mh-1)]\Big]$ converges to $\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n + \mathcal{H}_l^T\mathcal{H}_l)$ uniformly w.r.t. m and the sample paths a.s. as $h\to\infty$.

By the conditions (d.1) and (d.2), it follows that $\lambda_{\min}(\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n + \mathcal{H}_l^T\mathcal{H}_l)) > 0$. To see this, for any given $x \in \mathbb{R}^{Nn}$, $x \ne \mathbf{0}_{Nn\times 1}$, let $x = [x_1^T, \cdots, x_N^T]^T$, $x_i \in \mathbb{R}^n$; (i) if $x = \mathbf{1}_N \otimes a$, $\exists\ a \in \mathbb{R}^n$ and $a \ne \mathbf{0}_{n\times 1}$, i.e. $x_1 = x_2 = .. = x_N = a$, then by the condition (d.2), we have $x^T(\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n + \mathcal{H}_l^T\mathcal{H}_l))x = a^T[\sum_{i=1}^{N}\sum_{l=1}^{\infty}(\pi_l H_{i,l}^T H_{i,l})]a > 0$; (ii) otherwise, there must be $x_i \ne x_j$, $\exists\ i \ne j$. By the condition (d.1), we know that $\sum_{l=1}^{\infty}\pi_l\widehat{\mathcal{L}}_l$ is the Laplacian matrix of a connected graph. Then by Lemma A.7, we have $x^T(\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n+\mathcal{H}_l^T\mathcal{H}_l))x\geq x^T(\sum_{l=1}^{\infty}\pi_l\widehat{\mathcal{L}}_l\otimes I_n)x>0$. Combining (i) and (ii), we get $\lambda_{\min}(\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n+\mathcal{H}_l^T\mathcal{H}_l))>0$.

Since the function $\lambda_{\min}(\cdot)$, whose arguments are matrices, is continuous, we know that for a given constant $\mu\in(0,2\lambda_{\min}(\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n+\mathcal{H}_l^T\mathcal{H}_l)))$, there exists a constant $\delta>0$ such that for any given matrix L, $|\lambda_{\min}(L)-\lambda_{\min}(\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n+\mathcal{H}_l^T\mathcal{H}_l))|\leq\frac{\mu}{2}$ provided $\|L-\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n+\mathcal{H}_l^T\mathcal{H}_l)\|\leq\delta$. Since the convergence is uniform, we know that there exists an integer $h_0>0$ such that

$$\sup_{m\geq0}\Bigg\|\frac{1}{h}\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}\otimes I_n+\mathcal{H}^T(k)\mathcal{H}(k)|\mathcal{F}(mh-1)]-\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n+\mathcal{H}_l^T\mathcal{H}_l)\Bigg\|\leq\delta,\ h\geq h_0 \text{ a.s.,}$$

which gives

$$\sup_{m\geq0}\left|\frac{1}{h}\Lambda_m^h-\lambda_{\min}\left(\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n+\mathcal{H}_l^T\mathcal{H}_l)\right)\right|\leq\frac{\mu}{2}, h\geq h_0 \text{ a.s.}$$

Thus, we arrive at

$$\begin{aligned}\inf_{m\geq0}\Lambda_m^h &\geq \left[\lambda_{\min}\left(\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n+\mathcal{H}_l^T\mathcal{H}_l)\right)-\frac{\mu}{2}\right]h\\ &\geq \left[\lambda_{\min}\left(\sum_{l=1}^{\infty}\pi_l(\widehat{\mathcal{L}}_l\otimes I_n+\mathcal{H}_l^T\mathcal{H}_l)\right)-\frac{\mu}{2}\right]h_0\\ &> 0 \text{ a.s.}\end{aligned}$$

By Theorem IV.2, the proof is completed. □

Appendix C
Proofs in Section V

Proof of Lemma V.1: We adopt the mathematical induction method to prove the lemma. By (6) and (17), noting that $F(k)=I_{Nn}, -d\leq k\leq-1$, we have

$$\begin{aligned}F(0) &= I_{Nn}-[b(0)\mathcal{D}_{\mathcal{G}(0)}\otimes I_n+a(0)\mathcal{H}^T(0)\mathcal{H}(0)\\ &\quad -b(0)\sum_{q=0}^{d}\overline{A}(0,q)]\\ &= I_{Nn}-[b(0)\mathcal{D}_{\mathcal{G}(0)}\otimes I_n+a(0)\mathcal{H}^T(0)\mathcal{H}(0)\\ &\quad -b(0)\mathcal{A}_{\mathcal{G}(0)}\otimes I_n].\end{aligned}$$

Note that, under Condition **C1.d**, the set $\{\psi\in(0,1)|b(0)\leq f_{C_1,\beta_a,\beta_H,N,d}(\psi)\}$ is a nonempty and bounded closed set by the continuity of $f_{C_1,\beta_a,\beta_H,N,d}(\psi)$. Hence, ψ_1 exists. Then, by the definition of ψ_1, we have

$$b(0)\left[N\beta_a+C_1\beta_H^2+N\beta_a\frac{(1-\psi_1)^{-(d+1)}-1}{(1-\psi_1)^{-1}-1}\right]\leq\psi_1. \quad (61)$$

By the above, Assumption **A2.b** and Condition **C1.a**, we have

$$\begin{aligned}\|G(0)\| &= \|b(0)\mathcal{D}_{\mathcal{G}(0)}\otimes I_n+a(0)\mathcal{H}^T(0)\mathcal{H}(0)\\ &\quad -b(0)\mathcal{A}_{\mathcal{G}(0)}\otimes I_n\|\\ &\leq b(0)\sup_{k\geq0}\|\mathcal{D}_{\mathcal{G}(k)}\|+a(0)\sup_{k\geq0}\|\mathcal{H}^T(k)\mathcal{H}(k)\|\\ &\quad +b(0)\sup_{k\geq0}\|\mathcal{A}_{\mathcal{G}(k)}\|\\ &\leq b(0)[2N\beta_a+C_1\beta_H^2]\\ &\leq b(0)[N\beta_a+C_1\beta_H^2+N\beta_a\frac{(1-\psi_1)^{-(d+1)}-1}{(1-\psi_1)^{-1}-1}]\\ &\leq \psi_1 \text{ a.s.}\end{aligned}$$

By the above and Lemma A.1, noting $\psi_1\in(0,1)$, it follows that $F(0)$ is invertible a.s. and $\|F^{-1}(0)\|\leq(1-\psi_1)^{-1}$ a.s.

Assume that $F(k)$ is invertible a.s. and $\|F^{-1}(k)\|<(1-\psi_1)^{-1}$ a.s. for $k=0,1,2,\cdots$. By (61), Assumption **A2.b** and Condition **C1.a**, we have

$$\begin{aligned}\|G(k+1)\| &= \|b(k+1)\mathcal{D}_{\mathcal{G}(k+1)}\otimes I_n\\ &\quad +a(k+1)\mathcal{H}^T(k+1)\mathcal{H}(k+1)-b(k+1)\\ &\quad \times\sum_{q=0}^{d}\overline{A}(k+1,q)[\Phi_F(k,k-q+1)]^{-1}\|\\ &\leq b(k+1)[N\beta_a+C_1\beta_H^2]\\ &\quad +b(k+1)N\beta_a\sum_{q=0}^{d}(1-\psi_1)^{-q}\\ &\leq b(0)[N\beta_a+C_1\beta_H^2+N\beta_a\\ &\quad \times[(1-\psi_1)^{-(d+1)}-1]/[(1-\psi_1)^{-1}-1]]\\ &\leq \psi_1 \text{ a.s.}\end{aligned}$$

Then By Lemma A.1, we know that $F(k+1)$ is invertible a.s. and $\|F^{-1}(k+1)\|\leq(1-\psi_1)^{-1}$ a.s. By the mathematical induction, the proof is completed. □

Before proving Theorem V.1, we need the following lemma.

