Systems & Control Letters 112 (2018) 9–17

Contents lists available at ScienceDirect

Systems & Control Letters

journal homepage: www.elsevier.com/locate/sysconle

Convergence rates of discrete-time stochastic approximation consensus algorithms: Graph-related limit bounds

Huaibin Tang [a,b], Tao Li [c,*]

[a] School of Microelectronics, Shandong University, Jinan 250100, China
[b] Key Laboratory of Systems and Control, Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[c] Shanghai Key Laboratory of Pure Mathematics and Mathematical Practice, Department of Mathematics, East China Normal University, Shanghai 200241, China

ARTICLE INFO

Article history:
Received 9 August 2016
Received in revised form 29 November 2017
Accepted 5 December 2017

Keywords:
Consensus
Sensor network
Martingale difference sequence
Stochastic approximation
Convergence rate

ABSTRACT

In this paper, we study the convergence rates of the discrete-time stochastic approximation consensus algorithms over sensor networks with communication noises under general digraphs. Basic results of stochastic analysis and algebraic graph theory are used to investigate the dynamics of the consensus error, and the mean square and sample path convergence rates of the consensus error are both given in terms of the graph and noise parameters. Especially, calculation methods to estimate the mean square limit bounds are presented under balanced digraphs, and sufficient conditions on the network topology and the step sizes are given to achieve the fast convergence rate. For the sample path limit bounds, estimation methods are also presented under undirected graphs.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Recently, consensus algorithms with stochastic disturbances in sensor networks have been widely investigated, including measurement noises, time delay, quantized data and random link failures [1–7]. For consensus problems with communication noises, the stochastic approximation (SA) consensus algorithm with decreasing step sizes is an effective method to attenuate the influence of noises. Applications of SA consensus algorithms and theoretical development in consensus problems were reported in [8–12], etc. Under fixed or time-varying topologies, the condition that the network contains or jointly contains a spanning tree to guarantee consensus is well understood. This problem has been systematically investigated by the tools of the SA theory [12,13], the quadratic Lyapunov functions [8,9], the algebraic theory [5], and the ergodicity backward product approach [10,11]. Different from the SA-type consensus algorithm with decreasing step size, Amelina et al. [14] proposed the consensus algorithm with a nonvanishing stepsize for nonlinear agent dynamics over noisy networks to achieve the approximate mean square consensus.

It is worth noting that the convergence rate of the consensus algorithm, which characterizes how fast consensus can be achieved, is an important issue from the perspective of practical applications. For the case with precise communication, the consensus error vanishes exponentially with the rate governed by the second smallest eigenvalue of the Laplacian matrix [15,16]. The problems how to characterize and optimize the convergence rate are extensively studied via the optimization of the weighted adjacency matrix [17], local node state prediction [18], and filtering techniques [19]. Recently, Olshevsky and Tsitsiklis [20] investigated the convergence time of consensus algorithm under time-varying undirected graphs and proposed a linear time average-consensus protocol under fixed undirected graphs [21]. For the SA consensus algorithm, the convergence rate problem has also attracted much attention. For the average-consensus problem under undirected graphs, Kar and Moura [22] showed that the mathematical expectation of the state vector sequence converges exponentially to the consensus value, and Dasarathan et al. [12] derived the asymptotic covariance matrix of the consensus error when the step size $a(t) = \Theta(t^{-1})$. For the case with balanced digraphs, Li and Zhang [8] obtained the sample path convergence rate of finite step mean consensus error. For the leader-following topology case, Xu et al. [13] showed that the sample path convergence rate of the consensus error is $o(a^{\delta_1}(t))$ if the step size $a(t)$ satisfies $\lim_{t\to\infty}(a(t)-a(t+1))/(a(t)a(t+1)) \geq 0$, and the mean square convergence rate of the consensus error is $o(a^{\delta_2}(t))$ if $a(t) = \Theta(t^{-\alpha})$ with $\alpha \in (0.5, 1]$, $\delta_1, \delta_2 \in (0, 1)$. Wang et al. [23,24] investigated the convergence rate in the sense of convergence in distribution for multi-scale consensus modeling

* Corresponding author.
E-mail addresses: tanghuaibin@sdu.edu.cn (H. Tang), tli@math.ecnu.edu.cn (T. Li).

https://doi.org/10.1016/j.sysconle.2017.12.002
0167-6911/© 2017 Elsevier B.V. All rights reserved.

with Markovian regime switching. Compared with the case with precise communication, the convergence of SA algorithm is not exponentially fast any more and is in more complex relation to the step size $a(t)$ and network parameters. This motivates us to evaluate the impacts of the step size and network parameters on the algorithm, which is useful for the designers to improve the convergence rate.

In this paper, we consider the discrete-time SA consensus algorithm under general digraphs corrupted by martingale difference sequence communication noises. Different from [9–11] which focused on the consensus conditions of the SA consensus algorithm, the main goal of this paper is focused on the convergence rate analysis in relation to the step size and network graph parameters. For a class of typical step sizes, we apply basic results of stochastic analysis and algebraic graph theory to investigate the consensus error dynamic equation. Compared with the continuous-time SA consensus algorithm in [25], there is no Itô formula as the effective tool and we develop more technical tools of inequality theory to handle with the closed-form of the consensus error. Our contribution mainly includes the following three aspects.

- For the case of fixed topologies, we show that if the step size $a(t) = \Theta(t^{-\gamma}), \gamma \in (0.5, 1)$, then the mean square convergence rate of the consensus error is $O(t^{-\gamma})$; especially, for the case with balanced digraphs, the convergence rate is exactly $\Theta(t^{-\gamma})$. Furthermore, both upper and lower limit bounds of $t^{\gamma}E(\|\delta(t)\|^2)$ are explicitly given in terms of the noise intensity, the number of nodes, the smallest and the largest nonzero eigenvalues of the Laplacian matrix of the symmetrized graph.
- If $a(t) = \Theta(t^{-1})$, intuitively, the mean square convergence rate of the consensus error might be $O(t^{-1})$ and higher than the case with $a(t) = \Theta(t^{-\gamma}), \gamma \in (0.5, 1)$. Interestingly, we found that this is not always true, and the mean square convergence rate is $O(t^{-1})$ only if the Laplacian eigenvalues of the network topology graph satisfy certain conditions. It is observed that the fast convergence rate $O(t^{-1})$ depends on the step size $a(t)$ and eigenvalues of the Laplacian matrix. Especially, for the case with balanced graphs, choosing $a(t) = \Theta(t^{-1})$ with $\lambda_2(\widehat{L_{\mathcal{G}}}) \liminf_{t\to\infty}(ta(t)) \gtrsim 1$ will achieve the convergence rate $O(t^{-1})$, where $\lambda_2(\widehat{L_{\mathcal{G}}})$ is the algebraic connectivity of the symmetrized graph. For the case with undirected graphs, the condition on the step size $a(t)$ can be relaxed to $\lambda_2(L_{\mathcal{G}}) \liminf_{t\to\infty}(ta(t)) > 1/2$, where $\lambda_2(L_{\mathcal{G}})$ is the algebraic connectivity of the graph.
- We study the sample path behavior of the consensus error under undirected graphs. It is observed that the consensus error has a convergence rate slightly slower than $O(t^{-\gamma/2})$ almost surely. The upper limit bound of the sample path of the consensus error is calculated.

