Distributed Averaging With Random Network Graphs and Noises

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

***Abstract*— We consider a discrete-time distributed averaging algorithm over multi-agent networks with measurement noises and time-varying random graphs. Each agent updates its state by a weighted sum of pairwise state differences between its neighbors and itself with both additive and multiplicative measurement noises. The network structure is modeled by a sequence of time-varying random digraphs, which may be spatially and temporally dependent. By stochastic Lyapunov method and the combination of algebraic graph theory and martingale convergence theory, we obtain sufficient conditions for stochastic approximation type algorithms to achieve mean square and almost sure average consensus. We prove that all states of the agents converge to a common random variable, whose mathematical expectation is the average of initial values, in mean square and almost surely if the sequence of digraphs is conditionally balanced and uniformly conditionally jointly connected. An upper bound of the variance of the limit random variable, that is, the mean square steady-state error for stochastic average consensus is given quantitatively related to the weights, the algorithm gain and the energy level of the noises.**

***Index Terms*— Distributed averaging, multi-agent system, additive and multiplicative noise, time-varying random graph.**

I. Introduction

IN REAL networked systems, there exist various kinds of uncertain factors, such as channel noises, channel fading, random link failures and recreations. In recent years, stochastic multi-agent networks have attracted great attentions from scholars in various fields and become an active interdisciplinary research subject. For stochastic multi-agent networks, distributed averaging is one of the most fundamental problems and has wide application background, such as distributed computation [1], [2], distributed filtering [3], [4], information fusion over wireless sensor networks [5], distributed learning and optimization [6], [7], load balancing [8], etc.

Measurement or communication noises affect not only the decision-making of each individual agent, but also the overall performance of the whole system. Generally, measurement or communication noises are divided into two categories: additive and multiplicative noises. Additive noise corrupts signals in the form of superposition regardless of signals' own intensities, while, multiplicative noise has a different mechanism which can be represented by its coupling with signals. For example, the effects of coherent fading in imaging radar systems can be modeled by multiplicative noises [10]. For distributed averaging with additive measurement noises, Huang and Manton [11] proposed a discrete-time stochastic approximation type average-consensus protocol, and gave sufficient conditions for mean square consensus under fixed undirected graphs. Li and Zhang [12] studied a continuous-time distributed averaging algorithm with additive measurement noises and obtained necessary and sufficient conditions for mean square average-consensus under fixed balanced digraphs. For distributed averaging with multiplicative measurement noises, Li *et al.* [13] considered average consensus under fixed undirected graphs with nonlinear noise intensity functions, and gave necessary and sufficient conditions for mean square average consensus. Ni and Li [14] considered distributed consensus with multiplicative measurement noises where the noise intensities are absolute values of relative states.

Besides measurement and communication noises, the structure of a multi-agent network often randomly changes due to packet dropouts, link/node failures or recreations, which are particularly serious for wireless networks. The random switching of network structures has a strong impact on convergence and performance of distributed averaging algorithms. This topic also attracts extensive attentions from the community of distributed averaging. Distributed averaging and consensus with a sequence of independent identically distributed (i.i.d.) graphs were studied in[15]–[20]. Especially, Bajović *et al.* [19] proved that the product of i.i.d. symmetric stochastic matrices converges exponentially in probability. The cases with ergodic stationary and finite state homogeneous Markov chain type graph sequences were analyzed in [21] and [22], respectively, which both obtained necessary and sufficient conditions for almost sure consensus. Liu *et al.* [23] and Touri and Nedic [24] studied distributed consensus with more general random graph sequences. Liu *et al.* [23] obtained sufficient conditions for the pth order

Manuscript received February 6, 2017; revised October 8, 2017 and April 29, 2018; accepted July 21, 2018. Date of publication August 1, 2018; date of current version October 18, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61522310, and in part by the Shu Guang Project of the Shanghai Municipal Education Commission and the Shanghai Education Development Foundation under Grant 17SG26.

T. Li is with the Key Laboratory of Pure Mathematics and Mathematical Practice, School of Mathematical Sciences, East China Normal University, Shanghai 200241, China. This work was partially accomplished when he was with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China (e-mail: tli@math.ecnu.edu.cn).

J. Wang is with the School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China.

Communicated by R. La, Associate Editor for Communication Networks.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2018.2862161

0018-9448 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

moment of pairwise state differences to vanish by using the jointly-containing-spanning-tree type condition. Touri and Nedic [24] gave a more general condition for the convergence of weak periodic random matrix sequences.

Most of the above literature considered the effect of random changing of network structures or measurement noises on distributed algorithms separately. In real networks, various kinds of uncertainties may co-exist. For example, there may exist additive measurement noises and channel fading accompanied with random link changes. Many scholars have long been committed to developing distributed averaging algorithms with comprehensive uncertainties, establishing convergence conditions and quantitative relations between algorithm performances and network parameters. However, the theory of distributed averaging algorithms with all the random uncertainties mentioned above is still to be developed. Li and Zhang [25] considered distributed averaging with additive measurement or communication noises and deterministic switching graphs. They established a necessary and sufficient condition for mean square average consensus under fixed digraphs and the jointly-containing-spanning-tree type condition for mean square and almost sure average consensus under switching digraphs. Rajagopal and Wainwright [26] studied distributed averaging with additive storage noises, additive communication noises and data-constrained communication. Kar and Moura [27] gave sufficient conditions for almost sure consensus under a Markov chain type graph sequence with a fixed mean graph and additive measurement noises. Huang *et al.* [28] considered the case with spatial-temporal-independent additive measurement noises and random link gains under Markov and deterministic switching network graphs. They obtained sufficient conditions for mean square and almost sure consensus. Aysal and Barner [29] proposed a model of general consensus dynamics and gave conditions for almost sure convergence under additive disturbances and randomly switching graphs. Patterson *et al.* [30] considered distributed averaging with spatial-temporal-independent random link failures and random input noises. They gave the exponential mean square convergence rate for mean square average-consensus assuming that the underlying mean graph is always undirected and connected. Wang and Elia [31] focused on the system fragilities caused by communication constraints (additive input noises, communication delay and fading channels). They established a tight relationship among uncertainties of network channels, robust mean square stability and the appearance of Leǘy flight. They gave conditions for the difference between each pair of nodes' states vanishing in mean square, without additive input noises. Furthermore, Wang and Elia [32] studied how the model parameters affect the appearance of complex behaviour and provided an expression to verify system stability. Long *et al.* [33] considered distributed consensus with multiplicative noises and randomly switching graphs assuming that the mean graph is fixed and connected.

In this paper, we propose a discrete-time multi-agent distributed averaging algorithm with both additive and multiplicative measurement noises under time-varying random graphs. A time-varying algorithm gain is adopted to attenuate the noises. By stochastic Lyapunov method and the combination of algebraic graph theory and martingale convergence theory, we obtain sufficient conditions for the distributed approximation type algorithm to achieve mean square and almost sure average consensus. We prove that all states of the agents converge to a common random variable in mean square and almost surely if the sequence of random graphs is *conditionally balanced and uniformly conditionally jointly connected*. The mathematical expectation of the variable is the average of initial states of the agents. Moreover, we give an upper bound of the variance of the limit random variable, that is, the mean square steady-state error for stochastic average consensus, which is quantitatively related to the edge weights, the algorithm gain, the number of agents, the agents' initial states, the second-order moment and the intensity coefficients of the noises. Some preliminary results on distributed averaging with additive and multiplicative noises under fixed graphs have been presented in [34]. Compared with the relevant literature, main contributions of our paper are summarized as follows.

I. The measurement model covers both cases with additive and multiplicative noises. Different from the case with only multiplicative noises, due to the introduction of the time-varying algorithm gain to attenuate additive noises, the dynamic network associated with the algorithm becomes a time-varying stochastic system. The exponential convergence of stochastic Lyapunov energy function, which is essential to obtain the almost sure consensus conditions in [13], [14], and [33], cannot be used. Besides, different from the case with only additive measurement noises [11], [12], [25], multiplicative noises relying on the relative states between agents make states and noises coupled together in a distributed information structure. This leads to the fact that the martingale term induced by noises is coupled with states and network graphs in the system centroid equation. The estimation for the term results in more complex analysis for mean square steady-state error. To these ends, we further develop stochastic Lyapunov method. Firstly, by martingale convergence theory, we prove the boundedness of mean square consensus error. Then we obtain mean square average consensus from the result of substituting the boundedness back into the difference inequality of Lyapunov function. Furthermore, by tools of martingale convergence theory, we obtain almost sure average consensus. It is worth pointing out that though Wang and Elia [31], [32] considered both additive input noises and Bernoulli fading channels, they used fixed algorithm gain and ensured that the pairwise state differences vanish in mean square in absence of the additive input noises. In addition, different from the most existing literature, the noises in this paper are allowed to be spatially and temporally dependent.

II. In [34], the network graph is assumed to be fixed, balanced and strongly connected. In this way, the property of the Laplacian matrix of a connected graph can be directly used to get the contractive property of Lyapunov energy function. While this paper studies the case with time-varying random graphs, and the network graph is neither connected nor balanced instantaneously. Thus, the method of [34] is not applicable. In this paper, stochastic Lyapunov method is further developed for the case with a sequence of random graphs. In Huang [35], the lengths of the time intervals, over which the network is jointly connected, can randomly vary but must be bounded with probability one. The network graph condition given in [35] is essentially a deterministic type condition. However, for a sequence of random graphs, it is very difficult to verify whether its sample paths satisfy such kind of conditions with probability one. Particularly, the sample paths of Markovian switching graphs do not satisfy those conditions. In this paper, the network structure among agents is modeled by more general random graph sequences. The *generalized weighted adjacency matrices* are not required to have special statistical properties, such as independency with identical distribution, Markovian switching or stationarity, etc. By introducing the concept of conditional digraph and martingale convergence theory, we establish the *uniformly conditionally joint connectivity condition* to ensure stochastic average consensus. The joint connectivity conditions with respect to a sequence of i.i.d. graphs, Markovian and deterministic switching graphs in the existing literature are all special cases of our condition. Different from [25], which assumed that the digraphs are balanced, we only require that the conditional digraph is balanced; and different from [27] and [33], we do not require a fixed mean graph. Moreover, compared with [34], we do not require the instantaneous balance of the network graph. This leads to an additional martingale term in the system centroid equation, which needs more complex estimation by martingale convergence theory.

III. In real networks, there exist not only cooperative, but also antagonistic relations between agents [36]– [38]. Such relations can be modeled by links with positive or negative weights, respectively. Among most of the existing literature on distributed averaging, nonnegative edge weights are required. Liu *et al* [23] and Touri and Nedic [24] studied noise-free consensus algorithms under random graph sequences, and required nonnegative edge weights. Porfiri and Stilwell [15] considered noise-free distributed consensus with arbitrary weights in a sampled-data setting, however, the network graph sequence is required to be i.i.d. and the mean graph is always connected. In this paper, we show that under the uniformly conditionally joint connectivity condition, even though the random edge weights take negative values at some time instants, mean square and almost sure consensus can also be achieved.

The remaining parts of this paper are arranged as follows. Section II gives preliminaries and problem formulation. Sections III and IV give main results and the proof of the main theorem. In Section V, for two special cases of Markovian switching graph sequences with countable states and independently switching graph sequences with uncountable states, the sufficient conditions for mean square and almost average consensus are given. Section VI presents some numerical examples to demonstrate the theoretical results. Section VII gives concluding remarks and some future topics.

Notation and symbols:

$\mathbf{1}_N$: N-dimensional vector with all ones;
$\mathbf{0}_N$, N-dimensional vector with all zeros;
I_N: N-dimensional identity matrix;
$O_{m\times n}$: $m\times n$ dimensional zero matrix;
$\mathbb{R}$: set of real numbers;
$A\geq B$: matrix $A-B$ is positive semi-definite;
$A\succeq B$: matrix $A-B$ is a nonnegative matrix;
A^T: transpose of matrix A;
$diag(B_1,\ldots,B_n)$: block diagonal matrix with entries being $B_1,\ldots,B_n$;
$\|A\|$: 2-norm of matrix A;
$\|A\|_F$: Frobenius-norm of matrix A;
$\mathbb{P}\{A\}$: probability of event A;
$E[\xi]$: mathematical expectation of random variable ξ;
$Var(\xi)$: variance of ξ;
$|S|$: the cardinal number of set S;
$\lceil x\rceil$: the minimal integer greater than or equal to real number x;
$\lfloor x\rfloor$: the maximal integer smaller than or equal to x;
$b_n=O(r_n)$: $\limsup_{n\to\infty}\frac{|b_n|}{r_n}<\infty$, where $\{b_n,\ n\geq 0\}$ is a real sequence and $\{r_n,\ n\geq 0\}$ is a positive real sequence;
$b_n=o(r_n)$: $\lim_{n\to\infty}\frac{b_n}{r_n}=0$;
$b_n=\Theta(r_n)$: $\limsup_{n\to\infty}\frac{|b_n|}{r_n}<\infty$ and $\liminf_{n\to\infty}\frac{|b_n|}{r_n}>0$;
$\mathcal{F}_\eta(k)=\sigma(\eta(j),0\leq j\leq k)$, $k\geq 0$, $\mathcal{F}_\eta(-1)=\{\mathbf{\Omega},\emptyset\}$, where $\{\eta(k),\ k\geq 0\}$ is a random vector or matrix sequence.

II. Preliminaries and Problem Formulation

A. Preliminaries

Let the triple $\mathcal{G}=\{\mathcal{V},\mathcal{E}_\mathcal{G},\mathcal{A}_\mathcal{G}\}$ be a weighted digraph, where $\mathcal{V}=\{1,\ldots,N\}$ is the node set with node i representing agent i; $\mathcal{E}_\mathcal{G}$ is the edge set, and $(j,i)\in\mathcal{E}_\mathcal{G}$ if and only if agent j can send information to agent i directly. Denote the neighborhood of agent i by $\mathcal{N}_i=\{j\in\mathcal{V}|(j,i)\in\mathcal{E}_\mathcal{G}\}$. We call $\mathcal{A}_\mathcal{G}=[a_{ij}]\in\mathbb{R}^{N\times N}$ *the generalized weighted adjacency matrix* of $\mathcal{G}$, where $a_{ii}=0$, and $a_{ij}\neq 0\Leftrightarrow j\in\mathcal{N}_i$. Since $\mathcal{E}_\mathcal{G}$ is uniquely determined by $\mathcal{A}_\mathcal{G}$, the digraph can also be denoted by the pair $\mathcal{G}=\{\mathcal{V},\mathcal{A}_\mathcal{G}\}$. The in-degree and out-degree of agent i are denoted by $deg_{in}(i)=\sum_{j=1}^N a_{ij}$ and $deg_{out}(i)=\sum_{j=1}^N a_{ji}$, respectively. We call $L_\mathcal{G}=\mathcal{D}_\mathcal{G}-\mathcal{A}_\mathcal{G}$ *the generalized Laplacian matrix of* $\mathcal{G}$, where $\mathcal{D}_\mathcal{G}=diag(deg_{in}(1),\ldots,deg_{in}(N))$. By the definition, $L_\mathcal{G}\mathbf{1}_N=\mathbf{0}_N$. If $deg_{in}(i)=deg_{out}(i)$, $\forall\ i\in\mathcal{V}$, then $\mathcal{G}$ is balanced. We call $\widetilde{\mathcal{G}}=\{\mathcal{V},\mathcal{E}_{\widetilde{\mathcal{G}}},\mathcal{A}_{\widetilde{\mathcal{G}}}\}$ the reversed digraph of $\mathcal{G}$, where $(i,j)\in\mathcal{E}_{\widetilde{\mathcal{G}}}$ if and only if $(j,i)\in\mathcal{E}_\mathcal{G}$ and $\mathcal{A}_{\widetilde{\mathcal{G}}}=\mathcal{A}_\mathcal{G}^T$. Then, $\widehat{\mathcal{G}}=\{\mathcal{V},\mathcal{E}_\mathcal{G}\cup\mathcal{E}_{\widetilde{\mathcal{G}}},\frac{\mathcal{A}_\mathcal{G}+\mathcal{A}_\mathcal{G}^T}{2}\}$ is called the symmetrized graph of $\mathcal{G}$. Denote $\hat{L}_\mathcal{G}=\frac{L_\mathcal{G}+L_\mathcal{G}^T}{2}$. If $a_{ij}\geq 0$, $\forall\ i,\ j\in\mathcal{V}$, then the generalized weighted adjacency matrix $\mathcal{A}_\mathcal{G}$ and the generalized Laplacian matrix $L_\mathcal{G}$ degenerate to the weighted adjacency matrix and Laplacian matrix in usual sense, respectively. And $\hat{L}_\mathcal{G}$ is the Laplacian matrix of $\widehat{\mathcal{G}}$ if and only if $\mathcal{G}$ is balanced [39].