Lemma C.1: If Assumption **A2.b**, Conditions **C1.a** and **C1.d** hold, and there exist a positive integer h and a positive sequence $\{c(m),m\geq0\}$ such that $\widetilde{\Lambda}_m^h\geq c(m)$ a.s. with $c(m)$ satisfying

$$b^2(mh)=o(c(m)) \text{ and } \sum_{m=0}^{\infty}c(m)=\infty, \quad (62)$$

then

$$\lim_{k\to\infty}\left\|\mathbb{E}(\Phi_F(k,0)\Phi_F^T(k,0))\right\|=0.$$

Proof: Since Assumption **A2.b**, Conditions **C1.a** and **C1.d** hold, Lemma V.1 holds. Hence, $F(k)$ is invertible a.s., and (17) follows.

Similarly to (28)–(31) in the proof of Lemma B.1, there exists a positive integer m_1' such that

$$\begin{aligned}&\|\mathbb{E}[\Phi_F((m+1)h-1,mh)\\&\times\Phi_F^T((m+1)h-1,mh)|\mathcal{F}(mh-1)]\|\\&=1-\lambda_{\min}\left(\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[G(k)+G^T(k)|\mathcal{F}(mh-1)]\right)\\&+\left\|\mathbb{E}[\overline{M}_2(m)+\cdots+\overline{M}_{2h}(m)|\mathcal{F}(mh-1)]\right\|,\\&\forall m\geq m_1' \text{ a.s.}\end{aligned}\tag{63}$$

Here, the definitions of $\overline{M}_i(m), i=2,\cdots,2h$ are similar to (29).

By (18), (19) and $\widetilde{\Lambda}_m^h\geq c(m)$ a.s., we have

$$\begin{aligned}&1-\lambda_{\min}\left(\sum_{k=mh}^{(m+1)h-1}\mathbb{E}[G(k)+G^T(k)|\mathcal{F}(mh-1)]\right)\\&=1-\lambda_{\min}\left(\sum_{k=mh}^{(m+1)h-1}\mathbb{E}\Bigg[2b(k)\mathcal{D}_{\mathcal{G}(k)}\otimes I_n\right.\\&+2a(k)\mathcal{H}^T(k)\mathcal{H}(k)\\&-b(k)\sum_{q=0}^{d}[\overline{A}(k,q)[\Phi_F(k-1,k-q)]^{-1}\\&\left.+(\overline{A}(k,q)[\Phi_F(k-1,k-q)]^{-1})^T]\Bigg|\mathcal{F}(mh-1)\Bigg]\right)\\&=1-\lambda_{\min}\left(\sum_{k=mh}^{(m+1)h-1}\mathbb{E}\Bigg[2b(k)\widehat{\mathcal{L}}_{\mathcal{G}(k)}\otimes I_n\right.\\&+2a(k)\mathcal{H}^T(k)\mathcal{H}(k)\\&-b(k)\sum_{q=0}^{d}[\overline{A}(k,q)[[\Phi_F(k-1,k-q)]^{-1}-I_{Nn}]\\&+(\overline{A}(k,q)[[\Phi_F(k-1,k-q)]^{-1}-I_{Nn}])^T]\\&\left.\Bigg|\mathcal{F}(mh-1)\Bigg]\right)\\&=1-2\widetilde{\Lambda}_m^h\leq 1-c(m) \text{ a.s.}\end{aligned}\tag{64}$$

From (18), Assumption **A2.b**, Condition **C1.a** and Lemma V.1, we have

$$\begin{aligned}\|G(k)\|&\leq b(k)\|\mathcal{D}_{\mathcal{G}(k)}\otimes I_n\|+a(k)\|\mathcal{H}^T(k)\mathcal{H}(k)\|\\&\quad+b(k)\|\sum_{q=0}^{d}\overline{A}(k,q)[\Phi_F(k-1,k-q)]^{-1}\|\\&\leq b(k)\Bigg(N\beta_a+C_1\beta_H^2+N\beta_a\\&\quad\times\frac{1-(1-\psi_1)^{-(d+1)}}{1-(1-\psi_1)^{-1}}\Bigg) \text{ a.s.}, k\geq 0.\end{aligned}$$

By the above and the definition of $\overline{M}_i(m)$, $i=2,\cdots,2h$, we have

$$\begin{aligned}&\|\overline{M}_i(m)\|\leq b^2(mh)\mathbb{C}_{2h}^i\\&\times\left(N\beta_a+C_1\beta_H^2+N\beta_a\frac{1-(1-\psi_1)^{-(d+1)}}{1-(1-\psi_1)^{-1}}\right)^i \text{ a.s.},\end{aligned}$$

where $\mathbb{C}_m^p$ represent the combinatorial number of choosing p elements from m elements. Hence,

$$\begin{aligned}&\|\mathbb{E}[\overline{M}_2(m)+\cdots+\overline{M}_{2h}(m)|\mathcal{F}(mh-1)]\|\\&\leq b^2(mh)\sum_{i=2}^{2h}\mathbb{C}_{2h}^i\\&\times\left(N\beta_a+C_1\beta_H^2+N\beta_a\frac{1-(1-\psi_1)^{-(d+1)}}{1-(1-\psi_1)^{-1}}\right)^i\\&=b^2(mh)\gamma \text{ a.s.},\end{aligned}\tag{65}$$

where $\gamma=\left(\left(N\beta_a+C_1\beta_H^2+N\beta_a\frac{1-(1-\psi_1)^{-(d+1)}}{1-(1-\psi_1)^{-1}}\right)+1\right)^{2h}-1-2h\left(N\beta_a+C_1\beta_H^2+N\beta_a\frac{1-(1-\psi_1)^{-(d+1)}}{1-(1-\psi_1)^{-1}}\right)$.

By (63), (64) and (65), we have

$$\begin{aligned}&\|\mathbb{E}[\Phi_F((m+1)h-1,mh)\\&\times\Phi_F^T((m+1)h-1,mh)|\mathcal{F}(mh-1)]\|\\&\leq 1-c(m)+b^2(mh)\gamma \text{ a.s.}, m\geq m_1'.\end{aligned}\tag{66}$$

By (17) and Assumption **A2.b**, we know that there exists a positive constant $\overline{\eta}$ such that

$$\|F(k)\|\leq\overline{\eta} \text{ a.s.}, k\geq 0.\tag{67}$$

Denote $m_k=\lfloor\frac{k}{h}\rfloor$. By (67) and Lemma A.6, we have

$$\begin{aligned}&\|\mathbb{E}[\Phi_F(k,0)\Phi_F^T(k,0)]\|\\&\leq Nn\|\mathbb{E}[\Phi_F^T(k,0)\Phi_F(k,0)]\|\\&=Nn\|\mathbb{E}[\Phi_F^T(m_kh-1,0)\Phi_F^T(k,m_kh)\\&\times\Phi_F(k,m_kh)\Phi_F(m_kh-1,0)]\|\\&\leq Nn\|\mathbb{E}[\Phi_F^T(m_kh-1,0)\|\Phi_F(k,m_kh)\|^2\\&\times\Phi_F(m_kh-1,0)]\|\\&\leq\overline{\eta}^{2h}Nn\|\mathbb{E}[\Phi_F^T(m_kh-1,0)\Phi_F(m_kh-1,0)]\|\\&=\overline{\eta}^{2h}Nn\|\mathbb{E}[\Phi_F^T(m_1'h-1,0)\Phi_F^T(m_kh-1,m_1'h)\\&\times\Phi_F(m_kh-1,m_1'h)\Phi_F(m_1'h-1,0)]\|\\&\leq\overline{\eta}^{2h}Nn\|\mathbb{E}[\|\Phi_F(m_1'h-1,0)\|^2\Phi_F^T(m_kh-1,m_1'h)\\&\times\Phi_F(m_kh-1,m_1'h)]\|\\&\leq\overline{\eta}^{2(h+m_1'h)}Nn\|\mathbb{E}[\Phi_F^T(m_kh-1,m_1'h)\\&\times\Phi_F(m_kh-1,m_1'h)]\| \text{ a.s.}\end{aligned}\tag{68}$$