Compared with the existing related works [12,8,13,22,26,24], we systematically analyze the stochastic convergence rates of the distributed SA consensus algorithm in the sense that both the network topology and the class of step size are more general. Besides, the explicit limit bounds of the stochastic convergence rates are provided, which clearly show the impacts of various kinds of system parameters on the convergence rates, i.e. the number of nodes, the variance of noises, the maximal weight, the eigenvalues of the Laplacian matrix, etc. Also, sufficient conditions are given to achieve fast convergence rate $O(t^{-1})$. These will be all helpful for developing efficient and practical distributed algorithms over large scale sensor networks by designing the step sizes and network parameters.

This paper is organized as follows. In Section 2, we formulate the problem to be investigated. In Section 3, we investigate the dynamic consensus error equation and give the mean square and sample path convergence rates for the SA consensus algorithm. Numerical simulations to corroborate our analytical findings are presented in Section 4, and concluding remarks are given in Section 5. For the sake of conciseness, all the proofs are put in Appendix.

In this paper, we adopt the following notations. $\mathbf{1}_{N\times 1}$ and $\mathbf{0}_{N\times 1}$ denote $N \times 1$ column vectors with all ones and all zeros, respectively. For a given vector or matrix A, A^T denotes its transpose, and $\|A\|$ denotes its 2-norm. For any given complex number λ, $Re(\lambda)$ denotes its real part and $Im(\lambda)$ denotes its imaginary part. We denote $f(t) = o(g(t))$ if $\lim_{t\to\infty} |f(t)/g(t)| = 0$; $f(t) = O(g(t))$ if $\limsup_{t\to\infty} |f(t)/g(t)| < \infty$; $f(t) = \Omega(g(t))$ if $\liminf_{t\to\infty} |f(t)/g(t)| > 0$; and $f(t) = \Theta(g(t))$ if both $f(t) = O(g(t))$ and $f(t) = \Omega(g(t))$. For a differentiable function $f(t)$, $f^{(k)}(t)$ denotes its kth derivative and $f^{(0)}(t) = f(t)$.

2. Problem formulation

For a weighted digraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{A}\}$, $\mathcal{V} = \{1, \ldots, N\}$ denotes the set of N nodes, $\mathcal{E}$ denotes the set of edges, and $\mathcal{A} = [a_{ij}] \in \mathbf{R}^{N\times N}$ denotes the weighted adjacency matrix. The pair $(j, i) \in \mathcal{E} \Leftrightarrow$ node j can send information to node i directly. Then j is called the parent of i. Node i is called a source if it has no parent. The neighborhood of the ith node is denoted by $\mathcal{N}_i = \{j \in \mathcal{V} | (j, i) \in \mathcal{E}\}$. For any given $i, j \in \mathcal{V}$, $a_{ij} \geq 0$, and $a_{ij} > 0$ if and only if $j \in \mathcal{N}_i$. $L_{\mathcal{G}} = \mathcal{D} - \mathcal{A}$ is called the Laplacian matrix of $\mathcal{G}$, where $\mathcal{D} = diag(\sum_{j=1}^{N} a_{1j}, \ldots, \sum_{j=1}^{N} a_{Nj})$. The digraph $\mathcal{G}$ is balanced, if $\sum_{j=1}^{N} a_{ji} = \sum_{j=1}^{N} a_{ij}$ for all $i \in \mathcal{V}$. A directed tree is a digraph, where every node except the root has exactly one parent and the root is a source. A spanning tree of $\mathcal{G}$ is a directed tree whose node set is $\mathcal{V}$ and whose edge set is a subset of $\mathcal{E}$. If the digraph $\mathcal{G}$ contains a spanning tree, then $L_{\mathcal{G}}$ has a unique zero eigenvalue and all other $N-1$ eigenvalues have positive real parts. We denote $\lambda_1 = 0$ and all its distinct non-zero eigenvalues by $\lambda_2, \ldots, \lambda_l$. We denote $\lambda_2^* = \min\{Re(\lambda_m), 2 \leq m \leq l\}$. It is known that there exists a unique probability measure π^T which is the left eigenvector of $L_{\mathcal{G}}$ associated with λ_1, i.e., $\pi^T L_{\mathcal{G}} = \mathbf{0}_{N\times 1}$. If the digraph $\mathcal{G}$ is balanced, then $\pi^T = (1/N)\mathbf{1}_{N\times 1}$.

Consider the discrete-time SA consensus algorithm for a N nodes network

$$x_i(t+1) = x_i(t) + a(t) \sum_{j\in\mathcal{N}_i} a_{ij}(y_{ji}(t) - x_i(t)), \; t \geq 0, \; i \in \mathcal{V}, \tag{1}$$

here the step size $a(t) > 0$; $x_i(t) \in \mathbf{R}$ is the ith node's state, and the initial state $x_i(0)$ is deterministic; $y_{ji}(t)$ is the received information of the ith node from the jth node:

$$y_{ji}(t) = x_j(t) + \omega_{ji}(t), \; j \in \mathcal{N}_i, \tag{2}$$

where $\{\omega_{ji}(t), t \geq 0, i, j \in \mathcal{V}\}$ are the communication noises.