The union digraph of $\mathcal{G}_1=\{\mathcal{V},\mathcal{E}_{\mathcal{G}_1},\mathcal{A}_{\mathcal{G}_1}\}$ and $\mathcal{G}_2=\{\mathcal{V},\mathcal{E}_{\mathcal{G}_2},\mathcal{A}_{\mathcal{G}_2}\}$ with the common node set $\mathcal{V}$ is denoted by $\mathcal{G}_1+\mathcal{G}_2=\{\mathcal{V},\mathcal{E}_{\mathcal{G}_1}\cup\mathcal{E}_{\mathcal{G}_2},\mathcal{A}_{\mathcal{G}_1}+\mathcal{A}_{\mathcal{G}_2}\}$. By the definition of $L_\mathcal{G}$, we know that $L_{\sum_{j=1}^k\mathcal{G}_j}=\sum_{j=1}^k L_{\mathcal{G}_j}$. A sequence of edges (i_1,i_2), (i_2,i_3), $\ldots$, (i_{k-1},i_k) is called a directed path from i_1 to i_k. If for all $i,\ j\in\mathcal{V}$, there exists a directed path from i to j, then $\mathcal{G}$ is strongly connected.

B. Problem Formulation

Consider a multi-agent system of N agents whose information structure is described by a sequence of random digraphs with the identical node set $\{\mathcal{G}(k) = \{\mathcal{V}, \mathcal{A}_{\mathcal{G}(k)}\}, k \geq 0\}$. We consider the following distributed averaging algorithm:

$$x_i(k+1) = x_i(k) + c(k) \sum_{j \in \mathcal{N}_i(k)} a_{ij}(k)(y_{ji}(k) - x_i(k)), \quad k \geq 0, \quad i \in \mathcal{V}, \quad (1)$$

where $x_i(k) \in \mathbb{R}$ is the state of agent i at time instant k, and $x_i(0)$, $i = 1, 2, \ldots, N$ are the initial values. Here, $\mathcal{N}_i(k)$ denotes the neighborhood of agent i at time instant k, $c(k)$ is the time-varying algorithm gain, and $y_{ji}(k)$ denotes the measurement of agent j's state by its neighboring node i at time instant k, which is given by

$$y_{ji}(k) = x_j(k) + f_{ji}(x_j(k) - x_i(k))\xi_{ji}(k), \quad i \in \mathcal{V}, \quad j \in \mathcal{N}_i(k). \quad (2)$$

where $\{\xi_{ji}(k), k \geq 0\}$ is the measurement noise sequence on channel (j, i) and $f_{ji}(x_j(k) - x_i(k))$ is the noise intensity function. The combination of (1) and (2) is called the distributed stochastic approximation type consensus algorithm [11], [25], [27]. Let $\xi(k) = [\xi_{11}(k), \ldots, \xi_{N1}(k); \ldots; \xi_{1N}(k), \ldots, \xi_{NN}(k)]^T$, where $\xi_{ji}(k) \equiv 0$ if $j \notin \mathcal{N}_i(k)$ for all $k \geq 0$.

Remark 1: The information structure of the network is modeled by a stochastic process, i.e., a sequence of random digraphs $\{\mathcal{G}(k, \omega) = \{\{1, 2, \ldots, N\}, \mathcal{A}_{\mathcal{G}(k,\omega)}\}, k \geq 0\}$, where ω is a sample point of some sample space Ω. For a fixed ω, $\{\mathcal{G}(k, \omega) = \{\{1, 2, \ldots, N\}, \mathcal{A}_{\mathcal{G}(k,\omega)}\}, k \geq 0\}$ is a sequence of deterministic digraphs, and for a fixed $k \geq 0$, $\mathcal{G}(k, \omega) = \{\{1, 2, \ldots, N\}, \mathcal{A}_{\mathcal{G}(k,\omega)}\}$ is a random element, where $\mathcal{A}(k, \omega) \triangleq \mathcal{A}_{\mathcal{G}(k,\omega)} = [a_{ij}(k, \omega)]_{N \times N}$ is an N-dimensional random matrix with zero diagonal elements. Especially, if $a_{ij}(k, \omega) = a_{ji}(k, \omega)$, $i, j = 1, 2, \ldots, N$, and $a_{ij}(k, \omega)$, $i = 2, \ldots, N$, $j = 1, 2, \ldots, i-1$ are i.i.d $0-1$ valued random variables with the probability $P\{\omega : a_{12}(k, \omega) = 1\} = p_k$, then $\mathcal{G}(k, \omega)$ is the Erdös-Rényi random graph model $\mathcal{G}(N, p_k)$ [40]. Since there is a one to one correspondence between $\mathcal{G}(k, \omega)$ and $\mathcal{A}(k, \omega)$, the random graph sequence can also be viewed as a sequence of N-dimensional random matrix with zero diagonal elements. In stochastic process theory, the sample point ω is usually omitted. Besides the Erdös-Rényi random graph model, the readers may referred to [41] for more random graph models.

We introduce the concept of *conditional digraphs*. We call $E[\mathcal{A}_{\mathcal{G}(k)}|\mathcal{F}_{\mathcal{A}}(m)]$, $m \leq k-1$, the *conditional generalized weighted adjacency matrix* of $\mathcal{A}_{\mathcal{G}(k)}$ with respect to $\mathcal{F}_{\mathcal{A}}(m)$, and call its associated random graph the *conditional digraph* of $\mathcal{G}(k)$ with respect to $\mathcal{F}_{\mathcal{A}}(m)$, denoted by $\mathcal{G}(k|m)$, i.e., $\mathcal{G}(k|m) = \{\mathcal{V}, E[\mathcal{A}_{\mathcal{G}(k)}|\mathcal{F}_{\mathcal{A}}(m)]\}$. In this paper, we consider the sequence of balanced conditional digraphs as follows:

$$\Gamma_1 = \Big\{\{\mathcal{G}(k), k \geq 0\} | E[\mathcal{A}_{\mathcal{G}(k)}|\mathcal{F}_{\mathcal{A}}(k-1)] \succeq O_{N \times N} \text{ a.s.}, \quad \mathcal{G}(k|k-1) \text{ is balanced a.s.}, \ k \geq 0\Big\}.$$

For the measurement model (2) and the algorithm gain $c(k)$, we have the following assumptions.

(A1) For the noise intensity function $f_{ji}(\cdot)$: $\mathbb{R} \to \mathbb{R}$, there exist nonnegative constants σ_{ji} and b_{ji}, $i, j \in \mathcal{V}$, such that $|f_{ji}(x)| \leq \sigma_{ji}|x| + b_{ji}$, $\forall x \in \mathbb{R}$.

(A2) The noise process $\{\xi(k), \mathcal{F}_{\xi}(k), k \geq 0\}$ is a sequence of vector-valued martingale differences and there exists a positive constant β such that $\sup_{k \geq 0} E[\|\xi(k)\|^2 | \mathcal{F}_{\xi}(k-1)] \leq \beta$ a.s.

(A3) $c(k) > 0, \forall k \geq 0, \sum_{k=0}^{\infty} c(k) = \infty, \sum_{k=0}^{\infty} c^2(k) < \infty$.

(A4) $c(k)$ decreases monotonously, $c(k) = O(c(k+h))$, $k \to \infty$, $\forall h = 0, 1, 2, \ldots$

Remark 2: Assumption **(A1)** shows that the measurement model (2) covers both cases of additive and multiplicative measurement noises. Here, b_{ji}, $i, j \in \mathcal{V}$ and σ_{ji}, $i, j \in \mathcal{V}$ are additive and multiplicative noise intensity coefficients, respectively. The measurement models with additive noises in [11], [12], and [25] and those with multiplicative noises in [13], [14], and [33] are both special cases of model (2). In detail, the measurement model in [25] is $y_{ji}(k) = x_j(k) + \xi_{ji}(k)$, $j \in \mathcal{N}_i(k)$. The measurement model in [13] is $y_{ji}(k) = x_j(k) + f_{ji}(x_j(k) - x_i(k))\xi_{ji}(k)$, $j \in \mathcal{N}_i(k)$, where $|f_{ji}(x_j(k) - x_i(k))| \leq \sigma_{ji}|x_j(k) - x_i(k)|$. The measurement model in [14] and [33] is $y_{ji}(k) = x_j(k) + \sigma_{ji}|x_j(k) - x_i(k)|\xi_{ji}(k)$, $j \in \mathcal{N}_i(k)$. Obviously, all the noise intensity functions of the above three kinds of models all satisfy **(A1)**.

Remark 3: In Assumption **(A2)**, we assume that the overall noises constitute a martingale difference sequence without the requirement that the noises are spatial-temporal independent as in the most existing literature [13], [14], [28], [31]– [33]. This weaker assumption leads to difficulties in analyzing the algorithm, where the coupled term of states and noises cannot be simply separated as the case with independent noises. If $\{\xi(k), k \geq 0\}$ is an independent zero mean sequence with bounded second-order moments, then Assumption **(A2)** holds.

Remark 4: Existing literature showed that a fixed algorithm gain can ensure strong consensus [13], [14], [33] if only multiplicative measurement noises are considered. Here, we adopt the decaying algorithm gain $c(k)$ to attenuate the additive noises. In the field of distributed algorithms, Assumption **(A3)** ensures that $c(k)$ vanishes with a proper rate for attenuating noises and meanwhile the algorithm does not converge too early. If $c(k)$ decreases monotonically, and there are constants $\gamma \in (0.5, 1]$ and $\beta \geq -1$, $c_1 > 0$, $c_2 > 0$, such that for sufficiently large k, $\frac{c_1 \ln^{\beta}(k)}{k^{\gamma}} \leq c(k) \leq \frac{c_2 \ln^{\beta}(k)}{k^{\gamma}}$, then both Assumptions **(A3)** and **(A4)** hold.

We have the following assumption on the random graph sequence and the measurement noises.

(A5) The random graph sequence $\{\mathcal{G}(k), k \geq 0\}$ and the noise process $\{\xi(k), k \geq 0\}$ are mutually independent.

Remark 5: Here, Assumption **(A5)** requires that the graph sequence and the measurement noises are mutually independent. And different from the most existing works on distributed averaging under random network graphs, here, neither the graph sequence nor the process of measurement noises is required to be spatially or temporally independent. For the case with time-invariant random graphs, Porfiri and Stilwell [15] and Hatano and Mesbahi [18] assumed independent channels. For the case with time-varying random graphs, Boyd *et al*. [16], Kar and Moura [17], Tahbaz-Salehi and Jadbabaie [20] and Long *et al*. [33] assumed that $\{\mathcal{G}(k), k \geq 0\}$ is a sequence of independent random graphs. These spatial or temporal independency requirements cannot be always satisfied for real networks. Take a sensor network as the example. On the spatial scale, if a sensor node fails due to battery exhausted, then all channels between this node and its neighbors become inactive. This would happen randomly and the statistics of channels associated with this node are obviously spatially dependent. On the temporal scale, the unreliability of channels would increase due to aging of sensors as time goes on. Thus, the statistics of channels are also temporally dependent. In this paper, we do not require the spatial and temporal independency of the network graphs, which can cover more practical cases besides those in [15]–[18], [20], and [33]. To remove the independency between the noise process and the graph sequence would be more interesting and challenging.

Let $X(k) = [x_1(k), \cdots, x_N(k)]^T$, $D(k) = diag(\alpha_1^T(k), \cdots, \alpha_N^T(k))$ with $\alpha_i^T(k)$ being the ith row of $\mathcal{A}_{\mathcal{G}(k)}$, $Y(k) = diag(f_1(k), \cdots, f_N(k))$, where $f_i(k) = diag(f_{1i}(x_1(k) - x_i(k)), \cdots, f_{Ni}(x_N(k) - x_i(k)))$. Substituting (2) into (1) leads to the dynamic system associated with the algorithm (1) and (2) in the compact form

$$X(k+1) = (I_N - c(k)L_{\mathcal{G}(k)})X(k) + c(k)D(k)Y(k)\xi(k). \quad (3)$$

Remark 6: In [29], the dynamic system is described by $x(t+1) = A(t)x(t) + B(t)m(t)$, where $\{x(s) : s \leq t\}$ is independent of $A(t)$, $B(t)$ and $m(t)$ for all $t \geq 0$; and the disturbance process $m(t)$ is independent of $B(t)$. This assumption obviously fails for our model (3).

Definition 1 [25]: Stochastic average consensus: for the system (1) and (2), if for any given $X(0) \in \mathbb{R}^N$, there exists a random variable x^*, such that $E(x^*) = \frac{1}{N}\sum_{j=1}^N x_j(0)$, $Var(x^*) < \infty$, $\lim_{k\to\infty} E[x_i(k) - x^*]^2 = 0$, $i \in \mathcal{V}$, and $\lim_{k\to\infty} x_i(k) = x^*$ a.s., $i \in \mathcal{V}$, then we say that the system (1) and (2) achieves mean square and almost sure average consensus.

For consensus algorithms with random noises and randomly switching graphs, generally, the state limit is not a deterministic value but becomes some random variable [25]– [28]. Definition 1 is a generalization of the concept of deterministic average-consensus in [39]. Due to the stochastic uncertainties in the network, the limit value of the states is not the exact average $\frac{1}{N}\sum_{j=1}^N x_j(0)$, but becomes a random variable with its mathematical expectation being $\frac{1}{N}\sum_{j=1}^N x_j(0)$ and its variance characterizing mean square steady-state error.

In this paper, we aim at giving the conditions under which the system (1) and (2) achieves mean square and almost sure average consensus based on the models formulated above, i.e., the random digraph sequence and the measurement model with both additive and multiplicative noises. The following section gives the main result.

III. Main Results

Let $J_N = \frac{1}{N}\mathbf{1}\mathbf{1}^T$ and $P_N = I_N - J_N$. Denote the consensus error vector $\delta(k) = P_N X(k)$ and the Lyapunov energy function $V(k) = \|\delta(k)\|^2$. For any given $k \geq 0$ and positive integer h, denote

$$\lambda_k^h = \lambda_2\left(\sum_{i=k}^{k+h-1} E[\hat{L}_{\mathcal{G}(i)}|\mathcal{F}_{\mathcal{A}}(k-1)]\right), \quad (4)$$

where $\lambda_2(\cdot)$ denotes the second smallest eigenvalue. Since $E[\hat{L}_{\mathcal{G}(i)}|\mathcal{F}_{\mathcal{A}}(k-1)]$ is a real symmetric matrix a.s., λ_k^h is well defined.

We are now in the position for the main result.

Theorem 1: For the system (1)-(2) and the associated random graph sequence $\{\mathcal{G}(k), k \geq 0\} \in \Gamma_1$, assume that

(a) Assumptions **(A1)-(A5)** hold;

(b) there exist deterministic positive integer h and positive constants θ and ρ_0, such that (b.1) $\inf_{m\geq 0} \lambda_{mh}^h \geq \theta$ a.s.; (b.2) $\sup_{k\geq 0}\left[E[\|L_{\mathcal{G}(k)}\|^{2^{\max\{h,2\}}}|\mathcal{F}_{\mathcal{A}}(k-1)]\right]^{\frac{1}{2^{\max\{h,2\}}}} \leq \rho_0$ a.s.