From the properties of the conditional expectation and (66), it follows that

$$\begin{aligned}&\|\mathbb{E}[\Phi_F^T(m_kh-1,m_1'h)\Phi_F(m_kh-1,m_1'h)]\|\\&=\|\mathbb{E}[\Phi_F^T((m_k-1)h-1,m_1'h)\\&\times\Phi_F^T(m_kh-1,(m_k-1)h)\\&\times\Phi_F(m_kh-1,(m_k-1)h)\\&\times\Phi_F((m_k-1)h-1,m_1'h)]\|\end{aligned}$$

$$
\begin{aligned}
&= \|\mathbb{E}[\mathbb{E}[\Phi_F^T((m_k-1)h-1, m_1'h) \\
&\times \Phi_F^T(m_kh-1,(m_k-1)h) \\
&\times \Phi_F(m_kh-1,(m_k-1)h) \\
&\times \Phi_F((m_k-1)h-1, m_1'h)|\mathcal{F}((m_k-1)h-1)]]\| \\
&\leq \|\mathbb{E}[\Phi_F^T((m_k-1)h-1, m_1'h) \\
&\times \|\mathbb{E}[\Phi_F^T(m_kh-1,(m_k-1)h) \\
&\times \Phi_F(m_kh-1,(m_k-1)h)|\mathcal{F}((m_k-1)h-1)]\| \\
&\times \Phi_F((m_k-1)h-1, m_1'h)]\| \\
&\leq [1-c(m_k-1)+b^2((m_k-1)h)\gamma] \\
&\times \|\mathbb{E}[\Phi_F^T((m_k-1)h-1, m_1'h) \\
&\times \Phi_F((m_k-1)h-1, m_1'h)]\| \\
&\leq \prod_{s=m_1'}^{m_k-1}[1-c(s)+b^2(sh)\gamma] \text{ a.s.} \qquad (69)
\end{aligned}
$$

Combining (68) and (69) implies

$$
\begin{aligned}
&\|\mathbb{E}[\Phi_F(k,0)\Phi_F^T(k,0)]\| \\
&\leq Nn\overline{\eta}^{2(h+m_1'h)}\prod_{s=m_1'}^{m_k-1}[1-c(s)+b^2(sh)\gamma] \text{ a.s.}
\end{aligned}
$$

Similarly to (40)−(42) in the proof of Lemma B.1, by Condition **C1.a**, (62) and the above, we then have $\lim_{k\to\infty}\|\mathbb{E}[\Phi_F(k,0)\Phi_F^T(k,0)]\| = 0$. The proof is completed. □

Proof of Theorem V.1: By the conditions of the theorem, it follows that Lemmas V.1 and C.1 hold.

Denote the following block matrices: $\overline{r}(k) = [r^T(k), g^T(k), \cdots, g^T(k-d+1)]^T$, $\widehat{I} = [\mathbf{0}_{Nn\times Nn}, \widetilde{I}]^T$ and $\widetilde{I} = [I_{Nn}, \mathbf{0}_{Nn\times Nn}, \cdots, \mathbf{0}_{Nn\times Nn}]$, where $\widehat{I}$ and $\widetilde{I}$ are the $Nn(d+1)$ dimensional column block matrix and Nnd dimensional row block matrix with each block being the Nn dimensional matrix, respectively. Denote

$$
T(k) = \begin{pmatrix} F(k) & \widetilde{I} \\ \mathbf{0}_{Nnd\times Nn} & C(k) \end{pmatrix},
$$

which gives

$$
\Phi_T(k,0) = \begin{pmatrix} \Phi_F(k,0) & \sum_{i=0}^k \Phi_F(k,i+1)\widetilde{I}\Phi_C(i-1,0) \\ \mathbf{0}_{Nnd\times Nn} & \Phi_C(k,0) \end{pmatrix}.
$$

Denote

$$
C(k) = \begin{pmatrix} C_1(k+1) & C_2(k+1) & \cdots & C_d(k+1) \\ I_{Nn} & \mathbf{0}_{Nn\times Nn} & & \\ & \ddots & \ddots & \\ & & I_{Nn} & \mathbf{0}_{Nnd\times Nn} \end{pmatrix}. \qquad (70)
$$

By the state augmentation approach and (15), we have

$$
\begin{aligned}
\overline{r}(k+1) &= T(k)\overline{r}(k) + a(k+1)\widehat{I}\mathcal{H}^T(k+1)v(k+1) \\
&= \Phi_T(k,0)\overline{r}(0) + \sum_{i=1}^{k+1} a(i)\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)v(i), \\
&\quad k\geq 0.
\end{aligned}
$$

Premultiplying the $Nn(d+1)$ dimensional row block matrix $\overline{I} \triangleq [I_{Nn}, \mathbf{0}_{Nn\times Nn}, \cdots, \mathbf{0}_{Nn\times Nn}]$ on both sides of the above gives

$$
r(k+1) = \overline{I}\Phi_T(k,0)\overline{r}(0) + \sum_{i=1}^{k+1} a(i)\overline{I}\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)v(i),
$$

which leads to

$$
\begin{aligned}
&\mathbb{E}[r(k+1)r^T(k+1)] \\
&= \mathbb{E}[\overline{I}\Phi_T(k,0)\overline{r}(0)\overline{r}^T(0)\Phi_T^T(k,0)\overline{I}^T] \\
&+\mathbb{E}\Big[\overline{I}\Phi_T(k,0)\overline{r}(0)\Big(\sum_{i=1}^{k+1}a(i)v^T(i)\mathcal{H}(i)\widehat{I}^T\Phi_T^T(k,i)\overline{I}^T\Big)\Big] \\
&+\mathbb{E}\Big[\Big(\sum_{i=1}^{k+1}a(i)\overline{I}\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)v(i)\Big)\overline{r}^T(0)\Phi_T^T(k,0)\overline{I}^T\Big] \\
&+\mathbb{E}\Bigg[\Big[\sum_{i=1}^{k+1}a(i)\overline{I}\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)v(i)\Big] \\
&\times\Big[\sum_{i=1}^{k+1}a(i)[\overline{I}\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)v(i)]^T\Big]\Bigg]. \qquad (71)
\end{aligned}
$$

By Assumptions **A1.a** and **A1.b**, we know that the second and third terms on the right side of the above are both equal to zero.

By (45), we have

$$
\begin{aligned}
&\mathbb{E}\Bigg[\Big[\sum_{i=1}^{k+1}a(i)\overline{I}\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)v(i)\Big] \\
&\times\Big[\sum_{i=1}^{k+1}a(i)[\overline{I}\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)v(i)]^T\Big]\Bigg] \\
&= \sum_{i=1}^{k+1}a^2(i)\mathbb{E}[\overline{I}\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)v(i) \\
&\times v^T(i)\mathcal{H}(i)\widehat{I}^T\Phi_T^T(k,i)\overline{I}^T].
\end{aligned}
$$

Substituting the above into (71) and taking the 2-norm on both sides of (71), from Assumptions **A1.a**, **A1.b** and **A2.b**, it follows that

$$
\begin{aligned}
&\|\mathbb{E}[r(k+1)r^T(k+1)]\| \\
&\leq r_0\|\mathbb{E}[\overline{I}\Phi_T(k,0)\Phi_T^T(k,0)\overline{I}^T]\| + \Bigg\|\sum_{i=1}^{k+1}a^2(i) \\
&\times\mathbb{E}[\overline{I}\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)v(i)v^T(i)\mathcal{H}(i)\widehat{I}^T\Phi_T^T(k,i)\overline{I}^T]\Bigg\| \\
&= r_0\|\mathbb{E}[\overline{I}\Phi_T(k,0)\Phi_T^T(k,0)\overline{I}^T]\| \\
&+\Bigg\|\sum_{i=1}^{k+1}a^2(i)\mathbb{E}[\overline{I}\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)\mathbb{E}(v(i)v^T(i)) \\
&\times\mathcal{H}(i)\widehat{I}^T\Phi_T^T(k,i)\overline{I}^T]\Bigg\| \\
&\leq r_0\|\mathbb{E}[\overline{I}\Phi_T(k,0)\Phi_T^T(k,0)\overline{I}^T]\| + \sup_{k\geq 0}\|\mathbb{E}[v(k)v^T(k)]\| \\
&\times\Bigg\|\sum_{i=1}^{k+1}a^2(i)\mathbb{E}[\overline{I}\Phi_T(k,i)\widehat{I}\mathcal{H}^T(i)\mathcal{H}(i)\widehat{I}^T\Phi_T^T(k,i)\overline{I}^T]\Bigg\|
\end{aligned}
$$