Denote $X(t) = [x_1(t), \ldots, x_N(t)]^T$. Eq. (1) can be rewritten as follows:

$$X(t+1) = (I_N - a(t)L_{\mathcal{G}})X(t) + a(t)\Sigma_{\mathcal{G}}W(t). \tag{3}$$

Here, $W(t) = [w_1^T(t), \ldots, w_N^T(t)]^T$, $w_i(t) = [\omega_{1i}(t), \ldots, \omega_{Ni}(t)]^T$ and $\Sigma_{\mathcal{G}} = diag(\alpha_1^T, \ldots, \alpha_N^T)$ is an $N \times N^2$ dimensional block diagonal matrix with α_i^T being the ith row of the weighted adjacency matrix $\mathcal{A}$.

It was proved in [9] that if the digraph $\mathcal{G}$ contains a spanning tree and $a(t)$ satisfies the standard conditions $\sum_{t=0}^{\infty} a(t) = \infty$, $\sum_{t=0}^{\infty} a^2(t) < \infty$, then the SA consensus algorithm (1)–(2) can achieve both mean square and almost sure consensus, i.e., $E|x_i(t)|^2 < \infty$, and there exists a random variable x^* such that $\lim_{t\to\infty} E|x_i(t) - x^*|^2 = 0$ and $\lim_{t\to\infty} x_i(t) = x^*$ a.s., for all $i \in \mathcal{V}$. Hereinafter, to measure the disagreement among the nodes, we denote $J = \mathbf{1}_{N\times 1}\pi^T$ and the dynamic consensus error by $\delta(t) = (I_N - J)X(t)$.

In this paper, we will study the stochastic convergence rate of $\delta(t)$ and our main goal includes two aspects:

- To analyze the mean square and almost sure convergence rates in relation to the step size $a(t)$ and the network graph parameters.
- To seek the optimal step size for achieving fast convergence.

3. Main results

We make the following assumptions.

1. **Topology:**
   - (A1a) $\mathcal{G}$ contains a spanning tree.
   - (A1b) $\mathcal{G}$ is balanced and strongly connected.[1]
   - (A1c) $\mathcal{G}$ is undirected and connected.
2. **Step size sequence:**
   - (A2) There exists $\gamma \in (0.5, 1]$, $0 < \alpha \le \beta < \infty$ and $T_0 > 0$ such that
$$\alpha t^{-\gamma} \le a(t) \le \beta t^{-\gamma}, \forall\, t \ge T_0. \tag{4}$$
3. **Communication noises:**
   - (A3a) The noise sequence $W(t)$ is a martingale difference sequence, and $\sigma_W^2 \triangleq \sup_{t\ge 0} E(\|W(t)\|^2) < \infty$.
   - (A3b) The communication noises of different channels are uncorrelated, i.e. $E(W(t)W^T(t)) = diag(E(\omega_{11}^2(t)), \ldots, E(\omega_{N1}^2(t)), \ldots, E(\omega_{1N}^2(t)), \ldots, E(\omega_{NN}^2(t)))$.

It is worth noting that Assumption (A2) describes a typical class of step sizes $a(t)$. There are two equivalence forms of (4): (i) $a(t) = \Theta(t^{-\gamma})$; (ii) $0 < \liminf_{t\to\infty}\big(a(t)t^{\gamma}\big) \le \limsup_{t\to\infty}\big(a(t)t^{\gamma}\big) < \infty$. Especially, for any $\varepsilon > 0$, we may take $\alpha = \liminf_{t\to\infty}\big(a(t)t^{\gamma}\big) - \varepsilon$ and $\beta = \limsup_{t\to\infty}\big(a(t)t^{\gamma}\big) + \varepsilon$ with suitable T_0. Hereinafter, for any given $\{a(t), t \ge 0\}$, we will treat the parameters $\alpha, \beta, T_0, \gamma$ as fixed constants to offer more flexibility in theoretical analysis.

3.1. Mean square convergence

We first analyze the mean square convergence rate of the consensus error. Then limit bounds are estimated for the case with balanced digraphs. Finally, a sufficient condition will be given to achieve the fast convergence rate $O(t^{-1})$.

Theorem 3.1. *Suppose that Assumptions (A1a), (A2) and (A3a) hold. For the discrete-time SA consensus algorithm (1)–(2), (i) if $\gamma \in (0.5, 1)$, then $E(\|\delta(t)\|^2) = O(t^{-\gamma})$; (ii) if $\gamma = 1$, then $E(\|\delta(t)\|^2) = O(t^{-1})$ when $\lambda_2^*\alpha > 1$, and $E(\|\delta(t)\|^2) = O(t^{-\lambda\alpha})$ when $\lambda_2^*\alpha \le 1$. Here, λ is any given positive real number less than λ_2^*.*

Remark 3.1. For the SA consensus algorithms, it is known that under Assumptions (A1a) and (A3a), for all $\gamma \in (0.5, 1]$, the step size $a(t) = \Theta(t^{-\gamma})$ can guarantee consensus. One would intuitively believe that $\gamma = 1$ might lead to faster convergence than the case with $\gamma \in (0.5, 1)$ [26,13]. However, Theorem 3.1 tells us that this is not always true, especially, if $\lambda_2^*\alpha \le 1$, then the convergence may be slower than $\Theta(t^{-1})$ even if $\gamma = 1$.

If the network graph is balanced, then more details can be given for the relationship among the convergence rate, the step size $a(t)$ and the network parameters. We have the following theorem.

Theorem 3.2. *Suppose that Assumptions (A1b), (A2), (A3a) and (A3b) hold. For the discrete-time SA consensus algorithm (1)–(2), (i) if $\gamma \in (0.5, 1)$, then*
$$\limsup_{t\to\infty} t^{\gamma}E(\|\delta(t)\|^2) \le C_1\beta^2/(\lambda_2(\widehat{L}_{\mathcal{G}})\alpha); \tag{5}$$
(ii) if $\gamma = 1$, then
$$\limsup_{t\to\infty} tE(\|\delta(t)\|^2) \le C_1\beta^2/(\lambda_2(\widehat{L}_{\mathcal{G}})\alpha - 1) \text{ when } \lambda_2(\widehat{L}_{\mathcal{G}})\alpha > 1; \tag{6}$$
$$\limsup_{t\to\infty} t(\ln t)^{-1}E(\|\delta(t)\|^2) \le C_1\beta^2 \text{ when } \lambda_2(\widehat{L}_{\mathcal{G}})\alpha = 1; \tag{7}$$
and
$$\limsup_{t\to\infty} t^{\lambda_2(\widehat{L}_{\mathcal{G}})\alpha}E(\|\delta(t)\|^2) \le C_1\beta^2/(1 - \lambda_2(\widehat{L}_{\mathcal{G}})\alpha) \text{ when } \lambda_2(\widehat{L}_{\mathcal{G}})\alpha < 1. \tag{8}$$
Here, $C_1 = \max_{1\le i,j\le N} a_{ij}^2\sigma_W^2(N-1)/N$, and $\lambda_2(\widehat{L}_{\mathcal{G}})$ is the smallest nonzero eigenvalue of the Laplacian matrix of the symmetrized graph of $\mathcal{G}$ [15,8].