Then, as $k \to \infty$, the consensus error $\delta(k)$ vanishes in mean square and almost surely. Moreover, all states $x_i(k)$, $i \in \mathcal{V}$, converge to a common random variable x^*, in mean square and almost surely, with $E(x^*) = \frac{1}{N}\sum_{j=1}^N x_j(0)$ and

$$Var(x^*) \leq \frac{4c\beta b^2\rho_1}{N^2} + \frac{8\widetilde{c}\beta\sigma^2\rho_1}{N^2} + \frac{2c\rho_2\ q_x}{N}, \quad (5)$$

where

$$\begin{aligned}
c &= \sum_{k=0}^{\infty} c^2(k), \widetilde{c} = \sum_{k=0}^{\infty} E[V(k)]c^2(k),\\
\sigma &= \max_{1\leq i,j\leq N}\{\sigma_{ji}\}, b = \max_{1\leq i,j\leq N}\{b_{ji}\},\\
q_x &= \exp\left\{c\rho_0^2\right\}\left(\|X(0)\|^2 + 2c\beta\rho_1(2\sigma^2\ q_v + b^2)\right),\\
q_v &= \exp\left\{c(\rho_0^2 + 4\rho_1\beta\sigma^2)\right\}\left(\|\delta(0)\|^2 + 2c\beta\rho_1 b^2\right),
\end{aligned}$$

ρ_1 and ρ_2 are constants satisfying

$$\sup_{k\geq 0} E\left[|\mathcal{E}_{\mathcal{G}(k)}| \max_{1\leq i,j\leq N} a_{ij}^2(k)|\mathcal{F}_{\mathcal{A}}(k-1)\right] \leq \rho_1\ a.s.$$

$$\max_{1\leq i\leq N}\sup_{k\geq 0} E\left[\left(\sum_{j=1}^N a_{ij}(k) - \sum_{j=1}^N a_{ji}(k)\right)^2\middle|\mathcal{F}_{\mathcal{A}}(k-1)\right] \leq \rho_2\ a.s.$$

Remark 7: Most of existing literature on consensus-based distributed algorithms assumed that the edge weights, i.e., the entries of $\mathcal{A}_{\mathcal{G}(k)}$, are nonnegative. In Theorem 1, we assume that $\{\mathcal{G}(k), k \geq 0\} \in \Gamma_1$, which implies that the entries of $E[\mathcal{A}_{\mathcal{G}(k)}|\mathcal{F}_{\mathcal{A}}(k-1)]$ are nonnegative almost surely. This relaxation makes the algorithm more flexible at the price of more difficult analysis, since $L_{\mathcal{G}(k)}$ is not a Laplacian matrix anymore and some properties of Laplacian matrices are not applicable.

Remark 8: We call Condition (b.1) $\inf_{m\geq 0} \lambda_{mh}^h \geq \theta$ a.s. the *uniformly conditionally joint connectivity condition*, i.e., the conditional digraphs $\mathcal{G}(k|k-1)$ are jointly connected over the intervals $[mh, (m+1)h-1]$, $m \geq 0$, and the average algebraic connectivity is uniformly positive bounded away from zero.

Remark 9: The inequality (5) gives an upper bound of the mean square steady-state error. There are three terms on the right hand side of (5), which reflect the impacts of additive noises, multiplicative noises and the instantaneous unbalance of network graph on the final steady-state error, respectively. If the network graph is instantaneously balanced, i.e., $\sum_{j=1}^{N} a_{ij}(k) = \sum_{j=1}^{N} a_{ji}(k)$, $i = 1, 2, \ldots, N$, a.s., then the third term vanishes. Especially, if the measurement noise sequence $\{\xi_{ji}(k), k = 0, 1, \ldots, i, j = 1, 2 \ldots, N\}$ are both spatially and temporally independent, then from (44), we get

$$Var(x^*) \leq \frac{4c\beta b^2 \overline{\rho}_1}{N^2} + \frac{8\widetilde{c}\beta\sigma^2\overline{\rho}_1}{N^2}, \tag{6}$$

where $\overline{\rho}_1$ is a positive constant satisfying $\sup_{k\geq 0} \max_{1\leq i,j\leq N} E\left[a_{ij}^2(k)|\mathcal{F}_{\mathcal{A}}(k-1)\right] \leq \overline{\rho}_1$, a.s. Moreover, if $\beta = O(N)$ and $\overline{\rho}_1 = O(1)$ as $N \to \infty$, then $Var(x^*) = O(1/N)$, $N \to \infty$, which means that the larger the number of sensors is, the higher the accuracy of information fusion is. At the same time, a sensor network with a large number of nodes is definitely uneconomic, so there is a trade-off between the performance of the estimation and the cost of the system for selecting the number of nodes.

If the network graph is instantaneously balanced ($\rho_2 = 0$) and the measurement noise intensities are all zeros ($b = \sigma = 0$), then from (5), we get $Var(x^*) = 0$, which means $x^* = \frac{1}{N}\sum_{j=1}^{N} x_j(0)$ almost surely and Theorem 1 degenerates to the case for noise-free average consensus with balanced digraphs in [39].

IV. Proof of Theorem 1

To prove Theorem 1, firstly, we prove that the distance between each agent's state and the centroid of the system vanishes in mean square and almost surely asymptotically. Secondly, we prove that the centroid of the system converges in mean square and almost surely, which then means that each agent's state converges to the same random variable. Finally, we prove that the mathematical expectation of the limit random variable is just the average of the initial states and estimate its variance. Before proving Theorem 1, we have the following two lemmas, which also present some important properties of the consensus error themselves. Lemma 1, whose conditions are weaker than Theorem 1, shows that the mean square of the consensus error, i.e., the distance between each agent's state and the centroid of the system, is bounded. Lemma 2 shows that the consensus error vanishes in mean square and almost surely. Lemma 1 plays important roles in the proofs of Lemma 2 and Theorem 1.

Lemma 1: For the system (1)-(2) and the associated random graph sequence $\{\mathcal{G}(k), k \geq 0\} \in \Gamma_1$, if Assumptions **(A1)-(A3)** and **(A5)** hold and there exists a positive constant ρ_1 such that $\sup_{k\geq 0} E\left[|\mathcal{E}_{\mathcal{G}(k)}| \max_{1\leq i,j\leq N} a_{ij}^2(k)|\mathcal{F}_{\mathcal{A}}(k-1)\right] \leq \rho_1$ a.s., then the system (3) satisfies $\sup_{k\geq 0} E[V(k)] < \infty$.

Proof: By (3) and the definition of $\delta(k)$, we have

$$\begin{aligned}\delta(k+1) &= P_N(I_N - c(k)L_{\mathcal{G}(k)})X(k) \\ &\quad + c(k)P_N D(k)Y(k)\xi(k) \\ &= \delta(k) - c(k)P_N L_{\mathcal{G}(k)}X(k) \\ &\quad + c(k)P_N D(k)Y(k)\xi(k).\end{aligned}$$

By the definition of $L_{\mathcal{G}(k)}$, it follows that $L_{\mathcal{G}(k)}J_N = O_{N\times N}$, and thus $L_{\mathcal{G}(k)}X(k) = L_{\mathcal{G}(k)}\delta(k)$. Then from the above, we have

$$\begin{aligned}\delta(k+1) &= (I_N - c(k)P_N L_{\mathcal{G}(k)})\delta(k) \\ &\quad + c(k)P_N D(k)Y(k)\xi(k),\end{aligned} \tag{7}$$

which together with the definition of $V(k)$ leads to

$$\begin{aligned}V(k+1) &\leq V(k) - 2c(k)\delta^T(k)\frac{L_{\mathcal{G}(k)}^T P_N^T + P_N L_{\mathcal{G}(k)}}{2}\delta(k) \\ &\quad + c^2(k)\|L_{\mathcal{G}(k)}\|^2\|\delta(k)\|^2 \\ &\quad + c^2(k)\xi^T(k)Y^T(k)D^T(k)P_N D(k)Y(k)\xi(k) \\ &\quad + 2c(k)\xi^T(k)Y^T(k)D^T(k)P_N \\ &\quad \times (I_N - c(k)P_N L_{\mathcal{G}(k)})\delta(k).\end{aligned} \tag{8}$$

We now consider the mathematical expectation of each term on the RHS of (8). By Lemma A.1 and Assumption **(A2)**, we know that

$$E[\xi^T(k)Y^T(k)D^T(k)P_N(I_N - c(k)P_N L_{\mathcal{G}(k)})\delta(k)] = 0. \tag{9}$$

Noting that $\mathcal{G}(k|k-1)$ is balanced a.s., by Assumption **(A5)**, we get

$$\begin{aligned}&E\left[\frac{L_{\mathcal{G}(k)}^T P_N^T + P_N L_{\mathcal{G}(k)}}{2}\middle|\mathcal{F}_{\xi,\mathcal{A}}(k-1)\right] \\ &= E\left[\frac{L_{\mathcal{G}(k)}^T P_N^T + P_N L_{\mathcal{G}(k)}}{2}\middle|\mathcal{F}_{\mathcal{A}}(k-1)\right] \\ &= E[\hat{L}_{\mathcal{G}(k)}|\mathcal{F}_{\mathcal{A}}(k-1)] \geq O_{N\times N} \ a.s.,\end{aligned}$$

and then, by $\delta(k) \in \mathcal{F}_{\xi,\mathcal{A}}(k-1)$, we have

$$E\left[\delta^T(k)\frac{L_{\mathcal{G}(k)}^T P_N^T + P_N L_{\mathcal{G}(k)}}{2}\delta(k)\right] \geq 0. \tag{10}$$

By Assumption **(A1)** and the definitions of $Y(k)$ and $V(k)$, we get

$$\begin{aligned}\|Y(k)\|^2 &= \max_{1\leq i,j\leq N}\left(f_{ji}(x_j(k) - x_i(k))\right)^2 \\ &\leq \max_{1\leq i,j\leq N}[2\sigma^2(x_j(k) - x_i(k))^2 + 2b^2] \\ &\leq 4\sigma^2 \max_{1\leq j,i\leq N}\left[\left(x_j(k) - \frac{\sum_{i=1}^N x_i(k)}{N}\right)^2\right. \\ &\quad \left. + \left(x_i(k) - \frac{\sum_{i=1}^N x_i(k)}{N}\right)^2\right] + 2b^2 \\ &\leq 4\sigma^2 \sum_{j=1}^{N}\left(x_j(k) - \frac{\sum_{i=1}^N x_i(k)}{N}\right)^2 + 2b^2 \\ &= 4\sigma^2 V(k) + 2b^2,\end{aligned} \tag{11}$$

where the first "$\leq$" is by Assumption **(A1)** and the definition of $Y(k)$ while the last "$=$" is by the definition of $V(k)$. By Assumptions **(A2)**, the definition of $Y(k)$ and (1), we know that $Y(k)$ is adapted to $\mathcal{F}_{\mathcal{A},\xi}(k-1)$, then by **(A5)** and Lemma A.1, we have

$$\begin{aligned}
&E[\xi^T(k)Y^T(k)D^T(k)P_N D(k)Y(k)\xi(k)]\\
&\quad\leq E[\|Y(k)\|^2\|\xi(k)\|^2\|D^T(k)D(k)\|]\\
&\quad= E[E[\|Y(k)\|^2\|\xi(k)\|^2\|D^T(k)D(k)\||\mathcal{F}_{\mathcal{A},\xi}(k-1)]]\\
&\quad= E[\|Y(k)\|^2 E[\|\xi(k)\|^2|\mathcal{F}_{\xi}(k-1)]\\
&\qquad\times E[\|D^T(k)D(k)\||\mathcal{F}_{\mathcal{A}}(k-1)]]
\end{aligned}$$

which together with (11) and Assumption **(A2)** leads to

$$\begin{aligned}
&E[\xi^T(k)Y^T(k)D^T(k)P_N D(k)Y(k)\xi(k)]\\
&\quad\leq E[\|Y(k)\|^2\|\xi(k)\|^2\|D^T(k)D(k)\|]\\
&\quad\leq \beta E[(4\sigma^2 V(k)+2b^2)E[\|D^T(k)D(k)\||\mathcal{F}_{\mathcal{A}}(k-1)]]\\
&\quad= \beta E[(4\sigma^2 V(k)+2b^2)\\
&\qquad\times E[\lambda_{max}(D^T(k)D(k))|\mathcal{F}_{\mathcal{A}}(k-1)]]\\
&\quad= \beta E\Big[(4\sigma^2 V(k)+2b^2)\\
&\qquad\times E\Big[\max_{1\leq i\leq N}\lambda_{max}(\alpha_i(k)\alpha_i^T(k))|\mathcal{F}_{\mathcal{A}}(k-1)\Big]\Big]\\
&\quad= \beta E\Big[(4\sigma^2 V(k)+2b^2)\\
&\qquad\times E\Big[\max_{1\leq i\leq N}\mathbf{tr}(\alpha_i^T(k)\alpha_i(k))|\mathcal{F}_{\mathcal{A}}(k-1)\Big]\Big]\\
&\quad\leq \beta E\Big[(4\sigma^2 V(k)+2b^2)\\
&\qquad\times E\Big[|\mathcal{E}_{\mathcal{G}(k)}|\max_{1\leq i,j\leq N}a_{ij}^2(k)|\mathcal{F}_{\mathcal{A}}(k-1)\Big]\Big]\\
&\quad\leq 4\sigma^2\beta\rho_1 E[V(k)]+2b^2\beta\rho_1. \qquad (12)
\end{aligned}$$

From the above, taking the mathematical expectation on both sides of (8), by (9), (10) and $\sup_{k\geq 0} E\Big[|\mathcal{E}_{\mathcal{G}(k)}|\max_{1\leq i,j\leq N} a_{ij}^2(k)|\mathcal{F}_{\mathcal{A}}(k-1)\Big] \leq \rho_1$ a.s., we get

$$\begin{aligned}
E[V(k+1)] \leq{}& [1+c^2(k)(\rho_0^2+4\beta\sigma^2\rho_1)]E[V(k)]\\
&+2b^2\beta\rho_1 c^2(k), \quad k\geq 0. \qquad (13)
\end{aligned}$$

This together with Assumption **(A3)** and Lemma A.2 gives that $E[V(k)]$ is bounded (regarding $E[V(k)]$ as $x(k)$ in Lemma A.2). □

Lemma 2: For the system (1)-(2) and the associated random graph sequence $\{\mathcal{G}(k), k\geq 0\}\in\Gamma_1$, assume that (a) Assumptions **(A1)-(A5)** hold; (b) there exist deterministic positive integer h, positive constants θ and ρ_0, such that (b.1) $\inf_{m\geq 0}\lambda_{mh}^h \geq \theta$ a.s., (b.2) $\sup_{k\geq 0}\Big[E[\|L_{\mathcal{G}(k)}\|^{2^{\max\{h,2\}}}|\mathcal{F}_{\mathcal{A}}(k-1)]\Big]^{\frac{1}{2^{\max\{h,2\}}}} \leq \rho_0$ a.s. Then, $\lim_{k\to\infty} E[V(k)] = 0$ and $\lim_{k\to\infty} V(k) = 0$ a.s.

Proof: Let $\Phi(m,n) = (I_N - c(m-1)P_N L_{\mathcal{G}(m-1)})\cdots(I_N - c(n)P_N L_{\mathcal{G}(n)})$, $m > n \geq 0$, $\Phi(n,n) = I_N$, $n\geq 0$. By (7) and some iterative calculations, we get

$$\delta((m+1)h) = \Phi((m+1)h, mh)\delta(mh) + \tilde{\xi}_m^{mh}, \ m\geq 0,$$

where

$$\tilde{\xi}_m^{mh} = \sum_{j=mh}^{(m+1)h-1} c(j)\Phi((m+1)h, j+1) \times P_N D(j)Y(j)\xi(j). \qquad (14)$$

From the definition of $V(k)$, it follows that

$$\begin{aligned}
&V((m+1)h)\\
&\quad= \delta^T(mh)\Phi^T((m+1)h, mh)\Phi((m+1)h, mh)\delta(mh)\\
&\qquad+(\tilde{\xi}_m^{mh})^T(\tilde{\xi}_m^{mh})\\
&\qquad+2\delta^T(mh)\Phi^T((m+1)h, mh)\tilde{\xi}_m^{mh}\\
&\quad= \delta^T(mh)\Big[\Phi^T((m+1)h, mh)\Phi((m+1)h, mh) - I_N\\
&\qquad+\sum_{i=mh}^{(m+1)h-1} c(i)[P_N L_{\mathcal{G}(i)} + L_{\mathcal{G}(i)}^T P_N^T]\Big]\delta(mh)\\
&\qquad+V(mh)\\
&\qquad-\delta^T(mh)\sum_{i=mh}^{(m+1)h-1} c(i)[P_N L_{\mathcal{G}(i)} + L_{\mathcal{G}(i)}^T P_N^T]\delta(mh)\\
&\qquad+(\tilde{\xi}_m^{mh})^T(\tilde{\xi}_m^{mh})\\
&\qquad+2\delta^T(mh)\Phi^T((m+1)h, mh)\tilde{\xi}_m^{mh}. \qquad (15)
\end{aligned}$$

We now consider the mathematical expectation of each term on the RHS of (15). Noting that $\delta(mh)\in\mathcal{F}_{\xi,\mathcal{A}}(mh-1)$, by the properties of conditional expectation, we know that

$$\begin{aligned}
&E\Big[\delta^T(mh)\Phi^T((m+1)h, mh)\\
&\qquad\times\Phi((m+1)h, j+1)P_N D(j)Y(j)\xi(j)\Big]\\
&\quad= E\Big[\delta^T(mh)E\Big[\Phi^T((m+1)h, mh)\Phi((m+1)h, j+1)\\
&\qquad\times P_N D(j)Y(j)\xi(j)|\mathcal{F}_{\xi,\mathcal{A}}(j-1)\Big]\Big],\\
&\qquad mh\leq j\leq (m+1)h-1, \quad m\geq 0. \qquad (16)
\end{aligned}$$