$$
\begin{aligned}
&\le r_0\|\mathbb{E}[\overline{I}\Phi_T(k,0)\Phi_T^T(k,0)\overline{I}^T]\| \\
&+\beta_H^2\beta_v\left\|\sum_{i=1}^{k+1}a^2(i)\mathbb{E}[\overline{I}\Phi_T(k,i)\widehat{I}\widehat{I}^T\Phi_T^T(k,i)\overline{I}^T]\right\| \\
&\le r_0\|\mathbb{E}[\overline{I}\Phi_T(k,0)\Phi_T^T(k,0)\overline{I}^T]\| \\
&+\beta_H^2\beta_v\sum_{i=1}^{k+1}a^2(i)\|\mathbb{E}[\overline{I}\Phi_T(k,i)\Phi_T^T(k,i)\overline{I}^T]\|, \qquad (72)
\end{aligned}
$$

where $r_0\triangleq\|\overline{r}(0)\overline{r}^T(0)\|$. By the definitions of $\Phi_T(k,0)$ and $\overline{I}$, we have

$$
\overline{I}\Phi_T(k,0)=\Big(\Phi_F(k,0)\sum_{i=0}^{k}\Phi_F(k,i+1)\widetilde{I}\Phi_C(i-1,0)\Big).
$$

Substituting the above into (72) gives

$$
\begin{aligned}
&\|\mathbb{E}[r(k+1)r^T(k+1)]\| \\
&\le r_0\|\mathbb{E}[\Phi_F(k,0)\Phi_F^T(k,0)]\| \\
&+\beta_H^2\beta_v\sum_{i=1}^{k+1}a^2(i)\|\mathbb{E}[\Phi_F(k,i)\Phi_F^T(k,i)]\| \\
&+r_0\Big\|\mathbb{E}\Big[\Big\{\sum_{i=0}^{k}\Phi_F(k,i+1)\widetilde{I}\Phi_C(i-1,0)\Big\} \\
&\times\Big\{\sum_{i=0}^{k}\Phi_C^T(i-1,0)\widetilde{I}^T\Phi_F^T(k,i+1)\Big\}\Big]\Big\|+\beta_H^2\beta_v \\
&\times\sum_{i=1}^{k+1}a^2(i)\Big\|\mathbb{E}\Big[\Big\{\sum_{j=i}^{k}\Phi_F(k,j+1)\widetilde{I}\Phi_C(j-1,i)\Big\} \\
&\times\Big\{\sum_{j=i}^{k}\Phi_F(k,j+1)\widetilde{I}\Phi_C(j-1,i)\Big\}^T\Big]\Big\|. \qquad (73)
\end{aligned}
$$

By Lemma C.1, we know that the first term on the right side of the above converges to zero.

Denote $\widetilde{m}_i=\lceil\frac{i}{h}\rceil$. By (67) and noting the definition of m_k defined in the proof of Lemma C.1, we have

$$
\begin{aligned}
&\sum_{i=1}^{k-3h}a^2(i)\|\mathbb{E}[\Phi_F(k,i)\Phi_F^T(k,i)]\| \\
&=\sum_{i=0}^{k-3h-1}a^2(i+1)\|\mathbb{E}[\Phi_F(k,i+1)\Phi_F^T(k,i+1)]\| \\
&=\sum_{i=0}^{k-3h-1}a^2(i+1)\|\mathbb{E}[\Phi_F(k,m_kh) \\
&\times\Phi_F(m_kh-1,\widetilde{m}_{i+1}h)\Phi_F(\widetilde{m}_{i+1}h-1,i+1) \\
&\times\Phi_F^T(\widetilde{m}_{i+1}h-1,i+1)\Phi_F^T(m_kh-1,\widetilde{m}_{i+1}h) \\
&\times\Phi_F^T(k,m_kh)]\| \\
&\le\overline{\eta}^{2h}\sum_{i=0}^{k-3h-1}a^2(i+1)\|\mathbb{E}[\Phi_F(k,m_kh) \\
&\times\Phi_F(m_kh-1,\widetilde{m}_{i+1}h)\Phi_F^T(m_kh-1,\widetilde{m}_{i+1}h) \\
&\times\Phi_F^T(k,m_kh)]\| \\
&\le\overline{\eta}^{4h}\sum_{i=0}^{k-3h-1}a^2(i+1)\|\mathbb{E}[\Phi_F(m_kh-1,\widetilde{m}_{i+1}h) \\
&\times\Phi_F^T(m_kh-1,\widetilde{m}_{i+1}h)]\|,
\end{aligned}
$$

which together with Lemma A.6 and (69) leads to

$$
\begin{aligned}
&\sum_{i=1}^{k+1}a^2(i)\|\mathbb{E}[\Phi_F(k,i)\Phi_F^T(k,i)]\| \\
&\le\overline{\eta}^{4h}\sum_{i=0}^{k-3h-1}a^2(i+1)\|\mathbb{E}[\Phi_F(m_kh-1,\widetilde{m}_{i+1}h) \\
&\times\Phi_F^T(m_kh-1,\widetilde{m}_{i+1}h)]\| \\
&+\sum_{i=k-3h}^{k}a^2(i+1)\|\mathbb{E}[\Phi_F(k,i+1)\Phi_F^T(k,i+1)]\| \\
&\le Nn\overline{\eta}^{4h}\sum_{i=0}^{k-3h-1}a^2(i+1) \\
&\times\|\mathbb{E}[\Phi_F^T(m_kh-1,\widetilde{m}_{i+1}h)\Phi_F(m_kh-1,\widetilde{m}_{i+1}h)]\| \\
&+\sum_{i=k-3h}^{k}a^2(i+1)\|\mathbb{E}[\Phi_F(k,i+1)\Phi_F^T(k,i+1)]\| \\
&\le Nn\overline{\eta}^{4h}\sum_{i=0}^{k-3h-1}a^2(i+1)\prod_{s=\widetilde{m}_{i+1}}^{m_k-1}[1-c(s)+b^2(sh)\gamma] \\
&+\sum_{i=k-3h}^{k}a^2(i+1)\|\mathbb{E}[\Phi_F(k,i+1)\Phi_F^T(k,i+1)]\|.
\end{aligned}
$$

Similarly to (51)−(53) in the proof of Theorem IV.1, we have

$$
\lim_{k\to\infty}\sum_{i=1}^{k+1}a^2(i)\|\mathbb{E}[\Phi_F(k,i)\Phi_F^T(k,i)]\|=0. \qquad (74)
$$

Hence, the second term on the right side of (73) converges to zero.

From (16) and (17), we have

$$
C_i(k)=-b(k)\sum_{q=i}^{d}\overline{A}(k,q)[\Phi_F(k-1,k-q)]^{-1}, 1\le i\le d.
$$

By Assumption **A2.b** and Condition **C1.a**, then there exist $\epsilon\in(0,\frac{1-\psi_1}{\sqrt{Nnd}})$, where ψ_1 is defined in Lemma V.1 and a positive integer $k(\epsilon)$, such that for $\forall k\ge k(\epsilon)$, $\|C_i(k)\|_\infty\le\frac{\epsilon(\epsilon-1)}{\epsilon-\epsilon^{1-d}}$ a.s., $1\le i\le d$, where $\|\cdot\|_\infty$ represents the infinite norm of a matrix. If $d>1$, denote $Y=diag\{I_{Nn},\epsilon I_{Nn},\epsilon^2 I_{Nn},\cdots,\epsilon^{d-1}I_{Nn}\}$; if $d=1$, denote $Y=I_{Nn}$, which together with (70) leads to $YC(k)Y^{-1}=$

$$
\begin{pmatrix}
C_1(k+1) & \epsilon^{-1}C_2(k+1) & \cdots & \epsilon^{1-d}C_d(k+1)\\
\epsilon I_{Nn} & \mathbf{0}_{Nn\times Nn} & & \\
 & \ddots & \ddots & \\
 & & \epsilon I_{Nn} & \mathbf{0}_{Nn\times Nn}
\end{pmatrix}.
$$