If the network is an undirected graph, then more precise estimates of the upper limit bound for the consensus error can be given.

Theorem 3.3. *Suppose that Assumptions (A1c), (A2), (A3a) and (A3b) hold. For the discrete-time SA consensus algorithm (1)–(2), (i) if $\gamma \in (0.5, 1)$, then*
$$\limsup_{t\to\infty} t^{\gamma}E(\|\delta(t)\|^2) \le C_1\beta^2/(2\lambda_2\alpha); \tag{9}$$
(ii) if $\gamma = 1$, then
$$\limsup_{t\to\infty} tE(\|\delta(t)\|^2) \le C_1\beta^2/(2\lambda_2\alpha - 1) \text{ when } \lambda_2\alpha > 1/2; \tag{10}$$
$$\limsup_{t\to\infty} t(\ln t)^{-1}E(\|\delta(t)\|^2) \le C_1\beta^2 \text{ when } \lambda_2\alpha = 1/2; \tag{11}$$
and
$$\limsup_{t\to\infty} t^{\lambda_2\alpha}E(\|\delta(t)\|^2) \le C_1\beta^2/(1 - 2\lambda_2\alpha) \text{ when } \lambda_2\alpha < 1/2. \tag{12}$$
Here, C_1 is given in Theorem 3.2.

Theorem 3.4. *Suppose that Assumptions (A1b), (A2), (A3a) and (A3b) hold. For the discrete-time SA consensus algorithm (1)–(2), (i) if $\gamma \in (0.5, 1)$, then*
$$\liminf_{t\to\infty} t^{\gamma}E(\|\delta(t)\|^2) \ge \widetilde{C}_1\alpha^2/(2\lambda_N(\widehat{L}_{\mathcal{G}})\beta); \tag{13}$$
(ii) if $\gamma = 1$, then
$$\liminf_{t\to\infty} tE(\|\delta(t)\|^2) \ge \widetilde{C}_1\alpha^2/(2\lambda_N(\widehat{L}_{\mathcal{G}})\beta - 1) \text{ when } \lambda_N(\widehat{L}_{\mathcal{G}})\beta > 1/2; \tag{14}$$
and
$$\liminf_{t\to\infty} tE(\|\delta(t)\|^2) = \infty \text{ when } \lambda_N(\widehat{L}_{\mathcal{G}})\beta \le 1/2. \tag{15}$$
Here, $\widetilde{C}_1 = \inf_{t\ge 0}\Big\{\sum_{i,j=1}^N a_{ij}^2 E(\omega_{ji}^2(t))\Big\}(N-1)/N$, and $\lambda_N(\widehat{L}_{\mathcal{G}})$ is the largest eigenvalue of $\widehat{L}_{\mathcal{G}}$.

Remark 3.2. Theorems 3.2–3.4 jointly give the explicit convergence rates of the consensus error under balanced digraphs or undirected graphs. If $\gamma \in (0.5, 1)$, then $E(\|\delta(t)\|^2) = \Theta(t^{-\gamma})$; if $\gamma = 1$, then $E(\|\delta(t)\|^2) = \Theta(t^{-1})$ provided $\lambda_2(\widehat{L}_{\mathcal{G}})\alpha > 1$, and

[1] A digraph $\mathcal{G}$ is strongly connected, if for any $i, j \in \mathcal{V}$, there is a directed path from i to j. For a balanced digraph, containing a spanning tree implies being strongly connected.

$\lim_{t\to\infty} tE(\|\delta(t)\|^2) = \infty$ otherwise. Hence, the choice $a(t) = \Theta(t^{-1})$ with $\lambda_2(\widehat{L}_\mathcal{G})\liminf_{t\to\infty}(ta(t)) > 1$ suffices for achieving the fast convergence rate $O(t^{-1})$. And for the case with undirected graphs, the condition on $a(t)$ can be relaxed to $\lambda_2\liminf_{t\to\infty}(ta(t)) > 1/2$ by Theorem 3.3. It is noted that the eigenvalue λ_2, called algebraic connectivity, characterizes the degree of a graph's connectivity. To achieve the fast convergence rate, besides taking the step size $a(t) = \Theta(t^{-1})$, the connectivity degree should be large enough such that $\lambda_2 > 1/(2\liminf_{t\to\infty}(ta(t)))$.

3.2. Almost sure convergence

If the network is undirected, then the sample path convergence rates of $\delta(t)$ can be characterized, which illustrate the properties of the consensus algorithm with probability 1.

Theorem 3.5. *Suppose that Assumptions (A1a), (A1c), (A2) and (A3a) hold. For the discrete-time SA consensus algorithm (1)–(2), (i) if $\gamma \in (0.5, 1)$, then*

$$\limsup_{t\to\infty} t^{\frac{\gamma}{2}}(\Upsilon \log\log(S_t^{\lambda_N}\Upsilon))^{-\frac{1}{2}}\|\delta(t)\| \le \beta\sqrt{\frac{|\mathcal{E}|(N-1)\sum_{i,j}a_{ij}^2}{\lambda_2\alpha}} \quad a.s.; \tag{16}$$