By Assumptions **(A2)**, **(A5)** and Lemma A.1, we have

$$\begin{aligned}
&E\Big[\Phi^T((m+1)h, mh)\Phi((m+1)h, j+1)\\
&\qquad\times P_N D(j)Y(j)\xi(j)|\mathcal{F}_{\xi,\mathcal{A}}(j-1)\Big]\\
&\quad= E\Big[\Phi^T((m+1)h, mh)\Phi((m+1)h, j+1)\\
&\qquad\times P_N D(j)|\mathcal{F}_{\xi,\mathcal{A}}(j-1)\Big]Y(j)E[\xi(j)|\mathcal{F}_{\xi,\mathcal{A}}(j-1)]\\
&\quad= E\Big[\Phi^T((m+1)h, mh)\Phi((m+1)h, j+1)\\
&\qquad\times P_N D(j)|\mathcal{F}_{\mathcal{A}}(j-1)\Big]Y(j)E[\xi(j)|\mathcal{F}_{\xi}(j-1)]\\
&\quad= \mathbf{0}_{N\times N}, \quad mh\leq j\leq (m+1)h-1, \ m\geq 0,
\end{aligned}$$

where the second "$=$" is obtained from **(A5)** and Lemma A.1. This together with (14) and (16) gives

$$E\left[\delta^T(mh)\Phi^T((m+1)h, mh)\tilde{\xi}_m^{mh}\right] = 0. \qquad (17)$$

By Assumptions **(A3)** and **(A4)**, there exist positive integer m_0 and positive constant C_1, such that $c^2(mh) \le C_1 c^2((m+1)h)$, $\forall\ m \ge m_0$, and $c(k) \le 1$, $\forall\ k \ge m_0 h$. By Condition (b.2) and the conditional Lyapunov inequality, we obtain that

$$\begin{aligned}&\sup_{k\ge 0} E[\|L_{\mathcal{G}(k)}\|^i|\mathcal{F}_{\mathcal{A}}(k-1)]\\ &\le \sup_{k\ge 0}[E[\|L_{\mathcal{G}(k)}\|^{2^h}|\mathcal{F}_{\mathcal{A}}(k-1)]]^{\frac{i}{2^h}} \le \rho_0^i \text{ a.s.}, \quad \forall\ 2 \le i \le 2^h.\end{aligned} \tag{18}$$

Denote the combinatorial number of choosing i elements from $2h$ elements by M_{2h}^i. By termwise multiplication and using the Hölder inequality repeatedly, noting that $c(mh)$ decreasing monotonously as m increases and $E[\|L_{\mathcal{G}(k)}\|^l|\mathcal{F}_{\mathcal{A}}(mh-1)] = E[E[\|L_{\mathcal{G}(k)}\|^l|\mathcal{F}_{\mathcal{A}}(k-1)]|\mathcal{F}_{\mathcal{A}}(mh-1)]$, $2 \le l \le 2^h$, $k \ge mh$, from (18), we have

$$\begin{aligned}&E\Big[\Big\|\Phi^T((m+1)h, mh)\Phi((m+1)h, mh) - I_N\\ &\qquad + \sum_{i=mh}^{(m+1)h-1} c(i)(P_N L_{\mathcal{G}(i)} + L_{\mathcal{G}(i)}^T P_N^T)\Big\|\Big|\mathcal{F}_{\mathcal{A}}(mh-1)\Big]\\ &\le \Big(C_1 \sum_{i=2}^{2h} M_{2h}^i \rho_0^i\Big) c^2((m+1)h)\\ &= C_1[(1+\rho_0)^{2h} - 1 - 2h\rho_0]c^2((m+1)h),\\ &\quad m \ge m_0.\end{aligned} \tag{19}$$

Denote the symmetrized graph of $\mathcal{G}(i|mh-1)$ by $\hat{\mathcal{G}}(i|mh-1)$, $mh \le i \le (m+1)h-1$. Noting that $\mathcal{G}(i|i-1)$ is balanced a.s., we know that $\mathcal{G}(i|mh-1)$ is balanced a.s. Then, $E[\hat{L}_{\mathcal{G}(i)}|\mathcal{F}_{\mathcal{A}}(mh-1)]$ is the Laplacian matrix of $\hat{\mathcal{G}}(i|mh-1)$, a.s., $mh \le i \le (m+1)h-1$. Therefore, $\sum_{i=mh}^{(m+1)h-1} E[\hat{L}_{\mathcal{G}(i)}|\mathcal{F}_{\mathcal{A}}(mh-1)]$ is the Laplacian matrix of $\sum_{i=mh}^{(m+1)h-1} \hat{\mathcal{G}}(i|mh-1)$ a.s.. Furthermore, by Assumption **(A5)** and Lemma A.1, we have

$$\begin{aligned}&E\Big[\delta^T(mh)\Big[\sum_{i=mh}^{(m+1)h-1} c(i)(P_N L_{\mathcal{G}(i)}\\ &\qquad + L_{\mathcal{G}(i)}^T P_N^T)\Big]\delta(mh)\Big]\\ &= 2E\Big[\delta^T(mh)\Big[\sum_{i=mh}^{(m+1)h-1} c(i)E[\hat{L}_{\mathcal{G}(i)}|\mathcal{F}_{\xi,\mathcal{A}}(mh-1)]\Big]\\ &\qquad \times \delta(mh)\Big]\\ &= 2E\Big[\delta^T(mh)\Big[\sum_{i=mh}^{(m+1)h-1} c(i)E[\hat{L}_{\mathcal{G}(i)}|\mathcal{F}_{\mathcal{A}}(mh-1)]\Big]\\ &\qquad \times \delta(mh)\Big],\end{aligned}$$

which together with Assumption **(A4)** and Condition (b.1) leads to

$$\begin{aligned}&E\Big[\delta^T(mh)\Big[\sum_{i=mh}^{(m+1)h-1} c(i)(P_N L_{\mathcal{G}(i)} + L_{\mathcal{G}(i)}^T P_N^T)\Big]\delta(mh)\Big]\\ &\ge 2c((m+1)h)E\Big[\delta^T(mh)\\ &\qquad \times \Big[\sum_{i=mh}^{(m+1)h-1} E[\hat{L}_{\mathcal{G}(i)}|\mathcal{F}_{\mathcal{A}}(mh-1)]\Big]\delta(mh)\Big]\\ &\ge 2c((m+1)h)E\Big[\lambda_{mh}^h V(mh)\Big]\\ &\ge 2c((m+1)h)E\Big[\inf_{m\ge 0}(\lambda_{mh}^h)V(mh)\Big]\\ &\ge 2\theta c((m+1)h)E[V(mh)] \text{ a.s.}\end{aligned} \tag{20}$$

By Assumptions **(A2)**, **(A5)** and Lemma A.1, it follows that

$$\begin{aligned}&E[\xi^T(i)Y^T(i)D^T(i)P_N\Phi^T((m+1)h, i+1)\\ &\quad \times \Phi((m+1)h, j+1)P_N D(j)Y(j)\xi(j)]\\ &= E[E[\xi^T(i)Y^T(i)D^T(i)P_N\Phi^T((m+1)h, i+1)\\ &\quad \times \Phi((m+1)h, j+1) \mid \mathcal{F}_{\xi,\mathcal{A}}(i-1)]P_N D(j)Y(j)\xi(j)]\\ &= E[E[\xi^T(i)Y^T(i) \mid \mathcal{F}_{\xi,\mathcal{A}}(i-1)]\\ &\quad \times E[D^T(i)P_N\Phi^T((m+1)h, i+1)\\ &\quad \times \Phi((m+1)h, j+1) \mid \mathcal{F}_{\mathcal{A}}(i-1)]P_N D(j)Y(j)\xi(j)]\\ &= E[E[E[\xi^T(i) \mid \mathcal{F}_{\xi}(i-1)]Y^T(i) \mid \mathcal{F}_{\xi,\mathcal{A}}(i-1)]\\ &\quad \times E[D^T(i)P_N\Phi^T((m+1)h, i+1)\\ &\quad \times \Phi((m+1)h, j+1) \mid \mathcal{F}_{\mathcal{A}}(i-1)]P_N D(j)Y(j)\xi(j)]\\ &= 0,\ i > j,\end{aligned}$$

which together with the definition of $\tilde{\xi}_m^{mh}$ gives

$$\begin{aligned}&E[(\tilde{\xi}_m^{mh})^T(\tilde{\xi}_m^{mh})]\\ &= \sum_{i=mh}^{(m+1)h-1} c^2(i)E[\xi^T(i)Y^T(i)D^T(i)P_N\\ &\qquad \times\Phi^T((m+1)h, i+1)\Phi((m+1)h, i+1)\\ &\qquad \times P_N D(i)Y(i)\xi(i)]\\ &\le \sum_{i=mh}^{(m+1)h-1} c^2(i)E[\|\Phi^T((m+1)h, i+1)\\ &\qquad \times \Phi((m+1)h, i+1)\|\|D^T(i)D(i)\|\|Y(i)\|^2\|\xi(i)\|^2]\\ &= \sum_{i=mh}^{(m+1)h-1} c^2(i)E[\|Y(i)\|^2 E[\|\Phi^T((m+1)h, i+1)\\ &\qquad \times \Phi((m+1)h, i+1)\|\|D^T(i)D(i)\| \mid \mathcal{F}_{\mathcal{A}}(i-1)]\\ &\qquad \times E[\|\xi(i)\|^2|\mathcal{F}_{\xi}(i-1)]].\end{aligned} \tag{21}$$

By Condition (b.2), there is a constant ρ_1' such that

$$\sup_{k\ge 0}\Big[E[\|D^T(k)D(k)\|^2|\mathcal{F}_{\mathcal{A}}(k-1)]\Big]^{1/2} \le \rho_1' \text{ a.s.},$$

which together with the conditional Hölder inequality and Cr-inequality leads to

$$\begin{aligned}&E[\|\Phi^T((m+1)h, i+1)\Phi((m+1)h, i+1)\|\\ &\qquad \times \|D^T(i)D(i)\||\mathcal{F}_{\mathcal{A}}(i-1)]\\ &\le \rho_1'\{E[\|\Phi^T((m+1)h, i+1)\\ &\qquad \times \Phi((m+1)h, i+1)\|^2|\mathcal{F}_{\mathcal{A}}(i-1)]\}^{\frac{1}{2}}\\ &\le \rho',\ mh \le i \le (m+1)h-1,\ m \ge m_0,\end{aligned}$$

where $\rho' = \rho_1' \left\{ \left(\sum_{j=0}^{2(h-1)} M_{2(h-1)}^j \right) \sum_{l=0}^{2(h-1)} M_{2(h-1)}^l \rho_0^{2l} \right\}^{\frac{1}{2}}$. Then, by (11), (21) and the above, we get

$$\begin{aligned} &E[(\tilde{\xi}_m^{mh})^T(\tilde{\xi}_m^{mh})] \\ &\quad \le \rho' \sum_{i=mh}^{(m+1)h-1} c^2(i) E[4\sigma^2 V(i) E[\|\xi(i)\|^2 \mid \mathcal{F}_\xi(i-1)] \\ &\qquad + 2b^2 E[\|\xi(i)\|^2 \mid \mathcal{F}_\xi(i-1)]] \\ &\quad \le 4\sigma^2\beta\rho' \sum_{i=mh}^{(m+1)h-1} c^2(i) E[V(i)] \\ &\qquad + 2b^2\beta\rho' \sum_{i=mh}^{(m+1)h-1} c^2(i), \ m \ge m_0. \end{aligned} \tag{22}$$

Finally, by (15), (17), (19), (20) and (22), we have

$$\begin{aligned} &E[V((m+1)h)] \\ &\le \bigg(1 - 2\theta c((m+1)h) \\ &\quad + c^2((m+1)h) C_1[(1+\rho_0)^{2h} - 1 - 2h\rho_0]\bigg) E[V(mh)] \\ &\quad + 4\sigma^2\beta\rho' \sum_{i=mh}^{(m+1)h-1} c^2(i) E[V(i)] \\ &\quad + 2b^2\beta\rho' \sum_{i=mh}^{(m+1)h-1} c^2(i), \ m \ge m_0. \end{aligned} \tag{23}$$

We call (23) the difference inequality of stochastic Lyapunov function. Now we first prove that $E[V(mh)] \to 0$, $m \to \infty$, and then prove that $V(k) \to 0$, $k \to \infty$ a.s. By (23), Lemma 1 and (23), we have

$$\begin{aligned} &E[V((m+1)h)] \\ &\quad \le (1 - 2\theta c((m+1)h) \\ &\qquad + c^2((m+1)h) C_1[(1+\rho_0)^{2h} - 1 - 2h\rho_0]) E[V(mh)] \\ &\qquad + C_2 \sum_{i=mh}^{(m+1)h-1} c^2(i), \ m \ge m_0, \end{aligned} \tag{24}$$

where $C_2 = (4\sigma^2 \sup_{k\ge 0} E[V(k)] + 2b^2)\beta\rho'$.

By Assumption **(A3)**, there exists positive integer m_1 such that

$$\begin{aligned} 0 < 2\theta c((m+1)h) - c^2((m+1)h) C_1[(1+\rho_0)^{2h} \\ - 1 - 2h\rho_0] \le 1, \quad \forall\, m \ge m_1, \end{aligned} \tag{25}$$

and

$$\begin{aligned} &\sum_{m=0}^{\infty} \{2\theta c((m+1)h) \\ &\qquad - c^2((m+1)h) C_1[(1+\rho_0)^{2h} - 1 - 2h\rho_0]\} \\ &\quad = \infty. \end{aligned} \tag{26}$$

And by Assumption **(A4)**, we get

$$\begin{aligned} &\lim_{m\to\infty} \left\{ \left[C_2 \sum_{i=mh}^{(m+1)h-1} c^2(i) \right] / [2\theta c((m+1)h) \right. \\ &\quad \left. - c^2((m+1)h) C_1[(1+\rho_0)^{2h} - 1 - 2h\rho_0]] \right\} = 0. \end{aligned} \tag{27}$$

Then by Lemma A.3 and (24)-(27), we get $E[V(mh)] \to 0$, $m \to \infty$. Thus, for any given $\epsilon > 0$, there exists positive integer m_2 such that $E[V(mh)] < \epsilon$, $m \ge m_2$, and $\sum_{i=m_2h}^{\infty} c^2(i) < \epsilon$. Let $m_k = \lfloor \frac{k}{h} \rfloor$. Then for any given $k \ge m_2 h$, we have $m_k \ge m_2$ and $0 \le k - m_k h \le h$. Therefore, by (13) we have

$$\begin{aligned} &E[V(k+1)] \\ &\le \prod_{i=m_kh}^{k} [1 + c^2(i)(\rho_0^2 + 4\rho_1\beta\sigma^2)] E[V(m_kh)] \\ &\quad + 2\rho_1 b^2 \beta \sum_{i=m_kh}^{k} \prod_{j=i+1}^{k} [1 + c^2(j)(\rho_0^2 + 4\rho_1\beta\sigma^2)] c^2(i) \\ &\le \exp((\rho_0^2 + 4\rho_1\beta\sigma^2) \sum_{i=0}^{\infty} c^2(i))(1 + 2\rho_1 b^2\beta)\epsilon, \quad k \ge m_2 h, \end{aligned} \tag{28}$$

where $\prod_{j=k+1}^{k}[1+(\rho_0^2+4\rho_1\beta\sigma^2)c^2(j)]$ is defined as 1. Then, by the arbitrariness of ϵ, we get

$$E[V(k)] \to 0, k \to \infty. \tag{29}$$

Taking conditional expectation on both sides of (8) gives

$$\begin{aligned} &E[V(k+1) | \mathcal{F}_{\xi,\mathcal{A}}(k-1)] \\ &\qquad \le [1 + c^2(k)(\rho_0^2 + 4\sigma^2\rho_1\beta)] V(k) + 2b^2\rho_1\beta c^2(k). \end{aligned}$$

Then, by Lemma A.2 and Assumption **(A3)**, we obtain

$$V(k) \to \text{a finite random variable}, \ k \to \infty \text{ a.s.},$$

which together with (29) gives $V(k) \to 0$, $k \to \infty$ a.s. □

Proof of Theorem 1: Firstly, if Condition (b.2) holds, noting that

$$\begin{aligned} &|\mathcal{E}_{\mathcal{G}(k)}| \max_{1\le i,j\le N} a_{ij}^2(k) \\ &\qquad 0 \le N(N-1) \max_{1\le i,j\le N} a_{ij}^2(k) \le N(N-1) \| \| L_{\mathcal{G}(k)} \|_F^2, \end{aligned}$$

by the equivalence of 2-norm and Frobenius norm of matrices and the conditional Lyapunov inequality, we know that the deterministic constants ρ_1 and ρ_2 are both well defined. Secondly, by Lemma 2, we directly get that $\delta(k)$ vanishes in mean square and almost surely as $k \to \infty$. Then, this theorem is proved by three **Steps** as follows.