Then, it follows that

$$
\begin{aligned}
\|YC(k)Y^{-1}\|_\infty &\le \max\Big\{\sum_{i=1}^{d}\epsilon^{1-i}\|C_i(k+1)\|_\infty,\epsilon\Big\}\\
&\le \max\Big\{\frac{\epsilon(\epsilon-1)}{\epsilon-\epsilon^{1-d}}\frac{\epsilon-\epsilon^{1-d}}{\epsilon-1},\epsilon\Big\}=\epsilon \text{ a.s.}
\end{aligned}
$$

From the relation between infinite norm and 2-norm of a matrix, we have

$$
\begin{aligned}
\|YC(k)Y^{-1}\| &\le \sqrt{Nnd}\|YC(k)Y^{-1}\|_\infty\\
&\le \epsilon\sqrt{Nnd}<1-\psi_1 \text{ a.s.} \qquad (75)
\end{aligned}
$$

Noting that $F(k)$ is invertible a.s., we have

$$\begin{aligned}&\Big\|\mathbb{E}\Big[\Big\{\sum_{i=0}^{k}\Phi_F(k,i+1)\widetilde{I}\Phi_C(i-1,0)\Big\}\\&\times\Big\{\sum_{i=0}^{k}\Phi_F(k,i+1)\widetilde{I}\Phi_C(i-1,0)\Big\}^T\Big]\Big\|\\&\le\sum_{0\le i,j\le k}\|\mathbb{E}[\Phi_F(k,i+1)\widetilde{I}\Phi_C(i-1,0)\\&\times\Phi_C^T(j-1,0)\widetilde{I}^T\Phi_F^T(k,j+1)]\|\\&\le\sum_{0\le i,j\le k}\|\mathbb{E}[\Phi_F(k,0)[\Phi_F(i,0)]^{-1}\widetilde{I}\Phi_C(i-1,0)\\&\times\Phi_C^T(j-1,0)\widetilde{I}^T[\Phi_F(j,0)]^{-T}\Phi_F^T(k,0)]\|\\&\le\sum_{0\le i,j\le k}\Big\|\mathbb{E}[\Phi_F(k,0)\|[\Phi_F(i,0)]^{-1}\|\|\widetilde{I}\Phi_C(i-1,0)\\&\times\Phi_C^T(j-1,0)\widetilde{I}^T\|\|[\Phi_F(j,0)]^{-T}\|\Phi_F^T(k,0)]\Big\|.\end{aligned}\tag{76}$$

By Lemma V.1, it follows that

$$\begin{aligned}\|[\Phi_F(i,0)]^{-1}\|&\le(1-\psi_1)^{-(i+1)}\text{and}\\\|[\Phi_F(j,0)]^{-T}\|&\le(1-\psi_1)^{-(j+1)}\text{ a.s.}\end{aligned}\tag{77}$$

From (75), we obtain

$$\begin{aligned}&\|\widetilde{I}\Phi_C(i-1,0)\Phi_C^T(j-1,0)\widetilde{I}^T\|\\\le&\ \|\Phi_C(i-1,0)\|\|\Phi_C(j-1,0)\|\\=&\ \|Y^{-1}\Phi_{YCY^{-1}}(i-1,0)Y\|\\&\times\|Y^{-1}\Phi_{YCY^{-1}}(j-1,0)Y\|\\\le&\ (\epsilon\sqrt{Nnd})^{i+j-2}\text{ a.s.,}\end{aligned}\tag{78}$$

which combining (76) and (77) gives

$$\begin{aligned}&\Big\|\mathbb{E}\Big[\Big\{\sum_{i=0}^{k}\Phi_F(k,i+1)\widetilde{I}\Phi_C(i-1,0)\Big\}\\&\times\Big\{\sum_{i=0}^{k}\Phi_F(k,i+1)\widetilde{I}\Phi_C(i-1,0)\Big\}^T\Big]\Big\|\\&\le(1-\psi_1)^{-2}\|\mathbb{E}[\Phi_F(k,0)\Phi_F^T(k,0)]\|\\&\times\sum_{0\le i,j\le k}((1-\psi_1)^{-1}\epsilon\sqrt{Nnd})^{i+j}\text{ a.s.}\end{aligned}$$

Noting that $(1-\psi_1)^{-1}\epsilon\sqrt{Nnd}<1$, we have $\sum_{0\le i,j<\infty}((1-\psi_1)^{-1}\epsilon\sqrt{Nnd})^{i+j}<\infty$. Hence, by Lemma C.1, it follows that

$$\begin{aligned}&\lim_{k\to\infty}\Big\|\mathbb{E}\Big[\Big\{\sum_{i=0}^{k}\Phi_F(k,i+1)\widetilde{I}\Phi_C(i-1,0)\Big\}\\&\times\Big\{\sum_{i=0}^{k}\Phi_F(k,i+1)\widetilde{I}\Phi_C(i-1,0)\Big\}^T\Big]\Big\|=0.\end{aligned}$$

Thus, the third term on the right side of (73) converges to zero.

By (77)-(78) and similarly to (76), it follows that

$$\begin{aligned}&\sum_{i=1}^{k+1}a^2(i)\Big\|\mathbb{E}\Big[\Big\{\sum_{j=i}^{k}\Phi_F(k,j+1)\widetilde{I}\Phi_C(j-1,i)\Big\}\\&\times\Big\{\sum_{j=i}^{k}\Phi_F(k,j+1)\widetilde{I}\Phi_C(j-1,i)\Big\}^T\Big]\Big\|\\&=\sum_{i=1}^{k+1}a^2(i)\Big\|\sum_{i\le j_1,j_2\le k}\mathbb{E}[\Phi_F(k,j_1+1)\widetilde{I}\Phi_C(j_1-1,i)\\&\times\Phi_C^T(j_2-1,i)\widetilde{I}^T\Phi_F^T(k,j_2+1)]\Big\|\\&=\sum_{i=1}^{k+1}a^2(i)\Big\|\sum_{i\le j_1,j_2\le k}\mathbb{E}[\Phi_F(k,i)(\Phi_F(j_1,i))^{-1}\widetilde{I}\\&\times\Phi_C(j_1-1,i)\Phi_C^T(j_2-1,i)\widetilde{I}^T(\Phi_F^T(j_2,i))^{-1}\\&\times\Phi_F^T(k,i)]\Big\|\\&\le\sum_{i=1}^{k+1}a^2(i)\Big\|\sum_{i\le j_1,j_2\le k}\mathbb{E}[\Phi_F(k,i)\|(\Phi_F(j_1,i))^{-1}\widetilde{I}\\&\times\Phi_C(j_1-1,i)\Phi_C^T(j_2-1,i)\widetilde{I}^T(\Phi_F^T(j_2,i))^{-1}\|\\&\times\Phi_F^T(k,i)]\Big\|\\&\le\sum_{i=1}^{k+1}a^2(i)\|\mathbb{E}[\Phi_F(k,i)\Phi_F^T(k,i)]\|\\&\times\sum_{i\le j_1,j_2\le k}(1-\psi_1)^{-(j_1+j_2-2i+6)}(\epsilon\sqrt{Nnd})^{j_1+j_2-2i}\\&\le(1-\psi_1)^{-6}\sum_{i=1}^{k+1}a^2(i)\|\mathbb{E}[\Phi_F(k,i)\Phi_F^T(k,i)]\|\\&\times\sum_{i\le j_1,j_2\le k}((1-\psi_1)^{-1}\epsilon\sqrt{Nnd})^{(j_1+j_2-2i)}\\&=(1-\psi_1)^{-6}\sum_{i=1}^{k+1}a^2(i)\|\mathbb{E}[\Phi_F(k,i)\Phi_F^T(k,i)]\|\\&\times\left[\frac{1-((1-\psi_1)^{-1}\epsilon\sqrt{Nnd})^{k-i+1}}{1-(1-\psi_1)^{-1}\epsilon\sqrt{Nnd}}\right]^2\\&\le\frac{(1-\psi_1)^{-6}}{[1-(1-\psi_1)^{-1}\epsilon\sqrt{Nnd}]^2}\\&\times\sum_{i=1}^{k+1}a^2(i)\|\mathbb{E}[\Phi_F(k,i)\Phi_F^T(k,i)]\|\text{ a.s.}\end{aligned}$$

In the light of (74), the above converges to zero.