(ii) if $\gamma = 1$, then when $\lambda_2\alpha > 1/2$, we have

$$\limsup_{t\to\infty} t^{\frac{1}{2}}(\Upsilon \log\log(S_t^{\lambda_N}\Upsilon))^{-\frac{1}{2}}\|\delta(t)\| \le \beta\sqrt{\frac{2|\mathcal{E}|(N-1)\sum_{i,j}a_{ij}^2}{2\lambda_2\alpha-1}} \quad a.s.; \tag{17}$$

when $\lambda_2\alpha = 1/2$, we have

$$\limsup_{t\to\infty} t^{\frac{1}{2}}(\ln t)^{-\frac{1}{2}}(\Upsilon \log\log(S_t^{\lambda_N}\Upsilon))^{-\frac{1}{2}}\|\delta(t)\| \le \beta\sqrt{2|\mathcal{E}|(N-1)\sum_{i,j}a_{ij}^2} \quad a.s.; \tag{18}$$

and when $\lambda_2\alpha < 1/2$, we have

$$\limsup_{t\to\infty} t^{\lambda_2\alpha}(\Upsilon \log\log(S_t^{\lambda_N}\Upsilon))^{-\frac{1}{2}}\|\delta(t)\| \le \beta\sqrt{\frac{2|\mathcal{E}|(N-1)\sum_{i,j}a_{ij}^2}{1-2\lambda_2\alpha}} \quad a.s. \tag{19}$$

Here, $\Upsilon = \sup_{t\ge0}\max_{1\le i,j\le N}E(\omega_{ji}^2(t))$, $S_t^{\lambda_N} = \sum_{s=T}^t a^2(s)\prod_{r=T}^s(1-\lambda_N a(r))^{-2}$, λ_2 and λ_N are the smallest and largest nonzero eigenvalues of the Laplacian matrix $L_\mathcal{G}$, $T \ge \max\{T_1, T_2, T_5\}$ with T_1, T_2 and T_5 are given, respectively, in Lemma A.3, the proof of Theorem 3.1, and Lemma A.6.

Remark 3.3. For any given $\epsilon > 0$, we have $(\log\log S_t^{\lambda_N})^{1/2} = O(t^\epsilon)$. Thus if the communication noise is i.i.d. standard white noise, from Theorem 3.5, we can conclude that $\|\delta(t)\| = O(t^{-\frac{\gamma}{2}+\epsilon})$ almost surely.

4. Numerical example

In this section, we present numerical examples to corroborate our analytical findings. In all the simulations presented, the communication noises are Gaussian white noises with distribution $N(0, 1)$. To compute the mean square of the consensus error, we simulate 1000 runs of the SA consensus algorithm from the initial state. And the iteration time is 500.

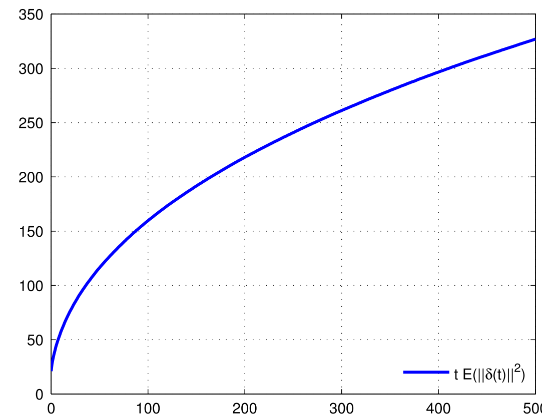

Fig. 1. The curve of $tE(\|\delta(t)\|^2)$ for the 4 nodes network with $a_1(t) = 0.2(t+1)^{-1}$.

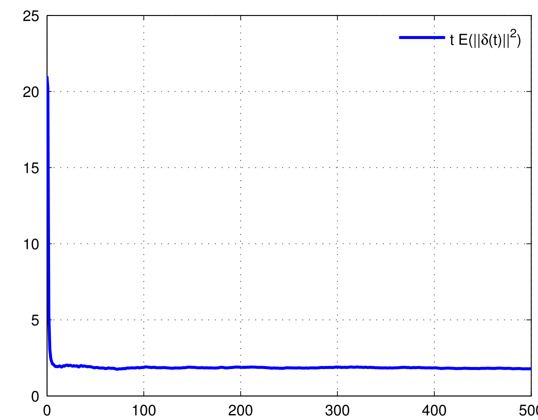

Fig. 2. The curve of $tE(\|\delta(t)\|^2)$ for the 4 nodes network with $a_2(t) = (t+1)^{-1}$.

4.1. The sufficient condition to achieve fast convergence rate

Example 4.1. Consider a sensor network of 4 nodes with the network topology graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{A}\}$, where $\mathcal{V} = \{1, 2, 3, 4\}$, $\mathcal{E} = \{(2, 1), (4, 2), (2, 3), (4, 3), (1, 4), (2, 4)\}$ and $\mathcal{A}$ is the $0-1$ weight adjacency matrix. The initial condition is given as $X(0) = [1, 6, 2, 3]^T$. We take the step size $a_1(t) = 0.2(t+1)^{-1}$ and $a_2(t) = (t+1)^{-1}$, respectively, for the SA consensus algorithm (1). Note that $\lambda_2^*(L_\mathcal{G}) = 2$, $\lambda_2^*(L_\mathcal{G})\lim_{t\to\infty}ta_1(t) = 0.4 < 1$, and $\lambda_2^*(L_\mathcal{G})\lim_{t\to\infty}ta_2(t) = 2 > 1$.

In Figs. 1 and 2, we show the convergence rate of the SA consensus algorithm by plotting $tE(\|\delta(t)\|^2)$ versus different step sizes $a_1(t)$ and $a_2(t)$, respectively. It can be seen that for the case with $a_1(t)$, $tE(\|\delta(t)\|^2)$ explodes, and for the case with $a_2(t)$, $tE(\|\delta(t)\|^2)$ decays fast initially and then reaches a steady state, which means that the convergence rate of $E(\|\delta(t)\|^2)$ is slower than $\Theta(t^{-1})$ for the case with $a_1(t)$, and $E(\|\delta(t)\|^2) = \Theta(t^{-1})$ for the case with $a_2(t)$, which are consistent with Theorem 3.1 and Remark 3.1.