Step 1: To prove that all $x_i(k)$, $i \in \mathcal{V}$ converge to x^* as $k \to \infty$ in mean square and almost surely.

Let $\tilde{L}_{\mathcal{G}(k)} = L_{\mathcal{G}(k)} - E[L_{\mathcal{G}(k)} | \mathcal{F}_\mathcal{A}(k-1)]$, $k \ge 0$. Noting that the associated digraph of Laplacian matrix $E[L_{\mathcal{G}(k)}|\mathcal{F}_\mathcal{A}(k-1)]$ is balanced a.s., we know that $\mathbf{1}^T E[L_{\mathcal{G}(k)} | \mathcal{F}_\mathcal{A}(k-1)] = \mathbf{0}_N^T$ a.s. Left multiplying with $\frac{1}{N}\mathbf{1}_N^T$ on both sides of (3), and then making a summation from 0 to $n-1$ with respect to k, we have

$$\begin{aligned} \frac{1}{N}\sum_{j=1}^{N} x_j(n) = \frac{1}{N}\sum_{j=1}^{N} x_j(0) - \frac{1}{N}\mathbf{1}^T \sum_{k=0}^{n-1} c(k) L_{\mathcal{G}(k)} X(k) \\ + \frac{1}{N}\mathbf{1}^T \sum_{k=0}^{n-1} c(k) D(k) Y(k) \xi(k) \end{aligned}$$

$$= \frac{1}{N}\sum_{j=1}^{N} x_j(0) - \frac{1}{N}\mathbf{1}^T \sum_{k=0}^{n-1} c(k)\tilde{L}_{\mathcal{G}(k)}X(k) + \frac{1}{N}\mathbf{1}^T \sum_{k=0}^{n-1} c(k)D(k)Y(k)\xi(k). \quad (30)$$

Noting that

$$\begin{aligned}
&E[\tilde{L}_{\mathcal{G}(m+i)}X(m+i)|\mathcal{F}_{\xi,\mathcal{A}}(m)] \\
&= E[E[\tilde{L}_{\mathcal{G}(m+i)}X(m+i)|\mathcal{F}_{\xi,\mathcal{A}}(m)]|\mathcal{F}_{\xi,\mathcal{A}}(m+i-1)] \\
&= E[E[\tilde{L}_{\mathcal{G}(m+i)}X(m+i)|\mathcal{F}_{\xi,\mathcal{A}}(m+i-1)]|\mathcal{F}_{\xi,\mathcal{A}}(m)] \\
&= E[E[\tilde{L}_{\mathcal{G}(m+i)}|\mathcal{F}_{\xi,\mathcal{A}}(m+i-1)]X(m+i)|\mathcal{F}_{\xi,\mathcal{A}}(m)], \\
&\quad 1 \le i \le n-m-1,
\end{aligned}$$

by the definition of $\tilde{L}_{\mathcal{G}(k)}$ and Assumption **(A5)**, it is known that $E[\tilde{L}_{\mathcal{G}(k)}|\mathcal{F}_{\xi,\mathcal{A}}(k-1)] = E[\tilde{L}_{\mathcal{G}(k)}|\mathcal{F}_{\mathcal{A}}(k-1)] = O_{N\times N}$, $k \ge 0$. Thus, from the above equality, we get

$$E[\tilde{L}_{\mathcal{G}(m+i)}X(m+i)|\mathcal{F}_{\xi,\mathcal{A}}(m)] = \mathbf{0}_N,\ 1 \le i \le n-m-1,$$

which gives

$$\begin{aligned}
&E\left[\sum_{k=0}^{n-1}\tilde{L}_{\mathcal{G}(k)}X(k)\middle|\mathcal{F}_{\xi,\mathcal{A}}(m)\right] \\
&= E\left[\sum_{i=0}^{m}\tilde{L}_{\mathcal{G}(i)}X(i)\middle|\mathcal{F}_{\xi,\mathcal{A}}(m)\right] + E\left[\sum_{i=m+1}^{n-1}\tilde{L}_{\mathcal{G}(i)}X(i)\middle|\mathcal{F}_{\xi,\mathcal{A}}(m)\right] \\
&= E\left[\sum_{i=0}^{m}\tilde{L}_{\mathcal{G}(i)}X(i)\middle|\mathcal{F}_{\xi,\mathcal{A}}(m)\right], \quad \forall\ m < n-1.
\end{aligned}$$

This together with the definition of martingales implies $\left\{\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{n} c(k)\tilde{L}_{\mathcal{G}(k)}X(k), \mathcal{F}_{\xi,\mathcal{A}}(n), n \ge 0\right\}$ is a martingale. On the other hand, by (30), we know that

$$\begin{aligned}
&\sup_{n\ge 0} E\left\|\sum_{k=0}^{n-1} c(k)\tilde{L}_{\mathcal{G}(k)}X(k)\right\|^2 \\
&\le \sup_{n\ge 0}\sum_{k=0}^{n-1} c^2(k)E[\|X(k)\|^2\|\tilde{L}_{\mathcal{G}(k)}\|^2] \\
&\le \sup_{k\ge 0} E[\|\tilde{L}_{\mathcal{G}(k)}\|^2|\mathcal{F}_{\mathcal{A}}(k-1)] \\
&\quad \times \sup_{k\ge 0} E\|X(k)\|^2 \sum_{k=0}^{\infty} c^2(k). \quad (31)
\end{aligned}$$

By Condition (b.2), we know that

$$\sup_{k\ge 0} E[\|\tilde{L}_{\mathcal{G}(k)}\|^2|\mathcal{F}_{\mathcal{A}}(k-1)] < \infty \text{ a.s.} \quad (32)$$

From (3), (12) and Condition (b.2), we get

$$\begin{aligned}
&E[\|X(k+1)\|^2] \\
&= E[X^T(k)(I_N - c(k)L^T_{\mathcal{G}(k)})(I_N - c(k)L_{\mathcal{G}(k)})X(k)] \\
&\quad + c^2(k)E[\xi^T(k)Y^T(k)D^T(k)D(k)Y(k)\xi(k)] \\
&\le E[\|X(k)\|^2] + c^2(k)E[\|X(k)\|^2\|L_{\mathcal{G}(k)}\|^2] \\
&\quad + c^2(k)E[\|Y(k)\|^2\|\xi(k)\|^2\|D^T(k)D(k)\|] \\
&\le E[\|X(k)\|^2] + c^2(k)\rho_0^2\ E[\|X(k)\|^2] \\
&\quad + c^2(k)\beta\rho_1 E[4\sigma^2 V(k) + 2b^2] \\
&\le (1 + c^2(k)\rho_0^2)E[\|X(k)\|^2] \\
&\quad + \beta\rho_1(4\sigma^2 \sup_{k\ge 0} E[V(k)] + 2b^2)c^2(k), \quad (33)
\end{aligned}$$

where the second term in the second inequality is by Condition (b.2), and the third term is similar to (12). This together with Lemma 1, Lemma A.2 and Assumption **(A3)** gives $\sup_{k\ge 0} E[\|X(k)\|^2] < \infty$. Then, by (31) and (32), we know that

$$\sup_{n\ge 0} E\left\|\sum_{k=0}^{n-1} c(k)\tilde{L}_{\mathcal{G}(k)}X(k)\right\|^2 < \infty.$$

This together with Lemma A.4 leads to the fact that

$$\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{n-1} c(k)\tilde{L}_{\mathcal{G}(k)}X(k) \quad \text{converges a.s. and in mean square.} \quad (34)$$

From Assumptions **(A2)** and **(A5)**, it follows that

$$\begin{aligned}
&E\left[\sum_{k=0}^{n-1} c(k)D(k)Y(k)\xi(k)|\mathcal{F}_{\xi,\mathcal{A}}(j)\right] \\
&= \sum_{k=0}^{j} c(k)D(k)Y(k)\xi(k) + \sum_{k=j+1}^{n-1} E[E(c(k)D(k) \\
&\quad \times Y(k)\xi(k)|\mathcal{F}_{\xi,\mathcal{A}}(k-1))|\mathcal{F}_{\xi,\mathcal{A}}(j)] \\
&= \sum_{k=0}^{j} c(k)D(k)Y(k)\xi(k), \forall\ j < n-1.
\end{aligned}$$

Thus, the adaptive sequence $\left\{\sum_{j=0}^{n} c(k)D_{\mathcal{G}(k)}Y(k)\xi(k), \mathcal{F}_{\xi,\mathcal{A}}(n), n \ge 0\right\}$ is a martingale. Then, by (11) and Condition (b.2), we have

$$\begin{aligned}
&\sup_{n\ge 0} E\left\|\sum_{k=0}^{n-1} c(k)D(k)Y(k)\xi(k)\right\|^2 \\
&= \sup_{n\ge 0}\sum_{k=0}^{n-1} E\left[c^2(k)\xi^T(k)Y^T(k)D^T(k)D(k)Y(k)\xi(k)\right] \\
&\le \beta \sup_{k\ge 0} E[\|D^T(k)D(k)\||\mathcal{F}_{\mathcal{A}}(k-1)] \\
&\quad \times \sup_{n\ge 0}\sum_{k=0}^{n-1} c^2(k)E\|Y(k)\|^2 \\
&\le \beta\rho_1 \sup_{n\ge 0}\sum_{k=0}^{n-1} c^2(k)(4\sigma^2\ E[V(k)] + 2b^2).
\end{aligned}$$

By Assumption **(A3)**, the boundedness of $E[V(k)]$ and the above, we get

$$\sup_{n\ge 0} E\left\|\sum_{k=0}^{n-1} c(k)D(k)Y(k)\xi(k)\right\|^2 < \infty,$$

which together with Lemma A.4 gives

$$\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{n-1}c(k)D(k)Y(k)\xi(k) \text{ converges, } k\to\infty \text{ a.s. and in mean square.} \quad (35)$$

Finally, by (30), (34) and (35) we know that

$$\frac{1}{N}\sum_{j=1}^{N}x_j(n)\to x^*,\ n\to\infty \text{ a.s. and in mean square,} \quad (36)$$

where

$$x^*=\frac{1}{N}\sum_{j=1}^{N}x_j(0)-\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{\infty}c(k)\tilde{L}_{\mathcal{G}(k)}X(k)+\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{\infty}c(k)D(k)Y(k)\xi(k). \quad (37)$$

Then, by the definition of $V(k)$, Lemma 2 and (36), we have

$$x_i(k)\to x^*,\ k\to\infty,\ \text{a.s. and in mean square, } i\in\mathcal{V}.$$

Step 2: To compute the mathematical expectation of x^*.

By (34), we have

$$E\left[\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{\infty}c(k)\tilde{L}_{\mathcal{G}(k)}X(k)\right]=\lim_{n\to\infty}E\left[\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{n-1}c(k)\tilde{L}_{\mathcal{G}(k)}X(k)\right]=0.$$

Similarly, by (35), we have

$$E\left[\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{\infty}c(k)D(k)Y(k)\xi(k)\right]=\lim_{n\to\infty}E\left[\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{n-1}c(k)D(k)Y(k)\xi(k)\right]=0.$$

This together with (37) gives

$$E(x^*)=\frac{1}{N}\sum_{j=1}^{N}x_j(0). \quad (38)$$

Step 3: To estimate the variance of x^*.

From (13), by iterative calculations, we have

$$E[V(k+1)]\le\prod_{i=0}^{k}[1+(\rho_0^2+4\beta\sigma^2\rho_1)c^2(i)]V(0)+2\rho_1 b^2\beta\sum_{i=0}^{k}c^2(i)\prod_{j=i+1}^{k}[1+(\rho_0^2+4\beta\sigma^2\rho_1)c^2(j)], \quad (39)$$

where $\prod_{j=k+1}^{k}[1+(\rho_0^2+4\beta\sigma^2\rho_1)c^2(j)]=1$. Actually, for all $k\ge j$, we have $\prod_{i=j}^{k}(1+(\rho_0^2+4\beta\sigma^2\rho_1)c^2(i))\le\exp\Big((\rho_0^2+4\beta\sigma^2\rho_1)\sum_{i=j}^{k}c^2(i)\Big)\le\exp\Big((\rho_0^2+4\beta\sigma^2\rho_1)\sum_{i=0}^{\infty}c^2(i)\Big)$. This together with (39) leads to

$$\sup_{k\ge0}E[V(k)]\le q_v. \quad (40)$$

Similarly, by (33) and the above, we have

$$E\|X(k+1)\|^2\le(1+c^2(k)\rho_0^2)E\|X(k)\|^2+\beta\rho_1(4\sigma^2 q_v+2b^2)c^2(k)\le q_x. \quad (41)$$

Then, by (34), (35), (37), (38), the dominated convergence theorem and Cr-inequality, we have

$$\begin{aligned}Var(x^*)&=E\Big[\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{\infty}c(k)D(k)Y(k)\xi(k)-\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{\infty}c(k)\tilde{L}_{\mathcal{G}(k)}X(k)\Big]^2\\&\le 2E\Big[\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{\infty}c(k)D(k)Y(k)\xi(k)\Big]^2+2E\Big[\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{\infty}c(k)\tilde{L}_{\mathcal{G}(k)}X(k)\Big]^2\\&\le 2\lim_{n\to\infty}E\Big[\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{n-1}c(k)\tilde{L}_{\mathcal{G}(k)}X(k)\Big]^2+2\lim_{n\to\infty}E\Big[\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{n-1}c(k)D(k)Y(k)\xi(k)\Big]^2. \end{aligned}\quad (42)$$

For the first term on the right hand side of (42), noting that $\Big\{\mathbf{1}_N^T\sum_{k=0}^{n}c(k)\tilde{L}_{\mathcal{G}(k)}X(k),\ \mathcal{F}_{\xi,\mathcal{A}}(n),\ n\ge0\Big\}$ is a martingale, we have

$$\begin{aligned}&\lim_{n\to\infty}E\left[\frac{1}{N}\mathbf{1}_N^T\sum_{k=0}^{n-1}c(k)\tilde{L}_{\mathcal{G}(k)}X(k)\right]^2\\&=\frac{1}{N^2}\lim_{n\to\infty}\sum_{k=0}^{n-1}\left\{c^2(k)E\left[\mathbf{1}_N^T\tilde{L}_{\mathcal{G}(k)}X(k)\right]^2\right\}\\&=\frac{1}{N^2}\lim_{n\to\infty}\sum_{k=0}^{n-1}\left\{c^2(k)E\left[\mathbf{1}_N^T L_{\mathcal{G}(k)}X(k)\right]^2\right\}\\&=\frac{1}{N^2}\lim_{n\to\infty}\sum_{k=0}^{n-1}\left\{c^2(k)E\Big[\sum_{i=1}^{N}x_i(k)\Big(\sum_{j=1}^{N}a_{ij}(k)-\sum_{j=1}^{N}a_{ji}(k)\Big)\Big]^2\right\}\\&\le\frac{1}{N}\lim_{n\to\infty}\sum_{k=0}^{n-1}\left\{c^2(k)\sum_{i=1}^{N}E\Big[x_i^2(k)\Big(\sum_{j=1}^{N}a_{ij}(k)-\sum_{j=1}^{N}a_{ji}(k)\Big)^2\Big]\right\}\\&\le\frac{\rho_2}{N}\sum_{k=0}^{\infty}c^2(k)E\|X(k)\|^2\le\frac{\rho_2 q_x}{N}\sum_{k=0}^{\infty}c^2(k),\end{aligned}\quad (43)$$

where the second "=" is by the definition of $\tilde{L}_{\mathcal{G}(k)}$ and $\{\mathcal{G}(k),k\ge0\}\in\Gamma_1$, the first "≤" is by Cr-inequality and the second is by (41).