So far, we have proved that all the four terms on the right side of (73) converge to zero. Thus, we have $\lim_{k\to\infty}\|\mathbb{E}(r(k+1)r^T(k+1))\|=0$, which, along with the facts that $\mathbb{E}\|r(k)\|^2=\mathbb{E}[\mathrm{Tr}(r(k)r^T(k))]=\mathrm{Tr}[\mathbb{E}(r(k)r^T(k))]$ and $r(k)$ is equivalent to $e(k)$, gives $\lim_{k\to\infty}\mathbb{E}\|e(k)\|^2=0$. The proof is completed. □

Proof of Corollary V.1: Following the lines in the proof of Lemma V.1, it can be verified that under $b(0)\le f_{C_1,\beta_a,\beta_H,N,d}(\psi_2)$, Assumption **A2.b** and Condition **C1.a**, $F(k)$ is invertible and $\|G(k)\|\le\psi_2$ a.s., $\forall k\ge0$.

Noting that $\mathcal{F}(mh-1) \subseteq \mathcal{F}(k-1), k \geq mh$, by the properties of the conditional expectation, we have

$$\begin{aligned}&\mathbb{E}[\overline{A}(k,q)[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}]|\mathcal{F}(mh-1)]\\ &= \mathbb{E}[\mathbb{E}[\overline{A}(k,q)\\ &\times[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}]|\mathcal{F}(k-1)]|\mathcal{F}(mh-1)]\\ &= \mathbb{E}[\mathbb{E}[\overline{A}(k,q)|\mathcal{F}(k-1)]\\ &\times[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}]|\mathcal{F}(mh-1)]. \end{aligned} \tag{79}$$

Since $\{\langle \mathcal{H}(k), \mathcal{A}_{\mathcal{G}(k)}, \lambda_{ji}(k), j, i \in \mathcal{V}\rangle, k \geq 0\}$ is an independent process, by Assumption **A1.a**, we know that $\overline{A}(k,q)$ is independent of $\mathcal{F}(k-1), q = 0, ..., d$. Then, by (79), we have

$$\begin{aligned}&\mathbb{E}[\overline{A}(k,q)[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}]|\mathcal{F}(mh-1)]\\ &= \mathbb{E}[\mathbb{E}[\overline{A}(k,q)]\\ &\times[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}]|\mathcal{F}(mh-1)]\\ &= \mathbb{E}[\overline{A}(k,q)]\\ &\times\mathbb{E}[[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}]|\mathcal{F}(mh-1)],\\ &k = mh, ..., (m+1)h-1, q = 0, ..., d. \end{aligned} \tag{80}$$

Let $\overline{G}_q(k) = I_{Nn} - \Phi_F(k-1, k-q), q = 0, ..., d$. Then, $\Phi_F(k-1, k-q) = I_{Nn} - \overline{G}_q(k)$. Noting that $\|G(k)\| \leq \psi_2 < 2^{\frac{1}{d}} - 1$, by the binomial expansion, we have $\|\overline{G}_q(k)\| = \|I_{Nn} - (I_{Nn} - G(k-1)\cdots(I_{Nn} - G(k-q)\| \leq [(1+\psi_2)^q - 1] < 1$. Hence, $[\Phi_F(k-1,k-q)]^{-1} = (I_{Nn} - \overline{G}_q(k))^{-1} = \sum_{i=0}^{\infty} \overline{G}_q^i(k)$. It follows that $[\Phi_F(k-1, k-q)]^{-1} - I_{Nn} = \sum_{i=1}^{\infty} \overline{G}_q^i(k)$. Therefore,

$$\begin{aligned}&\|[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}\|\\ &\leq \left\|\sum_{i=1}^{\infty}\overline{G}_q^i(k)\right\| \leq \sum_{i=1}^{\infty}[(1+\psi_2)^q - 1]^i\\ &= \frac{(1+\psi_2)^q - 1}{2-(1+\psi_2)^q}, q = 0, ..., d \text{ a.s.} \end{aligned} \tag{81}$$

Noting that for any symmetric matrix $B \in \mathbb{R}^{n\times n}$, $B \geq \lambda_{\min}(B)I_n, B \leq \|B\|I_n$, and for any matrix $B \in \mathbb{R}^{n\times n}$, $\|B\| = \|B^T\|$, by the definition of $\overline{\Lambda}_m^h$, we have

$$\begin{aligned}&\sum_{k=mh}^{(m+1)h-1}\Bigg(b(k)\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}] \otimes I_n + a(k)\mathbb{E}[\mathcal{H}^T(k)\mathcal{H}(k)]\\ &-\frac{b(k)}{2}\Bigg[\sum_{q=0}^{d}\mathbb{E}[\overline{A}(k,q)]\\ &\times\mathbb{E}[[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}]|\mathcal{F}(mh-1)]\\ &+\sum_{q=0}^{d}\mathbb{E}[[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}]|\mathcal{F}(mh-1)]^T\\ &\times\mathbb{E}[\overline{A}^T(k,q)]\Bigg]\Bigg)\\ &\geq \overline{\Lambda}_m^h I_{Nn} - \Bigg\|\sum_{k=mh}^{(m+1)h-1} b(k)\sum_{q=0}^{d}\mathbb{E}[\overline{A}(k,q)]\\ &\times\mathbb{E}[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}|\mathcal{F}(mh-1)]\Bigg\| I_{Nn}. \end{aligned} \tag{82}$$

By the above, (80), (81) and the definition of $\widetilde{\Lambda}_m^h$, we have

$$\begin{aligned}\widetilde{\Lambda}_m^h &\geq \overline{\Lambda}_m^h - \Bigg\|\sum_{k=mh}^{(m+1)h-1} b(k)\sum_{q=0}^{d}\mathbb{E}[\overline{A}(k,q)]\\ &\quad\times\mathbb{E}[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}|\mathcal{F}(mh-1)]\Bigg\|\\ &\geq \overline{\Lambda}_m^h - \sum_{k=mh}^{(m+1)h-1} b(k)\\ &\quad\times\sum_{q=0}^{d}\|\mathbb{E}[\overline{A}(k,q)]\|\frac{(1+\psi_2)^q - 1}{2-(1+\psi_2)^q} \geq c(m) \end{aligned}$$

where the last inequality follows by the condition (21). Hence, $\widetilde{\Lambda}_m^h \geq c(m)$. By Theorem V.1 and the conditions of the corollary, the proof is completed. □

Proof of Corollary V.2: We first prove the first part of the corollary. Let $c(m) = \min\{a((m+1)h), b((m+1)h)\}$. Since $\{\mathcal{G}(k), k \geq 0\} \in \Gamma_1$, we know that $\mathbb{E}[\widehat{\mathcal{L}}_{\mathcal{G}(k)}|\mathcal{F}(mh-1]$ is positive semi-definite, $k \geq mh$. Then, by the definitions of $\overline{\Lambda}_m^h$ and Λ_m^h, we have

$$\overline{\Lambda}_m^h \geq c(m)\Lambda_m^h. \tag{83}$$

Then, noting that $c(m) \geq \min\{1, 1/C_1\}a((m+1)h)$, by the definitions of C_2 and C_3, we have

$$\begin{aligned}b(mh) &\leq C_2 a(mh) \leq C_2(C_3)^h a((m+1)h)\\ &\leq C_2(C_3)^h\max\{1, C_1\}c(m). \end{aligned} \tag{84}$$

By the definitions of $\widetilde{\Lambda}_m^h$ and Σ_m^h, (83) and (84), similar to (82), we have

$$\begin{aligned}&\widetilde{\Lambda}_m^h\\ &\geq \overline{\Lambda}_m^h - \sum_{k=mh}^{(m+1)h-1} b(k)\Bigg(\sum_{q=0}^{d}\|\mathbb{E}[\overline{A}(k,q)\\ &\times([\Phi_F(k-1,k-q)]^{-1} - I_{Nn})|\mathcal{F}(mh-1)]\|\Bigg)\\ &\geq \overline{\Lambda}_m^h - b(mh)\sum_{k=mh}^{(m+1)h-1}\Bigg(\sum_{q=0}^{d}\|\mathbb{E}[\overline{A}(k,q)\\ &\times([\Phi_F(k-1,k-q)]^{-1} - I_{Nn})|\mathcal{F}(mh-1)]\|\Bigg)\\ &\geq c(m)\Lambda_m^h - c(m)\Sigma_m^h \geq c(m)\theta \text{ a.s.} \end{aligned}$$

where $\theta > 0$ by the condition (22). By Conditions **C1.a** and **C1.b**, similarly to (58)-(59), it follows that $\sum_{m=0}^{\infty} c(m) = \infty$ and $b^2(mh) = o(c(m))$. Then, the algorithm (7) converges in mean square by Theorem V.1.