Example 4.2. Consider a ring sensor network with 5 nodes where each node has two neighbors. The ith node's neighbor set $\mathcal{N}_i = \{i-1, i+1\}$ for $2 \le i \le 4$, $\mathcal{N}_1 = \{5, 2\}$, and $\mathcal{N}_5 = \{4, 1\}$. Clearly it is a 2-regular undirected graph. The initial condition is given as $x_i(0) = 2i - 1$ for $1 \le i \le 5$. We take the decreasing step size $a(t) = 0.3(t+1)^{-1}$. Noting that $\lambda_2(L_\mathcal{G}) = 1.382$, we have $\lambda_2(L_\mathcal{G})\lim_{t\to\infty}ta(t) = 0.4146 < 0.5$, which means that the algebraic connectivity $\lambda_2(L_\mathcal{G})$ of the network graph is not sufficiently large to meet the sufficient condition illustrated in

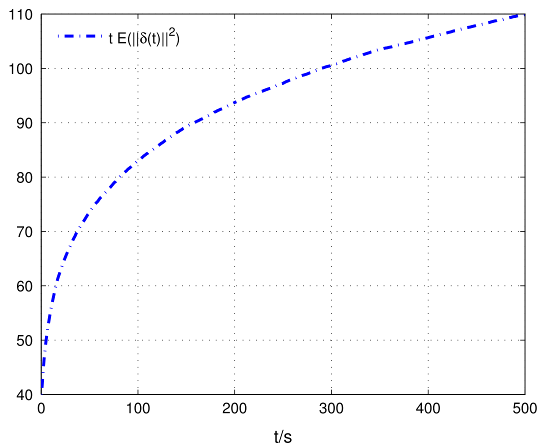

Fig. 3. The curve of $tE(\|\delta(t)\|^2)$ for the 5 nodes ring network.

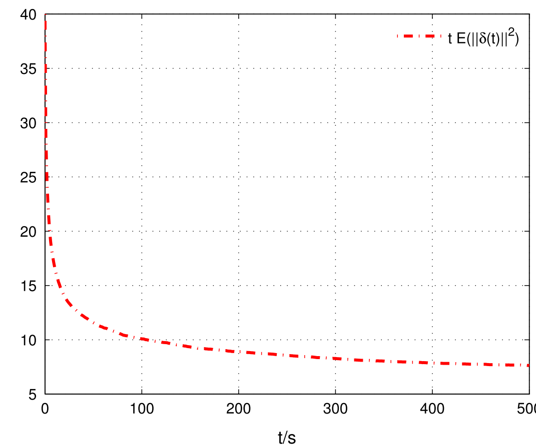

Fig. 5. The curve of $tE(\|\delta(t)\|^2)$ for the 5 nodes ring network with 2 added edges.

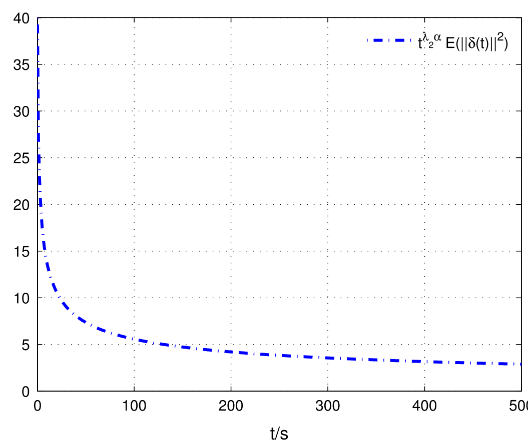

Fig. 4. The curve of $t^{\lambda_2(L_{\mathcal{G}})\alpha}E(\|\delta(t)\|^2)$ for the 5 nodes ring network.

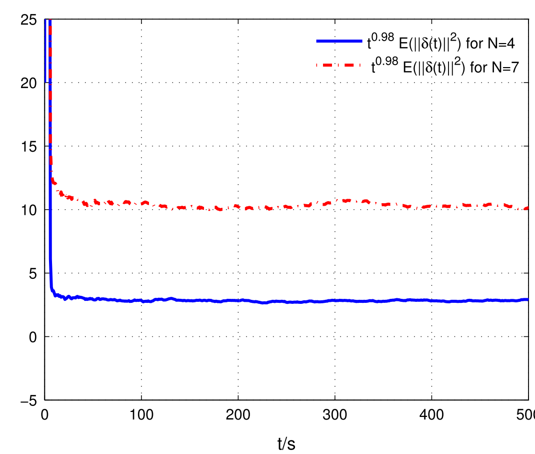

Fig. 6. The curve of $t^{0.98}E(\|\delta(t)\|^2)$ for the ring graph network with $a(t) = 2(t+1)^{-0.98}$.

Remark 3.2. From Figs. 3 and 4, it can be seen that $tE(\|\delta(t)\|^2)$ explodes, while $t^{\lambda_2(L_{\mathcal{G}})\alpha}E(\|\delta(t)\|^2)$ converges. The convergence rate of $E(\|\delta(t)\|^2)$ fails to attain $\Theta(t^{-1})$.

Now we increase the algebraic connectivity of the ring graph by adding four edges (3, 1), (1, 3), (4, 2) and (2, 4). The new graph is denoted by $\widetilde{\mathcal{G}} = \{\mathcal{V}, \widetilde{\mathcal{E}}, \widetilde{\mathcal{A}}\}$. Take the same step size $a(t) = 0.3(t+1)^{-1}$. Noting that $\lambda_2(L_{\widetilde{\mathcal{G}}}) = 2$, it can be verified that $\lambda_2(L_{\widetilde{\mathcal{G}}})\lim_{t\to\infty} ta(t) = 0.6 > 0.5$, which meets the sufficient condition to achieve the convergence rate $\Theta(t^{-1})$ illustrated in Remark 3.2. From Fig. 5, it can be seen that $tE(\|\delta(t)\|^2)$ converges, which means that $E(\|\delta(t)\|^2) = \Theta(t^{-1})$.

4.2. Graph-related limit bounds

Example 4.3. Consider a ring sensor network with N nodes where each node has two neighbors as in Example 4.2. The initial condition is also given as $x_i(0) = 2i - 1$ for $1 \le i \le N$. We take the decreasing step size $a(t) = 2(t+1)^{-0.98}$. The convergence rate of the consensus error is shown when $N = 4$ and $N = 7$, respectively, in Fig. 6, where the solid blue line denotes the curve of $t^{0.98}E(\|\delta(t)\|^2)$ for $N = 4$, whereas the dashed red line denotes the curve for $N = 7$. The two curves achieve steady values around 3 and 10, respectively.