For the second term, noting that $\{\mathbf{1}_N^T \sum_{j=0}^{n} c(k)D(k)Y(k)\xi(k), \mathcal{F}_{\xi,\mathcal{A}}(n), n \geq 0\}$ is a martingale, direct calculations gives

$$\begin{aligned}
&\lim_{n\to\infty} E\left[\frac{1}{N}\mathbf{1}_N^T \sum_{k=0}^{n-1} c(k)D(k)Y(k)\xi(k)\right]^2 \\
&= \frac{1}{N^2}\lim_{n\to\infty} E\left[\sum_{k=0}^{n-1}(\mathbf{1}_N^T c(k)D(k)Y(k)\xi(k))^2\right] \\
&\leq \frac{1}{N^2}\lim_{n\to\infty}\sum_{k=0}^{n-1} c^2(k)E\Bigg[\sum_{1\leq i,j\leq N}\xi_{ji}(k)a_{ij}(k) \\
&\quad \times(\sigma_{ji}(x_j(k)-x_i(k))+b_{ji})\Bigg]^2. \qquad (44)
\end{aligned}$$

Then by Cr-inequality, Assumptions **(A2)**, **(A5)** and Lemma A.1, we have

$$\begin{aligned}
&\lim_{n\to\infty} E\Big[\frac{1}{N}\mathbf{1}_N^T \sum_{k=0}^{n-1} c(k)D(k)Y(k)\xi(k)\Big]^2 \\
&\leq \frac{1}{N^2}\sum_{k=0}^{\infty}\Bigg\{c^2(k)\sum_{(i,j)\in\mathcal{E}_{\mathcal{G}(k)}} E\Big[|\mathcal{E}_{\mathcal{G}(k)}|\xi_{ji}^2(k)a_{ij}^2(k) \\
&\quad \times(\sigma_{ji}(x_j(k)-x_i(k))+b_{ji})^2\Big]\Bigg\} \\
&\leq \frac{2}{N^2}\sum_{k=0}^{\infty}\Bigg\{c^2(k)\sum_{(i,j)\in\mathcal{E}_{\mathcal{G}(k)}} E\Big[|\mathcal{E}_{\mathcal{G}(k)}|\xi_{ji}^2(k)a_{ij}^2(k) \\
&\quad \times(\sigma_{ji}^2(x_j(k)-x_i(k))^2+b_{ji}^2)\Big]\Bigg\} \\
&\leq \frac{2\beta b^2\rho_1}{N^2}\sum_{k=0}^{\infty}c^2(k) \\
&\quad +\frac{4\beta\sigma^2\rho_1}{N^2}\sum_{k=0}^{\infty}E[V(k)]c^2(k), \qquad (45)
\end{aligned}$$

where the first "$\leq$" is by Cr-inequality, and the last '$\leq$' is by Assumptions **(A2)**, **(A5)** and Lemma A.1. This together with (42) and (43) gives (5). □

Remark 10: The constant $\widetilde{c}$ in (5) and (6) can be replaced by $q_v c$ from the estimation (40). This removes the term $E[V(k)]$ in $\widetilde{c}$, however, makes the upper bound of the mean square steady-state error more conservative.

Remark 11: Lemma 1 plays important roles in the proof of Theorem 1.

- In [34], the network is assumed to be a fixed, balanced and strongly connected digraph. Then the property of the Laplacian matrix of a connected graph was directly used to the first-order difference inequality of the Lyapunov energy function. For the case with time-varying random graphs of this paper, the network graph is neither connected nor balanced instantaneously, and thus the method of [34] is not applicable. We further develop the stochastic Lyapunov method for the case with a sequence of random graphs and compound noises. By Lemma 1, the high-order difference inequality (23), where $E[V(i)]$ $i = mh+1,\ldots,(m+1)h-1$ are involved, is transformed into the h-step ahead first-order difference inequality (24).
- The system centroid equation (30) is different from those in [34] and [25]. First, the term $\frac{1}{N}\mathbf{1}^T\sum_{k=0}^{n-1}c(k)D(k)Y(k)\xi(k)$ induced by the noises is coupled with the state and the random graph sequence. Second, there is an additional term $\frac{1}{N}\mathbf{1}^T\sum_{k=0}^{n-1}c(k)\tilde{L}_{\mathcal{G}(k)} X(k)$ induced by the instantaneous unbalance of the network graph. By Lemma 1, we prove that the sequences $\{\mathbf{1}^T\sum_{k=0}^{n-1}c(k)D(k)Y(k)\xi(k), k \geq 0\}$ and $\{\mathbf{1}^T\sum_{k=0}^{n-1}c(k)\tilde{L}_{\mathcal{G}(k)} X(k), k \geq 0\}$ are both square integrable martingales.

Remark 12: Here, Assumption **(A3)** is a standard assumption on the step size in stochastic approximation. In practice, different from distributed averaging aiming at estimating the average of initial values, if the quantity to be estimated changes over time, then non-vanishing step size is often used. If the step size $c(k)$ is a sufficiently small constant, then from (31), one may see that the centroid of the system will diverge due to the additive noises and thus the mean square and almost average consensus will not be achieved.

V. Special Cases

In this section, we consider two special classes of random graph sequences: (i) $\{\mathcal{G}(k), k \geq 0\}$ is a Markov chain with countable state space; (ii) $\{\mathcal{G}(k), k \geq 0\}$ is an independent process with uncountable state space. By the stochastic Lyapunov method based on random graph sequences, we obtain sufficient conditions for mean square and almost sure average consensus. For these two special cases, Condition (b.1) of Theorem 1 becomes more intuitive and Condition (b.2) is weakened.

A. Markovian Switching Graph Sequence

Definition 2 [45]: A Markov chain on a countable state space $\mathcal{S}$ with a stationary distribution π, and transition probability function $\mathbb{P}(x,\cdot)$ is called uniformly ergodic, if there exist positive constants $r > 1$ and R such that for all $x \in \mathcal{S}$,

$$\|\mathbb{P}^n(x,\cdot)-\pi\|_1 \leq Rr^{-n}.$$

Here, $\|\mathbb{P}^n(x,\cdot)-\pi\|_1 = \sum_{y\in\mathcal{S}}|\mathbb{P}^n(x,y)-\pi(y)|$.

Denote $S_1 = \{A_j, j = 1,2,\ldots\}$, which is a countable set of generalized weighted adjacency matrices and denote the associated generalized Laplacian matrix of A_j by L_j. Let $\hat{L}_j = \frac{L_j+L_j^T}{2}$. In this subsection, we consider the class of random graph sequences defined by Γ_2 below, each element of which is a homogeneous and uniformly ergodic Markov chain with countable states and unique stationary distribution, i.e.

$$\begin{aligned}
\Gamma_2 = \Big\{&\{\mathcal{G}(k), k\geq 0\}|\{\mathcal{A}_{\mathcal{G}(k)},\ k\geq 0\}\subseteq S_1, \\
&\text{and is a homogeneous and uniformly ergodic} \\
&\text{Markov chain with unique stationary distribution}\pi; \\
&E[\mathcal{A}_{\mathcal{G}(k)}|\mathcal{A}_{\mathcal{G}(k-1)}]\succeq O_{N\times N},\ \text{a.s.}, \\
&\text{and the associated digraph of } E[\mathcal{A}_{\mathcal{G}(k)}|\mathcal{A}_{\mathcal{G}(k-1)}] \\
&\text{is balanced a.s.},\ k\geq 0.\Big\}.
\end{aligned}$$

Here, $\pi = [\pi_1, \pi_2, \ldots]^T$, $\pi_j \geq 0$, $\sum_{j=1}^{\infty} \pi_j = 1$, where π_j denotes $\pi(A_j)$.

We have the following theorem.

Theorem 2: For the system (1)-(2) and the associated random graph sequence $\{\mathcal{G}(k), k \geq 0\} \in \Gamma_2$, assume that

(i) Assumptions **(A1)-(A5)** hold;

(ii) the associated graph of the Laplacian matrix $\sum_{j=1}^{\infty} \pi_j L_j$ contains a spanning tree;

(iii) $\sup_{j\geq 1} \|\hat{L}_j\| < \infty$.

Then the system (1)-(2) achieves mean square and almost sure average consensus.

Proof: Since $\{\mathcal{A}_{\mathcal{G}(k)}, k \geq 0\}$ is a Markov chain, by the Markov property, we know that $E[\mathcal{A}_{\mathcal{G}(k)}|\mathcal{F}_{\mathcal{A}}(k-1)] = E[\mathcal{A}_{\mathcal{G}(k)}|\mathcal{A}_{\mathcal{G}(k-1)}]$. Thus, $\{\mathcal{G}(k), k \geq 0\} \in \Gamma_1$.

By the one-to-one correspondence among $\mathcal{A}_{\mathcal{G}(k)}$, $L_{\mathcal{G}(k)}$ and $\hat{L}_{\mathcal{G}(k)}$, we know that $\{L_{\mathcal{G}(k)}, k \geq 0\}$ and $\{\hat{L}_{\mathcal{G}(k)}, k \geq 0\}$ are both homogeneous and uniformly ergodic Markov chains with the unique stationary distribution π, whose state spaces are $S_2 = \{L_1, L_2, L_3, \ldots\}$ and $S_3 = \{\hat{L}_1, \hat{L}_2, \hat{L}_3, \ldots\}$, respectively. From (4), we know that

$$\begin{aligned}\lambda_{mh}^h &= \lambda_2 \left\{ \sum_{i=mh}^{mh+h-1} E[\hat{L}_{\mathcal{G}(i)}|\hat{L}_{\mathcal{G}(mh-1)} = \hat{L}_0] \right\} \\ &= \lambda_2 \left\{ \sum_{i=1}^{h} \sum_{j=1}^{\infty} \hat{L}_j \mathbb{P}^i(\hat{L}_0, \hat{L}_j) \right\}, \ \forall\ \hat{L}_0 \in S_3, \\ &\quad \forall\ m \geq 0,\ h \geq 1. \end{aligned} \tag{46}$$

Noting the uniform ergodicity of $\{\hat{L}_{\mathcal{G}(k)}, k \geq 0\}$ and the uniqueness of the stationary distribution π, by Condition (iii), we have

$$\begin{aligned}&\left\| \frac{\sum_{i=1}^{h} \sum_{j=1}^{\infty} \hat{L}_j \mathbb{P}^i(\hat{L}_0, \hat{L}_j)}{h} - \sum_{j=1}^{\infty} \pi_j \hat{L}_j \right\| \\ &= \left\| \frac{\sum_{i=1}^{h} \sum_{j=1}^{\infty} (\hat{L}_j \mathbb{P}^i(\hat{L}_0, \hat{L}_j) - \pi_j \hat{L}_j)}{h} \right\| \\ &= \left\| \frac{\sum_{i=1}^{h} \sum_{j=1}^{\infty} \hat{L}_j (\mathbb{P}^i(\hat{L}_0, \hat{L}_j) - \pi_j)}{h} \right\| \\ &\leq \sup_j \|\hat{L}_j\| \frac{\sum_{i=1}^{h} R r^{-i}}{h} \to 0, \quad h \to \infty.\end{aligned}$$

Furthermore, by the definition of uniform convergence, we know that

$$\frac{1}{h}\left[\sum_{i=mh}^{mh+h-1} E[\hat{L}_{\mathcal{G}(i)}|\hat{L}_{\mathcal{G}(mh-1)}]\right] \text{ converges to} \sum_{j=1}^{\infty} \pi_j \hat{L}_j \text{ a.s.,}$$

uniformly with respect to m, as $h \to \infty$. Denote $\alpha = \lambda_2(\sum_{j=1}^{\infty} \pi_j \hat{L}_j)$. By Condition (ii), it follows that $\alpha > 0$. Since the function $\lambda_2(\cdot)$, whose arguments are matrices, is continuous, we know that for the given $\frac{\alpha}{2}$, there exists a constant $\delta > 0$ such that for any given Laplacian matrix L, $|\lambda_2(L) - \lambda_2(\sum_{j=1}^{\infty} \pi_j \hat{L}_j)| \leq \frac{\alpha}{2}$, provided $\|L - \sum_{j=1}^{\infty} \pi_j \hat{L}_j\| \leq \delta$. Since the convergence is uniform, there exists a positive integer h_0 such that $\|\frac{1}{h}[\sum_{i=mh}^{mh+h-1} E[\hat{L}_{\mathcal{G}(i)}|\hat{L}_{\mathcal{G}(mh-1)}]] - \sum_{j=1}^{\infty} \pi_j \hat{L}_j\| \leq \delta$, $h \geq h_0$, a.s., which leads to $|\lambda_2(\frac{1}{h}[\sum_{i=mh}^{mh+h-1} E[\hat{L}_{\mathcal{G}(i)}|\hat{L}_{\mathcal{G}(mh-1)}]]) - \lambda_2(\sum_{j=1}^{\infty} \pi_j \hat{L}_j)| \leq \frac{\alpha}{2}$, $h \geq h_0$, a.s. Thus,

$$\lambda_2\left(\frac{1}{h}\left[\sum_{i=mh}^{mh+h-1} E[\hat{L}_{\mathcal{G}(i)}|\hat{L}_{\mathcal{G}(mh-1)}]\right]\right) \geq \frac{\alpha}{2} > 0, \text{ a.s.}$$

Then, by (46), we have $\lambda_{mh}^h \geq \frac{h\alpha}{2} > 0$, $h \geq h_0$ a.s. Thus, Condition (b.1) of Theorem 1 holds. Then, by Condition (iii), we know that Condition (b.2) of Theorem 1 holds. Finally, by Theorem 1, we get the conclusion of the theorem. □

B. Independent Graph Sequence

Consider the independent graph sequence

$$\begin{aligned}\Gamma_3 = \Big\{ &\{\mathcal{G}(k), k \geq 0\} | \{\mathcal{G}(k), k \geq 0\} \\ &\text{is an independent process, } E[\mathcal{A}_{\mathcal{G}(k)}] \succeq O_{N\times N}, \text{ a.s.} \\ &\text{and the associated digraph of } E[\mathcal{A}_{\mathcal{G}(k)}] \\ &\text{is balanced a.s., } k \geq 0 \Big\}.\end{aligned}$$

We have the following theorem.

Theorem 3: For the system (1)-(2) and the associated random graph sequence $\{\mathcal{G}(k), k \geq 0\} \in \Gamma_3$, assume that

(i) Assumptions **(A1)-(A5)** hold;

(ii) there exists a positive integer h such that

$$\inf_{m\geq 0} \left\{ \lambda_2 \left[\sum_{i=mh}^{(m+1)h-1} E[\hat{L}_{\mathcal{G}(i)}] \right] \right\} > 0;$$

(iii) $\sup_{k\geq 0} E\left[\|L_{\mathcal{G}(k)}\|^2\right] < \infty$.

Then the system (1)-(2) achieves mean square and almost sure average consensus.

Proof: From $\mathcal{G}(k) \in \Gamma_3$, we know that $\mathcal{G}(k) \in \Gamma_1$, and $E[\hat{L}_{\mathcal{G}(k)}]$ is positive semi-definite. By the independence of $\{\mathcal{G}(k), k \geq 0\}$, we have $E[\mathcal{A}_{\mathcal{G}(k)}|\mathcal{F}_{\mathcal{A}}(k-1)] = E[\mathcal{A}_{\mathcal{G}(k)}]$, $E[L_{\mathcal{G}(k)}|\mathcal{F}_{\mathcal{A}}(k-1)] = E[L_{\mathcal{G}(k)}]$, which together with Assumption **(A5)** gives

$$\begin{aligned}&E\left[\delta^T(k) \frac{L_{\mathcal{G}(k)}^T P_N^T + P_N L_{\mathcal{G}(k)}}{2} \delta(k)\right] \\ &= E\left[\delta^T(k) E\left[\frac{L_{\mathcal{G}(k)}^T P_N^T + P_N L_{\mathcal{G}(k)}}{2}\middle| \mathcal{F}_{\xi,\mathcal{A}}(k-1)\right] \delta(k)\right] \\ &= E\left[\delta^T(k) \frac{E[L_{\mathcal{G}(k)}^T] + E[L_{\mathcal{G}(k)}]}{2} \delta(k)\right] \\ &= E\left[\delta^T(k) E[\hat{L}_{\mathcal{G}(k)}] \delta(k)\right] \geq 0.\end{aligned}$$

Then, similar to the proof of **Step** 1 of Theorem 1, we get that $E[V(k)]$ is bounded. Denote $\sup_{k\geq 0} \left[E[\|L_{\mathcal{G}(k)}\|^2]\right]^{\frac{1}{2}}$ by ρ_4. Since $L_{\mathcal{G}(i)}$ is independent of $L_{\mathcal{G}(j)}$, $i \neq j$, we do not have to use the conditional Hölder inequality as in (19). Here, by the conditional Lyapunov inequality and Condition (iii), we have $\sup_{k\geq 0} E[\|L_{\mathcal{G}(k)}\|] \leq \sup_{k\geq 0}\{E[\|L_{\mathcal{G}(k)}\|^2]\}^{\frac{1}{2}} \leq \rho_4$. Then, similar to (19), we obtain

$$\begin{aligned}
&E\Big[\Big\|\Phi^T((m+1)h, mh)\Phi((m+1)h, mh)\\
&\qquad - I_N + \sum_{i=mh}^{(m+1)h-1} c(i)(P_N L_{\mathcal{G}(i)} + L^T_{\mathcal{G}(i)} P_N^T)\Big\|\Big]\\
&\leq \Big(C_1 \sum_{i=2}^{2h} M_{2h}^i \rho_4^i\Big) c^2((m+1)h)\\
&= C_1[(1+\rho_4)^{2h} - 1 - 2h\rho_4] c^2((m+1)h).
\end{aligned}$$

Also, by the independence of $\{\mathcal{G}(k), k \geq 0\}$ and Condition (ii), similarly to (20), we have

$$\begin{aligned}
&E\Bigg[\delta^T(mh) \sum_{i=mh}^{(m+1)h-1} c(i)\Big[P_N L_{\mathcal{G}(i)}\\
&\qquad + L^T_{\mathcal{G}(i)} P_N^T\Big]\delta(mh)\Bigg]\\
&= 2E\left[\delta^T(mh)\left(\sum_{i=mh}^{(m+1)h-1} c(i) E[\hat{L}_{\mathcal{G}(i)}]\right)\delta(mh)\right]\\
&\geq 2c((m+1)h) \inf_{m\geq 0}\left\{\lambda_2\left[\sum_{i=mh}^{(m+1)h-1} E[\hat{L}_{\mathcal{G}(i)}]\right]\right\}\\
&\qquad \times E[V(mh)].
\end{aligned}$$

Then, similarly to the proof of **Step** 2 of Theorem 1, we get $E[V(k)] \to 0, k \to \infty$.