We next prove the second part of the corollary. Since $\{\langle \mathcal{H}(k), \mathcal{A}_{\mathcal{G}(k)}, \lambda_{ji}(k), j, i \in \mathcal{V}\rangle, k \geq 0\}$ is an independent process, by (80) and (81), we have

$$\begin{aligned}&\|\mathbb{E}[\overline{A}(k,q)[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}]|\mathcal{F}(mh-1)]\|\\ &= \|\mathbb{E}[\overline{A}(k,q)]\\ &\times\mathbb{E}[[[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}]|\mathcal{F}(mh-1)]\|\\ &\leq \|\mathbb{E}[\overline{A}(k,q)]\|\frac{(1+\psi_2)^q - 1}{2-(1+\psi_2)^q}, q = 0, ..., d. \end{aligned}$$

Noting the definition of Σ_m^h, we then have

$$\Sigma_m^h \leq C_2(C_3)^h \max\{1, C_1\} \times \sup_{m\geq 0} \sum_{k=mh}^{(m+1)h-1} \left(\sum_{q=0}^{d} \|\mathbb{E}[\overline{A}(k,q)]\| \frac{(1+\psi_2)^q - 1}{2-(1+\psi_2)^q} \right).$$

By the above and the condition (23), we know that $\inf_{m\geq 0}(\Lambda_m^h - \Sigma_m^h) \geq \theta$ where

$$\theta \triangleq \inf_{m\geq 0} \Lambda_m^h - C_2(C_3)^h \max\{1, C_1\} \sup_{m\geq 0} \sum_{k=mh}^{(m+1)h-1} \left(\sum_{q=0}^{d} \|\mathbb{E}[\overline{A}(k,q)]\| \frac{(1+\psi_2)^q - 1}{2-(1+\psi_2)^q} \right) > 0.$$

Then, the proof is completed. □

Proof of Corollary V.3: Following the lines of the proof of Lemma V.1, it can be verified that by $b(0) \leq f_{C_1,\beta_a,\beta_H,N,d}(\psi_3)$, Assumption **A2.b** and Condition **C1.a**, $F(k)$ is invertible a.s. and $\|G(k)\| \leq \psi_3$ a.s., $\forall k \geq 0$.

Let $c(m) = \min\{a((m+1)h), b((m+1)h)\}$. Recalling the definition of Σ_m^h in Corollary V.2, by (83) and (84), we have

$$\begin{aligned} \widetilde{\Lambda}_m^h \;\geq\; & \overline{\Lambda}_m^h - \sum_{k=mh}^{(m+1)h-1} b(k) \Bigg(\sum_{q=0}^{d} \|\mathbb{E}[\overline{A}(k,q) \\ & \times ([\Phi_F(k-1,k-q)]^{-1} - I_{Nn}) | \mathcal{F}(mh-1)]\| \Bigg) \\ \geq\; & c(m)(\Lambda_m^h - \Sigma_m^h) \geq c(m)(\theta - \Sigma_m^h), \end{aligned} \tag{85}$$

where the last inequality follows by $\inf_{m\geq 0}\Lambda_m^h \geq \theta$ a.s. We next prove that $\theta - \Sigma_m^h$ has a positive lower bound under the conditions of the corollary.

By the definition of ψ_3, similar to (81), we have

$$\|[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}\| \leq \frac{(1+\psi_3)^q - 1}{2-(1+\psi_3)^q}, q = 0, ..., d \text{ a.s.}$$

By the above, we have

$$\begin{aligned} & \frac{\Sigma_m^h}{C_2(C_3)^h \max\{1, C_1\}} \\ &= \sum_{k=mh}^{(m+1)h-1} \sum_{q=0}^{d} \|\mathbb{E}[\overline{A}(k,q) \\ &\quad \times ([\Phi_F(k-1,k-q)]^{-1} - I_{Nn}) | \mathcal{F}(mh-1)]\| \\ &\leq \sum_{k=mh}^{(m+1)h-1} \sum_{q=0}^{d} \mathbb{E}[\|\overline{A}(k,q)\| \\ &\quad \times \|[\Phi_F(k-1,k-q)]^{-1} - I_{Nn}\| | \mathcal{F}(mh-1)] \\ &\leq \sum_{k=mh}^{(m+1)h-1} \sum_{q=0}^{d} \mathbb{E}[\|\overline{A}(k,q)\| | \mathcal{F}(mh-1)] \\ &\quad \times \frac{(1+\psi_3)^q - 1}{2-(1+\psi_3)^q} \\ &\leq \frac{(1+\psi_3)^d - 1}{2-(1+\psi_3)^d} \times \sum_{k=mh}^{(m+1)h-1} \left(\sum_{q=0}^{d} \mathbb{E}[\|\overline{A}(k,q)\| | \mathcal{F}(mh-1)] \right) \\ &\leq N\beta_a dh \frac{(1+\psi_3)^d - 1}{2-(1+\psi_3)^d}. \end{aligned}$$

This together with

$$\psi_3 < \left(1 + \frac{\theta}{\theta + NC_2(C_3)^h \max\{1, C_1\}\beta_a dh}\right)^{\frac{1}{d}} - 1$$

gives

$$\begin{aligned} & \theta - \Sigma_m^h \\ & \geq \theta - NC_2(C_3)^h \max\{1, C_1\}\beta_a dh \frac{(1+\psi_3)^d - 1}{2-(1+\psi_3)^d} \\ & > 0. \end{aligned}$$

Then, by (85), we have $\widetilde{\Lambda}_m^h \geq c'(m)$, $m \geq 0$, where

$$c'(m) = c(m) \left[\theta - NC_2(C_3)^h \max\{1, C_1\}\beta_a dh \times \frac{(1+\psi_3)^d - 1}{2-(1+\psi_3)^d} \right].$$

Similarly to (58)-(59), by Conditions **C1.a** and **C1.b** it is known that $\sum_{m=0}^{\infty} c'(m) = \infty$ and $b^2(mh) = o(c'(m))$. By Theorem V.1, we get the conclusion of the corollary. □

Appendix D
The Deterministic Observation Matrices in the Simulation

$H_1' = [\widetilde{H}_1, \mathbf{0}_{5\times 9}], H_2' = [\widetilde{H}_2, \mathbf{0}_{7\times 5}], H_3' = [\mathbf{0}_{6\times 4}, \widetilde{H}_3], H_4' = [\mathbf{0}_{4\times 7}, \widetilde{H}_4]$, where

$$\widetilde{H}_1 = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & -1 \\ -1 & 0 & 0 & -1 \\ -1 & 0 & -1 & 3 \end{pmatrix},$$

$$\widetilde{H}_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\widetilde{H}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix},$$

$$\widetilde{H}_4 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 & -1 & 2 \\ 0 & 1 & -1 & 0 & 0 & 0 \end{pmatrix}.$$

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

[1] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*. Boca Raton, FL, USA: CRC Press, 2004.

[2] Y. B. Shalom, X. R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation*. New York, NY, USA: Wiley, 2001.

[3] P. O. Arambel, C. Rago, and R. K. Mehra, "Covariance intersection algorithm for distributed spacecraft state estimation," in *Proc. Amer. Control Conf.*, Arlington, VA, USA, Jun. 2001, pp. 4398–4403.