The theoretical value of $\lim_{t\to\infty} t^{0.98}E(\|\delta(t)\|^2)$ predicted by our analysis is given by Theorems 3.3 and 3.4. When $N = 4$, it is easy to check that $C_1 = \tilde{C}_1 = 2(N-1) = 6$, $\lambda_2 = 2$, $\lambda_4 = 4$, and $\lim_{t\to\infty} t^{0.98}E(\|\delta(t)\|^2) \in [1.5, 3]$. When $N = 7$, it is easy to check that $C_1 = \tilde{C}_1 = 12$, $\lambda_2 = 0.753$, $\lambda_7 = 3.8019$, and $\lim_{t\to\infty} t^{0.98}E(\|\delta(t)\|^2) \in [3.1563, 15.9385]$. Here, we take $\alpha = \beta = \lim_{t\to\infty}(t^{0.98}a(t)) = 2$ for the simplicity of computation. The numerical result demonstrated in Fig. 6 shows that the limit bounds given by the theoretical results are quite reasonable.

5. Conclusions

In this paper, we have considered the stochastic convergence rates for SA consensus algorithms with martingale difference sequence communication noises under directed topologies. By the tools of stochastic analysis and algebraic graph theory, we have quantified the relationship among the convergence rate, the step size $a(t)$, and the network topology. We found that if $a(t) = \Theta(t^{-\gamma})$ with $\gamma \in (0.5, 1]$, then for the case with general digraphs with spanning trees, the mean square of the consensus error has a convergence rate of $O(t^{-\gamma})$ if $\gamma \in (0.5, 1)$ and $O(t^{-1})$ if $\gamma = 1$ and $\lambda_2^*\alpha > 1$. These results are further extended to the case with balanced digraphs and undirected graphs, which indicates the optimal choice of $\{a(t), t \ge 0\}$ to achieve fast convergence rate. Furthermore, we give the calculation methods for the limit bounds of the convergence rate of the consensus error by the network topology parameters, e.g., the number of nodes and edges. The sample path convergence rate performances are also derived for undirected graphs.

There are many interesting topics deserving further investigation, such as to get the convergence time for any given accuracy and extend our analysis to networks with nonlinear dynamics, random link failures, time-varying topologies, and time delays.

Acknowledgments

This work was supported by National Natural Science Foundation of China under Grants 61522310 and 61603215, the China Postdoctoral Science Foundation 2016M601152, Young Scholars Program of Shandong University, Shanghai Rising-Star Program under Grant 15QA1402000 and the Shu Guang project of Shanghai Municipal Education Commission and Shanghai Education Development Foundation under grant 17SG26.

Appendix

Lemma A.1 ([9]). *Suppose that Assumption (A1a) holds. Then for any given matrix $\phi_{N\times(N-1)}$ satisfying span$\{\phi_{N\times(N-1)}\}$ = span$\{L_{\mathcal{G}}\}$, the matrix $\Phi=(\mathbf{1}_{N\times 1}\ \phi_{N\times(N-1)})$ is nonsingular, Φ^{-1} has the representation $\Phi^{-1}=\begin{pmatrix}\pi^T\\ \psi_{(N-1)\times N}\end{pmatrix}$ and $\Phi^{-1}L_{\mathcal{G}}\Phi=\begin{pmatrix}0 & \mathbf{0}_{1\times(N-1)}\\ \mathbf{0}_{(N-1)\times 1} & \tilde{L}_{\mathcal{G}}\end{pmatrix}$. Here, all $N-1$ eigenvalues of $\tilde{L}_{\mathcal{G}}$, which are also nonzero eigenvalues of $L_{\mathcal{G}}$, have positive real parts.*

Lemma A.2. *Under Assumption (A2), for any given $t>s>T_0$ and $\lambda>0$, if $\gamma\in(0.5,1)$, we have*

$$e^{\frac{\lambda\beta}{1-\gamma}s^{1-\gamma}}e^{-\frac{\lambda\beta}{1-\gamma}t^{1-\gamma}} \le e^{-\lambda\sum_{r=s+1}^{t}a(r)} \le e^{\frac{\lambda\alpha}{1-\gamma}(s+1)^{1-\gamma}}e^{-\frac{\lambda\alpha}{1-\gamma}(t+1)^{1-\gamma}}, \tag{A.1}$$

and if $\gamma=1$, then

$$s^{\lambda\beta}t^{-\lambda\beta} \le e^{-\lambda\sum_{r=s+1}^{t}a(r)} \le (s+1)^{\lambda\alpha}(t+1)^{-\lambda\alpha}. \tag{A.2}$$

The proof is straightforward and is omitted here.

Lemma A.3. *Under Assumption (A2), if $\gamma\in(0.5,1)$, for any given real number $\lambda>0$, there exists $T_1>1$ such that for any given $T\ge T_1$, we have*

$$\limsup_{t\to\infty} t^{\gamma}\sum_{s=T}^{t}a^2(s)e^{-\lambda\sum_{r=s+1}^{t}a(r)} \le \beta^2/(\lambda\alpha), \tag{A.3}$$

and

$$\liminf_{t\to\infty} t^{\gamma}\sum_{s=T}^{t}a^2(s)e^{-\lambda\sum_{r=s+1}^{t}a(r)} \ge \alpha^2/(\lambda\beta). \tag{A.4}$$

Proof. By Assumption (A2) and Lemma A.2, for any given $T\ge T_0$, we have

$$\sum_{s=T}^{t}a^2(s)e^{-\lambda\sum_{r=s+1}^{t}a(r)} \le \beta^2 e^{-\frac{\lambda\alpha}{1-\gamma}(t+1)^{1-\gamma}}\sum_{s=T}^{t}s^{-2\gamma}e^{\frac{\lambda\alpha}{1-\gamma}(s+1)^{1-\gamma}}. \tag{A.5}$$

Denote $\delta=\lambda\alpha/(1-\gamma)$. It is easy to check that there exists a constant $C>0$ such that for any given $s>C$, $f(s)=s^{-2\gamma}e^{\sigma(s+1)^{1-\gamma}}$ is increasing. Therefore for all $T>C$, we have

$$\sum_{s=T}^{t}s^{-2\gamma}e^{\delta(s+1)^{1-\gamma}} \le \int_T^{t+1}s^{-2\gamma}e^{\delta(s+1)^{1-\gamma}}ds. \tag{A.6}$$

By L'Hôpital's rule, it follows that

$$\lim_{t\to\infty}\frac{\int_T^{t+1}s^{-2\gamma}e^{\delta(s+1)^{1-\gamma}}ds}{(t+1)^{-\gamma}e^{\delta(t+2)^{1-\gamma}}}=\frac{1}{\lambda\alpha}. \tag{A.7}$$

Denote $T_1=\max\{T_0,C\}$. By Eqs. (A.5), (A.6) and (A.7), and noting that $\lim_{t\to\infty}t^{\gamma}e^{-\delta(t+1)^{1-\gamma}}(t+1)^{-\gamma}e^{\delta(t+2)^{1-\gamma}}=1$, Eq. (A.3) follows. (A.4) can be obtained similarly.