By the independence of $\mathcal{G}(k)$, $k \geq 0$ and Assumption **(A5)**, we know that the adaptive sequences $\{\mathbf{1}_N^T \sum_{j=0}^n c(k) D_{\mathcal{G}(k)} Y(k)\xi(k),\ \mathcal{F}_{\xi,\mathcal{A}}(n),\ n \geq 0\}$ and $\{\mathbf{1}_N^T \sum_{k=0}^n c(k)\tilde{L}_{\mathcal{G}(k)} X(k),\ \mathcal{F}_{\xi,\mathcal{A}}(n), n \geq 0\}$ are both martingale sequences. Then, similar to **Steps** 1, 2 and 3 of Theorem 1, we get the conclusion of the theorem. □

Remark 13: In Theorem 3, the associated digraph of $E[\mathcal{A}_{\mathcal{G}(k)}]$, i.e., the mean graph at each time instant, is balanced, so the symmetrized mean graph is undirected. Condition (ii) of Theorem 3 means that the symmetrized mean graphs are jointly-connected (the mean graph has a spanning tree) over consecutive fixed-length time intervals and the average algebraic connectivity is uniformly positive bounded away from zero.

The gossip algorithm [16] is a special distributed averaging algorithm with a sequence of i.i.d network graphs. For distributed averaging algorithms with a sequence of i.i.d network graphs, the mean square steady-state error can be estimated more precisely with sufficiently small initial algorithm gains. Moreover, the almost sure convergence rate of the n-step mean consensus error can be estimated.

Consider the i.i.d graph sequence

$$\begin{aligned}
\Gamma_4 = \Big\{&\{\mathcal{G}(k), k \geq 0\} | \{\mathcal{G}(k), k \geq 0\}\\
&\text{is an i.i.d process with } E[\mathcal{A}_{\mathcal{G}(0)}] \succeq O_{N\times N}, \text{and}\\
&\text{the associated digraph of } E[\mathcal{A}_{\mathcal{G}(0)}] \text{ is balanced.}\Big\}.
\end{aligned}$$

Theorem 4: For the system (1)-(2) and the associated random graph sequence $\mathcal{G}(k) \in \Gamma_4$, assume that

(i) Assumptions **(A1)**-**(A5)** hold;

(ii) the associated digraph of the Laplacian matrix $E[L_{\mathcal{G}(0)}]$ has a spanning tree;

(iii) $E\left[\|L_{\mathcal{G}(0)}\|^2\right] < \infty$.

Then, all states $x_i(k)$, $i \in \mathcal{V}$, converge to a common random variable x^*, in mean square and almost surely, with $E(x^*) = \frac{1}{N}\sum_{j=1}^N x_j(0)$ and

$$Var(x^*) \leq \frac{4c\beta b^2 \bar{\rho}_1}{N^2} + \frac{8\widetilde{c}\beta\sigma^2\bar{\rho}_1}{N^2} + \frac{2c\bar{\rho}_2\ q_x}{N},$$

where b, σ, c, $\widetilde{c}$, q_x are constants defined in (5) and

$$\begin{aligned}
\bar{\rho}_1 &= E\left[|\mathcal{E}_{\mathcal{G}(0)}| \max_{1\leq i,j\leq N} a_{ij}^2(0)\right],\\
\bar{\rho}_2 &= \max_{1\leq i\leq N} E\left[\left(\sum_{j=1}^N a_{ij}(0) - \sum_{j=1}^N a_{ji}(0)\right)^2\right].
\end{aligned}$$

The convergence rate of n-step mean consensus error is estimated by

$$\frac{1}{n}\sum_{k=0}^n \|\delta(k)\| = o\left(\frac{1}{\sqrt{c(n)n}}\right) \text{ a.s.} \tag{47}$$

Furthermore, if the initial algorithm gain is so small that

$$c(0) < \frac{2\lambda_2\left(E[\hat{L}_{\mathcal{G}(0)}]\right)}{E\left[\|L_{\mathcal{G}(0)}\|^2\right] + 4\sigma^2\beta\bar{\rho}_1}, \tag{48}$$

then

$$\widetilde{c} \leq \frac{c(0)E[V(0)] + 2b^2\beta\bar{\rho}_1\sum_{k=0}^\infty c^3(k)}{2\lambda_2\left(E[\hat{L}_{\mathcal{G}(0)}]\right) - \left(E\left[\|L_{\mathcal{G}(0)}\|^2\right] + 4\sigma^2\beta\bar{\rho}_1\right)c(0)}. \tag{49}$$

Proof: It is obvious that $\Gamma_4 \subseteq \Gamma_3$, so $\mathcal{G}(k) \in \Gamma_3$. By Condition (ii) and $\mathcal{G}(k) \in \Gamma_4$, we know that $\lambda_2\left(E[\hat{L}_{\mathcal{G}(0)}]\right) > 0$ and Condition (ii) of Theorem 3 holds with $h = 1$. Obviously, Condition (iii) together with $\mathcal{G}(k) \in \Gamma_4$ implies Condition (iii) of Theorem 3. Then, by Theorem 3, the closed-loop system achieves mean square and almost sure average consensus. From (8), we have

$$\begin{aligned}
&E[V(k+1)|\mathcal{F}_{\xi,\mathcal{A}}(k)]\\
&\quad \leq V(k) - 2c(k)\lambda_2(E[\hat{L}_{\mathcal{G}(0)}])V(k)\\
&\qquad + E[\|L_{\mathcal{G}(0)}\|^2]c^2(k)V(k)\\
&\qquad + 4\sigma^2\beta\bar{\rho}_1 c^2(k)V(k) + 2b^2\beta\bar{\rho}_1 c^2(k) \text{ a.s.},
\end{aligned} \tag{50}$$

which together with $\lambda_2\left(E[\hat{L}_{\mathcal{G}(0)}]\right) > 0$ and Lemma A.2 leads to

$$\sum_{k=0}^\infty c(k)V(k) < \infty \text{ a.s.} \tag{51}$$

Then, by Assumption **(A4)** and Kronecker lemma ([42]), we have

$$\lim_{n\to\infty} c(n)\sum_{k=0}^{n} V(k) = 0 \text{ a.s.},$$

which together with Cauchy inequality $\sum_{k=0}^{n}\|\delta(k)\| \leq \sqrt{n}\sqrt{\sum_{k=0}^{n} V(k)}$ results in (47).

From (50), we have

$$\begin{aligned} &E[V(k+1)] \\ &\quad\leq E[V(k)] - 2c(k)\lambda_2(E[\hat{L}_{\mathcal{G}(0)}])E[V(k)] \\ &\qquad + E[\|L_{\mathcal{G}(0)}\|^2]c^2(k)E[V(k)] \\ &\qquad + 4\sigma^2\beta\bar{\rho}_1 c^2(k)E[V(k)] + 2b^2\beta\bar{\rho}_1 c^2(k). \end{aligned}$$

Then, by Assumption **(A4)**, we have

$$\begin{aligned} &(2\lambda_2(E[\hat{L}_{\mathcal{G}(0)}]) - E[\|L_{\mathcal{G}(0)}\|^2]c(0) \\ &\quad - 4\sigma^2\beta\bar{\rho}_1\ c(0))c^2(k)E[V(k)] \\ &\quad\leq c(k)E[V(k)] - c(k+1)E[V(k+1)] + 2b^2\beta\bar{\rho}_1 c^3(k). \end{aligned}$$

Taking summation on both sides of the above inequality from $k=0$ to $k=n$ gives

$$\begin{aligned} &(2\lambda_2(E[\hat{L}_{\mathcal{G}(0)}]) - E[\|L_{\mathcal{G}(0)}\|^2]c(0) \\ &\qquad - 4\sigma^2\beta\bar{\rho}_1\ c(0))\sum_{k=0}^{n} c^2(k)E[V(k)] \\ &\quad\leq c(0)E[V(0)] - c(n+1)E[V(n+1)] \\ &\qquad + 2b^2\beta\bar{\rho}_1\sum_{k=0}^{n} c^3(k). \end{aligned}$$

Then, by (48) and let $n\to\infty$, we have (49). □

Remark 14: Theorem 4 shows that if the step size $c(k) = \Theta\left(\frac{\ln^{\beta}(k)}{k^{\gamma}}\right)$, $\gamma\in(0.5,1]$ and $\beta\geq -1$, then the n-step mean consensus error is $o\left(\frac{1}{\sqrt{n^{1-\gamma}\ln^{\beta} n}}\right)$ almost surely under the i.i.d graph sequence. Here, Theorem 4 only gives a rough estimate of the convergence rate of n-step mean consensus error in the sense that if $c(k)=\Theta(1/k)$, then the trivial estimate $o(1)$ is given. To get the exact convergence rate of the consensus error in probability one is challenging. Some preliminary results have been presented in [47] for the case with only additive noises and fixed network graph, especially, if $c(k)=\Theta(1/k)$, it was shown that $\|\delta(k)\| = O\left(\sqrt{\frac{\ln\ln k}{k}}\right)$ a.s., provided the algebraic connectivity of the network graph is sufficiently large [47].

VI. Numerical Examples

We consider a simple random multi-agent network with three nodes, whose states are $x_1(k)$, $x_2(k)$ and $x_3(k)$, $k\geq 0$, respectively. The initial values are given by $x_1(0)=9$, $x_2(0)=7$, $x_3(0)=6$. At each time instant, the network graph has six random edges. Here, the noise intensity function $f_{ji}(x_j(k)-x_i(k)) = \sigma|x_i(k)-x_j(k)|+b$, $i,j=1,2,3$. Take $c(k)=\frac{1}{k}$. Then by the algorithm (1)-(2), the state updating rule is given by

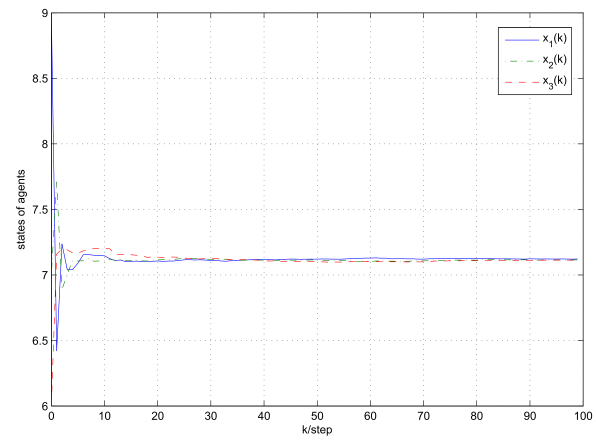

Fig. 1. Trajectories of states when $h=1$.

$$\begin{aligned} x_1(k+1) &= x_1(k) + \frac{1}{k}\sum_{i=2,3} a_{1i}(k)\Big(x_i(k)-x_1(k) \\ &\qquad + \sigma|x_i(k)-x_1(k)|\xi_{i1}(k) + b\xi_{i1}(k)\Big), \\ x_2(k+1) &= x_2(k) + \frac{1}{k}\sum_{i=1,3} a_{2i}(k)\Big(x_i(k)-x_2(k) \\ &\qquad + \sigma|x_i(k)-x_2(k)|\xi_{i2}(k) + b\xi_{i2}(k)\Big), \\ x_3(k+1) &= x_3(k) + \frac{1}{k}\sum_{i=1,2} a_{3i}(k)\Big(x_i(k)-x_3(k) \\ &\qquad + \sigma|x_i(k)-x_3(k)|\xi_{i3}(k) + b\xi_{i3}(k)\Big). \end{aligned}$$

The random weights $\{a_{ij}(k), i,j=1,2,3, k\geq 0\}$ are selected by the following rules. For some positive integer h, when $k=mh$, $m\geq 0$, the random weights are uniformly distributed on the interval $[0,1]$; when $k\neq mh$, $m\geq 0$, the random weights are uniformly distributed on $[-0.5,0.5]$. So, here, the random weights may be negative at some time instants. Here, $\{a_{ij}(k), i,j=1,2,3, k\geq 0\}$ are spatially and temporally independent. Then the conditional graph degenerates to the mean graph. It can be verified that when $k=mh$, $m\geq 0$, the mean graph is balanced and connected and when $k\neq mh$, $m\geq 0$, the mean graphs are empty. Thus, the mean graphs are jointly connected on the time interval $[mh,(m+1)h)$. Assume that the communication noises $\{\xi_{ji}(k), i,j=1,2,3, k\geq 0\}$ are independent standard normally distributed random variables and independent of the random graphs. Let $\sigma=0.1$ and $b=0.1$. By Theorem 3, the states of these nodes would asymptotically converge to a random variable whose mathematical expectation is the average of initial values.

Now we demonstrate that the states of the agents agree asymptotically. Take $h=1,2,3$, and the states of agents are shown in Figures 1, 2 and 3, respectively. It is shown that the agreement is asymptotically achieved and smaller h (the length of the intervals over which the network graphs are jointly connected) gives faster convergence. Take $h=1$ and

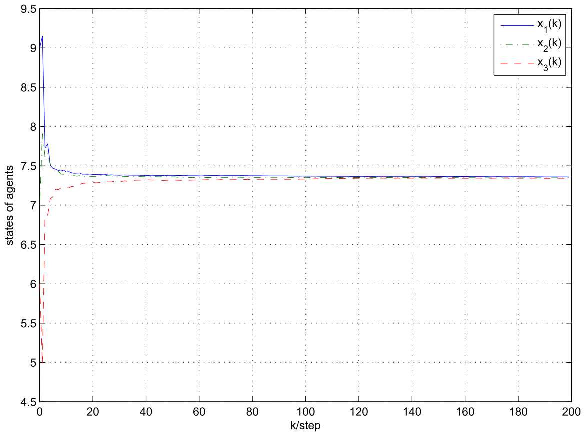

Fig. 2. Trajectories of states when $h = 2$.

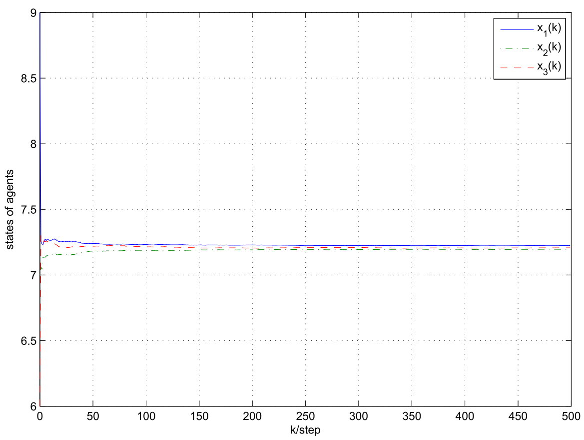

Fig. 3. Trajectories of states when $h = 3$.