[4] N. E. Leonard and A. Olshevsky, "Cooperative learning in multiagent systems from intermittent measurements," *SIAM. J. Control Optim.*, vol. 53, no. 1, pp. 1–29, 2015.

[5] G. Rigatos, P. Siano, and N. Zervos, "A distributed state estimation approach to condition monitoring of nonlinear electric power systems," *Asian J. Control*, vol. 15, no. 3, pp. 1–12, Jul. 2012.

[6] D. M. Falcao, F. F. Wu, and L. Murphy, "Parallel and distributed state estimation," *IEEE Trans. Power Syst.*, vol. 10, no. 2, pp. 724–730, May 1995.

[7] 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.

[8] S. Das and J. M. F. Moura, "Consensus+innovations distributed Kalman filter with optimized gains," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 467–481, Jan. 2017.

[9] N. Nahi, "Optimal recursive estimation with uncertain observation," *IEEE Trans. Inf. Theory*, vol. IT-15, no. 4, pp. 457–462, Jul. 1969.

[10] V. Ugrinovskii, "Distributed robust estimation over randomly switching networks using H_∞ consensus," *Automatica*, vol. 49, no. 1, pp. 160–168, Jan. 2013.

[11] S. Kar and J. M. F. Moura, "Gossip and distributed Kalman filtering: Weak consensus under weak detectability," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1766–1786, Apr. 2011.

[12] A. K. Sahu, D. Jakovetić, and S. Kar, "$\mathcal{CIRFE}$: A distributed random fields estimator," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4980–4995, Sep. 2018.

[13] A. Simões and J. Xavier, "FADE: Fast and asymptotically efficient distributed estimator for dynamic networks," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2080–2092, Apr. 2019.

[14] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3122–3136, Jul. 2008.

[15] 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.

[16] S. Al-Sayed, A. M. Zoubir, and A. H. Sayed, "Robust distributed estimation by networked agents," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3909–3921, Aug. 2017.

[17] M. R. Gholami, M. Jansson, E. G. Ström, and A. H. Sayed, "Diffusion estimation over cooperative multi-agent networks with missing data," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 3, pp. 276–289, Sep. 2016.

[18] R. Abdolee, B. Champagne, and A. H. Sayed, "Diffusion adaptation over multi-agent networks with wireless link impairments," *IEEE Trans. Mobile Comput.*, vol. 15, no. 6, pp. 1362–1376, Jun. 2016.

[19] M. J. Piggott and V. Solo, "Diffusion LMS with correlated regressors I: Realization-wise stability," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5473–5484, Nov. 2016.

[20] M. J. Piggott and V. Solo, "Diffusion LMS with correlated regressors II: Performance," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 3934–3947, Aug. 2017.

[21] J. Y. Ishihara and S. A. Alghunaim, "Diffusion LMS filter for distributed estimation of systems with stochastic state transition and observation matrices," in *Proc. Amer. Control Conf.*, Seattle, WA, USA, May 2017, pp. 5199–5204.

[22] 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.

[23] S. Kar and J. M. F. Moura, "Consensus+innovations distributed inference over networks: Cooperation and sensing in networked systems," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 99–109, May 2013.

[24] 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.

[25] J. Zhang, X. He, and D. Zhou, "Distributed filtering over wireless sensor networks with parameter and topology uncertainties," *Int. J. Control*, vol. 93, no. 4, pp. 910–921, Apr. 2020.

[26] M. Mahmoud, *Robust Control and Filtering for Time-Delay Systems*. New York, NY, USA: Marcel Dekker, 2000.

[27] C. Peng and J. Zhang, "Delay-distribution-dependent load frequency control of power systems with probabilistic interval delays," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3309–3317, Jul. 2016.

[28] Y.-P. Tian, "Time synchronization in WSNs with random bounded communication delays," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 5445–5450, Oct. 2017.

[29] Y. Zhang, F. Li, and Y. Chen, "Leader-following-based distributed Kalman filtering in sensor networks with communication delay," *J. Franklin Inst.*, vol. 354, no. 16, pp. 7504–7520, Nov. 2017.

[30] P. Millán, L. Orihuela, C. Vivas, and F. R. Rubio, "Distributed consensus-based estimation considering network induced delays and dropouts," *Automatica*, vol. 48, no. 10, pp. 2726–2729, Oct. 2012.

[31] Y. Chen and Y. Shi, "Consensus for linear multiagent systems with time-varying delays: A frequency domain perspective," *IEEE Trans. Cybern.*, vol. 47, no. 8, pp. 2143–2150, Aug. 2017.

[32] S. Liu, T. Li, and L. Xie, "Distributed consensus for multiagent systems with communication delays and limited data rate," *SIAM J. Control Optim.*, vol. 49, no. 6, pp. 2239–2262, Jan. 2011.

[33] S. Liu, L. Xie, and H. Zhang, "Distributed consensus for multi-agent systems with delays and noises in transmission channels," *Automatica*, vol. 47, no. 5, pp. 920–934, May 2011.

[34] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*. New York, NY, USA: Wiley, 2012.

[35] F. Chung, "Laplacians and the Cheeger inequality for directed graphs," *Ann. Combinatorics*, vol. 9, no. 1, pp. 1–19, Apr. 2005.

[36] T. Li and J. Wang, "Distributed averaging with random network graphs and noises," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7063–7080, Nov. 2018.

[37] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1520–1533, Sep. 2004.

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

[39] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. London, U.K.: Springer-Verlag, 1993.

[40] K. Zhou and J. C. Doyle, *Essentials of Robust Control*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

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

[42] H. Robbins and D. Siegmund, "A convergence theorem for nonnegative almost supermartingales and some applications," in *Selected Papers*, T. L. Lai and D. Siegmund, Eds. New York, NY, USA: Springer-Verlag, 1985.

[43] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed. New York, NY, USA: Springer-Verlag, 2002.

[44] H. Chan and Ü. Özgüner, "Closed-loop control of systems over communications network with queues," *Int. J. Control*, vol. 62, no. 3, pp. 493–510, 1995.

[45] L. Zhang, Y. Shi, T. Chen, and B. Huang, "A new method for stabilization of networked control systems with random delays," *IEEE Trans. Autom. Control*, vol. 50, no. 8, pp. 1177–1181, Aug. 2005.

Jiexiang Wang received the B.E. degree in automation from East China Jiaotong University, Nanchang, China, in 2014. He is currently pursuing the Ph.D. degree in control theory and engineering with the School of Mechatronic Engineering and Automation, Shanghai University. His current research interests include multi-agent systems and distributed state estimation. He received the 28th Zhang Siying (CCDC) Outstanding Youth Paper Award in 2016.

Tao Li (Senior Member, IEEE) received the B.E. degree in automation from Nankai University, Tianjin, China, in 2004, and the Ph.D. degree in systems theory from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China, in 2009. Since January 2017, he has been with East China Normal University, Shanghai, China, where he is currently a Professor and the Director of the Department of Mathematical Intelligence Sciences, School of Mathematical Sciences. His current research interests include stochastic systems, cyber-physical multi-agent systems, distributed algorithms and game theory. He is a member of the IFAC Technical Committee on Networked Systems. He received the 28th "Zhang Siying" Outstanding Youth Paper Award in 2016, the Best Paper Award of the 7th Asian Control Conference in 2009, and honorable mentioned as one of five finalists for Young Author Prize of the 17th IFAC Congress in 2008. He also received the 2009 Singapore Millennium Foundation Research Fellowship and the 2010 Australian Endeavor Research Fellowship. He was entitled Dongfang Distinguished Professor from Shanghai Municipality in 2012 and received the Excellent Young Scholar Fund from NSFC in 2015. He currently serves as an Associate Editor for several journals, including *Systems and Control Letters* and *Nonlinear Analysis: Hybrid Systems*.

Xiwei Zhang received the B.S. degree in mathematics and applied mathematics from East China Normal University, Shanghai, China, in 2018, where he is currently pursuing the Ph.D. degree in operational research and cybernetics with the School of Mathematical Sciences. His current research interests include distributed estimation theory and stochastic time-varying systems.