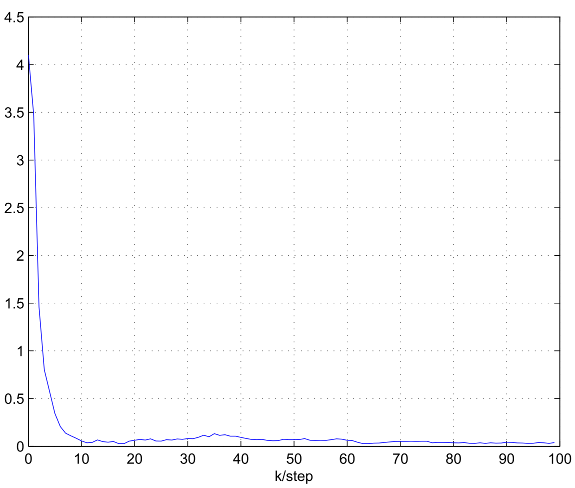

Fig. 4. Trajectories of $\|\delta(k)\|\sqrt{\frac{k+1}{\ln(1+\ln(k+2))}}$ when $h = 1$.

the trajectory of $\|\delta(k)\|\sqrt{\frac{k+1}{\ln(1+\ln(k+2))}}$ is shown in 4, from which one may see that the convergence rate is no slower than $O\left(\frac{\ln\ln k}{k}\right)$.

VII. Conclusion

We have considered discrete-time stochastic approximation type distributed averaging algorithms with random measurement noises and time-varying random graphs. Compared with the existing literature, our model is more widely applicable in the sense that i) the measurement covers both additive and multiplicative noises; ii) the network graphs and noises are not required to be spatially and temporally independent; iii) the edge weights of network graphs are not necessarily nonnegative with probability one. By further developing stochastic Lyapunov method and the combination of algebraic graph theory and martingale convergence theory, sufficient conditions have been given to achieve mean square and almost sure average consensus. It has been shown that all states of agents converge to a common variable in mean square and almost surely if the graph sequence is *conditionally balanced and uniformly conditionally jointly connected*. The mathematical expectation of the common random variable is just the average of initial values. Moreover, an upper bound of the mean square steady-state error has been given in relation to the edge weights, the time-varying algorithm gain, the number of agents, the agents' initial values, the second-order moment and the intensity coefficients of noises. Especially, if the measurement noises are both spatially and temporally independent, then the mean square steady-state error vanishes as the number of nodes increases to infinity under mild conditions on the network graphs.

Convergence rate is an important performance for distributed averaging algorithms. Different from the fixed-gain algorithms for noise-free cases [19], [30], [48], here, the non-zero off-diagonal elements of the closed-loop state matrix are not uniformly bounded away from zero, which results in much more difficulties to get the exact stochastic convergence rates of the algorithm. For the case with a sequence of i.i.d random graphs, we have given a rough estimate for the n-step mean consensus error with probability one. It is interesting to develop effective tools to give the exact stochastic convergence rates of our algorithms.

Appendix

In this paper, the following basic inequalities will be used. For the conditional Lyapunov inequality and the conditional Hölder inequality, the readers may be referred to Theorem 6.4 and its next paragraph in [49, Ch. 6].

Denote the probability space by $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_1$ be a sub σ-algebra of $\mathcal{F}$.

Conditional Lyapunov inequality. Let ξ be a random variable on $(\Omega, \mathcal{F}, P)$. Then

$$(E[|\xi|^s|\mathcal{F}_1])^{1/s} \le (E[|\xi|^t|\mathcal{F}_1])^{1/t} \quad \text{a.s.,} \ \ 0 < s < t.$$

Conditional Hölder inequality. 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 $E[|\xi|^p] < \infty$ and $E[|\eta|^q] < \infty$, then

$$E[|\xi\eta||\mathcal{F}_1] \le (E[|\xi|^p|\mathcal{F}_1])^{1/p}(E[|\eta|^q|\mathcal{F}_1])^{1/q} \ \text{a.s.}$$

If $\mathcal{F}_1$ is the trivial σ-algebra $\{\Omega, \Phi\}$, then the conditional Lyapunov inequality and conditional Hölder inequality degenerate to the usuanl Lyapunov inequality and Hölder inequality, respectively.

Cr-inequality. Let $a_i \ge 0$, $i = 1, 2, \ldots, n$. Then $\left(\sum_{i=1}^n a_i\right)^r \le n^{r-1}\sum_{i=1}^n a_i^r$, $r > 1$.

Lemma A.1: Let $\{Z_k, k \geq 0\}$ and $\{W_k, k \geq 0\}$ be mutually independent random vector sequences. Then $\sigma(Z_j, Z_{j+1}, \ldots)$ and $\sigma(W_j, W_{j+1}, \ldots)$ are conditionally independent given $\sigma(Z_0, \ldots, Z_{j-1}, W_0, \ldots, W_{j-1})$, $\forall\ j \geq 1$.

Proof: Denote $Z_{m\sim n} = \{Z_m = z_m, \ldots, Z_n = z_n\}$ and $Z_{m\sim\infty} = \{Z_m = z_m, Z_{m+1} = z_{m+1}, \ldots\}$ where z_k denotes the possible values of Z_k. By the definition of conditional probability, we have

$$\begin{aligned}\mathbb{P}\{Z_{j\sim\infty}, W_{j\sim\infty}|Z_{0\sim j-1}, W_{0\sim j-1}\}\\ = \mathbb{P}\{W_{j\sim\infty}|Z_{0\sim j-1}, W_{0\sim j-1}\}\\ \times \mathbb{P}\{Z_{j\sim\infty}|Z_{0\sim j-1}, W_{0\sim\infty}\}.\end{aligned} \quad \text{(A.1)}$$

Noting that $\sigma(Z_{0\sim\infty}) = \sigma(\sigma(Z_{j\sim\infty}) \cup \sigma(Z_{0\sim j-1}))$ and $\sigma(Z_{0\sim\infty})$ is independent of $\sigma(W_{0\sim\infty})$, by [42, Sec. 7.3, Corollary 3], we have $\mathbb{P}\{Z_{j\sim\infty}|Z_{0\sim j-1}, W_{0\sim\infty}\} = \mathbb{P}\{Z_{j\sim\infty}|Z_{0\sim j-1}\} = \mathbb{P}\{Z_{j\sim\infty}|Z_{0\sim j-1}, W_{0\sim j-1}\}$, which together with (A.1) gives $\mathbb{P}\{Z_{j\sim\infty}, W_{j\sim\infty}|Z_{0\sim j-1}, W_{0\sim j-1}\} = \mathbb{P}\{W_{j\sim\infty}|Z_{0\sim j-1}, W_{0\sim j-1}\}\mathbb{P}\{Z_{j\sim\infty}|Z_{0\sim j-1}, W_{0\sim j-1}\}$. By the definition of conditional independence, we get the conclusion. □

Lemma A.2: [46] Let $\{x(k), \mathcal{F}(k)\}$, $\{\alpha(k), \mathcal{F}(k)\}$, $\{\beta(k), \mathcal{F}(k)\}$ and $\{\gamma(k), \mathcal{F}(k)\}$ be nonnegative adaptive sequences satisfying

$$E(x(k+1)|\mathcal{F}(k)) \leq (1+\alpha(k))x(k) - \beta(k) + \gamma(k), k \geq 0\ a.s.,$$

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

Lemma A.3: [43] Let $\{u(k),\ k \geq 0\}$, $\{q(k),\ k \geq 0\}$ and $\{\alpha(k),\ k \geq 0\}$ be real sequences, where $0 < q(k) \leq 1$, $\alpha(k) \geq 0$, $k \geq 0$, $\sum_{k=0}^{\infty} q(k) = \infty$, $\frac{\alpha(k)}{q(k)} \to 0$, $k \to \infty$, and $u(k+1) \leq (1-q(k))u(k) + \alpha(k)$. Then $\limsup_{k\to\infty} u(k) \leq 0$. Especially, if $u(k) \geq 0$, $k \geq 0$, then $u(k) \to 0$, $k \to \infty$.

Lemma A.4: ([44]) Let $\{X(k), \mathcal{F}(k)\}$ be a martingale sequence satisfying $\sup_{k\geq 0} E[\|X(k)\|^2] < \infty$. Then $X(k)$ converges in mean square and almost surely.

Acknowledgement

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

References

[1] N. A. Lynch, *Distributed Algorithms*. San Matero, CA, USA: Morgan Kaufmann, 1996.

[2] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, 2004.

[3] R. Olfati-Saber and P. Jalalkamali, "Coupled distributed estimation and control for mobile sensor networks," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2609–2614, Oct. 2012.

[4] A. T. Kamal, J. A. Farrell, and A. K. Roy-Chowdhury, "Information weighted consensus filters and their application in distributed camera networks," *IEEE Trans. Autom. Control*, vol. 58, no. 12, pp. 3112–3125, Dec. 2013.

[5] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw.*, Apr. 2005, pp. 63–70.

[6] A. H. Sayed, "Adaptive networks," *Proc. IEEE*, vol. 102, no. 4, pp. 460–497, Apr. 2014.

[7] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. AC-31, no. 9, pp. 803–812, Sep. 1986.

[8] N. Amelina, A. Fradkov, Y. Jiang, and D. J. Vergados, "Approximate consensus in stochastic networks with application to load balancing," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1739–1752, Apr. 2015.

[9] N. Elia, "Remote stabilization over fading channels," *Syst. Control Lett.*, vol. 54, no. 3, pp. 237–249, 2005.

[10] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 157–166, Mar. 1982.

[11] M. Huang and J. H. Manton, "Coordination and consensus of networked agents with noisy measurements: Stochastic algorithms and asymptotic behavior," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 134–161, 2009.

[12] T. Li and J.-F. Zhang, "Mean square average-consensus under measurement noises and fixed topologies: Necessary and sufficient conditions," *Automatica*, vol. 45, no. 8, pp. 1929–1936, 2009.

[13] T. Li, F. Wu, and J.-F. Zhang, "Multi-agent consensus with relative-state-dependent measurement noises," *IEEE Trans. Autom. Control*, vol. 59, no. 9, pp. 2463–2468, Sep. 2014.

[14] Y.-H. Ni and X. Li, "Consensus seeking in multi-agent systems with multiplicative measurement noises," *Syst. Control Lett.*, vol. 62, no. 5, pp. 430–437, 2013.

[15] M. Porfiri and D. J. Stilwell, "Consensus seeking over random weighted directed graphs," *IEEE Trans. Autom. Control*, vol. 52, no. 9, pp. 1767–1773, Sep. 2007.

[16] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[17] S. Kar and J. M. F. Moura, "Sensor networks with random links: Topology design for distributed consensus," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3315–3326, Jul. 2008.

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

[19] D. Bajović, J. Xavier, J. M. F. Moura, and B. Sinopoli, "Consensus and products of random stochastic matrices: Exact rate for convergence in probability," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2557–2571, May 2013.

[20] A. Tahbaz-Salehi and A. Jadbabaie, "A necessary and sufficient condition for consensus over random networks," *IEEE Trans. Autom. Control*, vol. 53, no. 3, pp. 791–795, Apr. 2008.

[21] A. Tahbaz-Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 225–230, Jan. 2010.

[22] I. Matei, J. S. Baras, and C. Somarakis, "Convergence results for the linear consensus problem under Markovian random graphs," *SIAM J. Control Optim.*, vol. 51, no. 2, 1574–1591, 2013.

[23] B. Liu, W. Lu, and T. Chen, "Consensus in networks of multiagents with switching topologies modeled as adapted stochastic processes," *SIAM J. Control Optim.*, vol. 49, no. 1, pp. 227–253, 2011.

[24] B. Touri and A. Nedić, "Product of random stochastic matrices," *IEEE Trans. Autom. Control*, vol. 59, no. 2, pp. 437–448, Feb. 2014.

[25] T. Li and J.-F. Zhang, "Consensus conditions of multi-agent systems with time-varying topologies and stochastic communication noises," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2043–2057, Sep. 2010.

[26] R. Rajagopal and M. J. Wainwright, "Network-based consensus averaging with general noisy channels," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 373–385, Jan. 2011.

[27] S. Kar and J. M. F. Moura, "Distributed consensus algorithms in sensor networks with imperfect communication: Link failures and channel noise," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 355–369, Jan. 2009.

[28] M. Huang, S. Dey, G. N. Nair, and J. H. Manton, "Stochastic consensus over noisy networks with Markovian and arbitrary switches," *Automatica*, vol. 46, no. 10, pp. 1571–1583, 2010.

[29] T. C. Aysal and K. E. Barner, "Convergence of consensus models with stochastic disturbances," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 4101–4113, Aug. 2010.

[30] S. Petterson, B. Bamieh, and A. El Abbadi, "Convergence rates of distributed average consensus with stochastic link failures," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 880–892, Apr. 2010.

[31] J. Wang and N. Elia, "Distributed averaging under constraints on information exchange: Emergence of Lévy flights," *IEEE Trans. Autom. Control*, vol. 57, no. 10, pp. 2435–2449, Oct. 2012.

[32] J. Wang and N. Elia, "Mitigation of complex behavior over networked systems: Analysis of spatially invariant structures," *Automatica*, vol. 49, no. 6, pp. 1626–1638, 2013.

[33] Y. Long, S. Liu, and L. Xie, "Distributed consensus of discrete-time multi-agent systems with multiplicative noises," *Int. J. Robust Nonlinear Control*, vol. 25, no. 16, pp. 3113–3131, 2014.

[34] J. Wang and T. Li, "Sufficient conditions on distributed averaging with compound noises and fixed topologies," in *Proc. 28th Chin. Control Decis. Conf.*, Yinchuan, China, May 2016, pp. 832–837.

[35] M. Huang, "Stochastic approximation for consensus: A new approach via ergodic backward products," *IEEE Trans. Autom. Control*, vol. 57, no. 12, pp. 2994–3008, Dec. 2012.

[36] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[37] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[38] C. Altafini, "Consensus problems on networks with antagonistic interactions," *IEEE Trans. Autom. Control*, vol. 58, no. 4, pp. 935–946, Apr. 2013.

[39] 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.

[40] R. Diestel, *Graph Theory*, 3rd ed. Berlin, Germany: Springer-Verlag, 2006.

[41] B. Bollobás, *Modern Graph Theory*. New York, NY, USA: Springer-Verlag, 1998.

[42] Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, 3rd ed. New York, NY, USA: Springer-Verlag, 1997.

[43] B. T. Polyak, *Introduction to Optimization*. New York, NY, USA: Optimization Software, 1987.

[44] R. B. Ash, *Real Analysis and Probability*. New York, NY, USA: Academic, 1972.

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

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

[47] H. Tang and T. Li, "Convergence rates of discrete-time stochastic approximation consensus algorithms: Graph-related limit bounds," *Syst. Control Lett.*, vol. 112, pp. 9–17, Feb. 2018.

[48] A. Olshevsky and J. N. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 33–55, 2009.

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

Tao Li (M'09–SM'14) 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 (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2009. Since January 2017, he has been with East China Normal University, Shanghai, China, where now he is a Professor and Director of the Department of Intelligent Mathematical Sciences, School of Mathematical Sciences. Dr. Li's current research interests include stochastic systems, cyber-physical multi-agent systems and game theory. He received the 28th "Zhang Siying" (CCDC) Outstanding Youth Paper Award in 2016, the Best Paper Award of the 7th Asian Control Conference in 2009, and honourable mentioned as one of five finalists for Young Author Prize of the 17th IFAC Congress in 2008. He received the 2009 Singapore Millennium Foundation Research Fellowship and the 2010 Australian Endeavor Research Fellowship. He was entitled Dongfang Distinguished Professor by Shanghai Municipality in 2012, received the Excellent Young Scholar Fund from National Natural Science Foundation of China in 2015 and was elected to the Chang Jiang Scholars Program, Ministry of Education, China, Youth Scholar in 2018. He now serves as an Associate Editor of *SCIENCE CHINA Information Sciences*, *International Journal of System Control and Information Processing*, *Control and Decision*, and *Journal of Systems Science and Mathematical Sciences*. He is a member of IFAC Technical Committee on Networked Systems and a member of Technical Committee on Control and Decision of Cyber-Physical Systems, Chinese Association of Automation.

Jiexiang Wang was born in Jiangxi, China, in September 1991. He received his B.E. degree in Automation from East China Jiaotong University, Nanchang, China, in 2014, and is now working towards the Ph.D. degree in control theory and engineering in the School of Mechatronic Engineering and Automation, Shanghai University. His current research interests include stochastic systems and multi-agent systems. He received the 28th "Zhang Siying" (CCDC) Outstanding Youth Paper Award in 2016.