Lemma A.4. *Under Assumption (A2), if $\gamma=1$, when $\lambda\alpha>1$, then for any given $T>T_0$, we have*

$$\limsup_{t\to\infty} t\sum_{s=T}^{t}a^2(s)e^{-\lambda\sum_{r=s+1}^{t}a(r)} \le \beta^2/(\lambda\alpha-1); \tag{A.8}$$

when $\lambda\alpha=1$, we have

$$\limsup_{t\to\infty} t(\ln t)^{-1}\sum_{s=T}^{t}a^2(s)e^{-\lambda\sum_{r=s+1}^{t}a(r)} \le \beta^2; \tag{A.9}$$

and when $\lambda\alpha<1$, we have

$$\limsup_{t\to\infty} t^{\lambda\alpha}\sum_{s=T}^{t}a^2(s)e^{-\lambda\sum_{r=s+1}^{t}a(r)} \le \beta^2/(1-\lambda\alpha). \tag{A.10}$$

Proof. By Assumption (A2) and Lemma A.2, we have

$$\sum_{s=T}^{t}a^2(s)e^{-\lambda\sum_{r=s+1}^{t}a(r)} \le \beta^2(t+1)^{-\lambda\alpha}\int_{T-1}^{t+1}(s^{\lambda\alpha-2}+O(s^{\lambda\alpha-3}))ds.$$

It is easy to check that

$$\int_{T-1}^{t+1}s^{\lambda\alpha-2}ds = \begin{cases}\ln(t+1)-\ln(T-1), & \text{if } \lambda\alpha=1;\\ 1/(\lambda\alpha-1)((t+1)^{\lambda\alpha-1}-(T-1)^{\lambda\alpha-1}) & \text{if } \lambda\alpha\ne 1.\end{cases}$$

Note that $(T-1)^{\lambda\alpha-1}<1$ if $\lambda\alpha\le 1$ and $T>T_0$. And the conclusion follows.

Lemma A.5. *Under Assumption (A2), if $\gamma=1$, then for any given $T>T_0$, when $\lambda\beta>1$, we have*

$$\liminf_{t\to\infty} t\sum_{s=T}^{t}a^2(s)e^{-\lambda\sum_{r=s+1}^{t}a(r)} \ge \alpha^2/(\lambda\beta-1); \tag{A.11}$$

and when $\lambda\beta\le 1$, we have

$$\liminf_{t\to\infty} t\sum_{s=T}^{t}a^2(s)e^{-\lambda\sum_{r=s+1}^{t}a(r)} = \infty. \tag{A.12}$$

The proof is similar to Lemma A.4.

Lemma A.6. *Under Assumption (A2), for any given positive real number λ, there exists $T_5>0$ such that*

$$\lim_{t\to\infty}\sum_{s=T}^{t}a^2(s)\prod_{r=T}^{s}(1-\lambda a(r))^{-2}=\infty, \forall T\ge T_5. \tag{A.13}$$

Proof. For all $T\ge T_0$, with T_0 given in Assumption (A2), we denote $S_t^{\lambda}=\sum_{s=T}^{t}a^2(s)\prod_{r=T}^{s}(1-\lambda a(r))^{-2}$. In view of $(1-x)^{-1}\ge e^x$ for all $0<x<1$, we have

$$S_t^{\lambda}\ge\alpha^2\sum_{s=T}^{t}s^{-2\gamma}e^{2\lambda\alpha\sum_{r=T}^{s}r^{\gamma}}\ge\alpha^2e^{-\frac{2\lambda\alpha}{1+\gamma}T^{1+\gamma}}\sum_{s=T}^{t}s^{-2\gamma}e^{\frac{2\lambda\alpha}{1+\gamma}s^{1+\gamma}}.$$

It is easy to check that $f(s)=s^{-2\gamma}e^{\frac{2\lambda\alpha}{1+\gamma}s^{1+\gamma}}$ is increasing when $s>\gamma/(\lambda\alpha)$. Hence, for all $T>\gamma/(\lambda\alpha)$, we have $S_t^{\lambda}\ge\alpha^2 e^{-\frac{2\lambda\alpha}{1+\gamma}T^{1+\gamma}}\int_T^t s^{-2\gamma}e^{\frac{2\lambda\alpha}{1+\gamma}s^{1+\gamma}}ds$. Note that $\lim_{t\to\infty}s^{1-2\gamma}e^{\frac{2\lambda\alpha}{1+\gamma}s^{1+\gamma}}=\infty$. By Cauchy criteria of improper integral, we have $\lim_{t\to\infty}\int_T^t s^{-2\gamma}$

$e^{\frac{2\lambda\alpha}{1+\gamma}s^{1+\gamma}}ds = \infty$. Denote $T_5 = \max\{T_0, \gamma/(\lambda\alpha)\}$. And (A.13) follows.

Proof of Theorem 3.1. (i) By (3) and the equation $(I_N - J)(I_N - a(t)L_{\mathcal{G}}) = (I_N - a(t)L_{\mathcal{G}})(I_N - J)$, it is easy to check that $\delta(t+1) = (I_N - a(t)L_{\mathcal{G}})\delta(t) + a(t)(I_N - J)\Sigma_{\mathcal{G}}W(t)$. Define $\tilde{\delta}(t) = \Phi^{-1}\delta(t)$, with Φ^{-1} given in Lemma A.1. Denote $\eta(t) = (\tilde{\delta}_2(t), \ldots, \tilde{\delta}_N(t))^T$. In view of that $\pi^T(I_N - J) = \pi^T\big(I_N - \mathbf{1}_{N\times 1}\pi^T\big) = \mathbf{0}_{1\times N}$, it follows that for any given $T \ge T_0$ with T_0 is given in Assumption (A2), the closed-form solution can be written as

$$\tilde{\delta}_1(t+1) = \tilde{\delta}_1(0) = \pi^T\delta(0) = \pi^T(I_N - J)X(0) = 0, \tag{A.14}$$

$$\begin{aligned}\eta(t+1) = &\prod_{s=T}^{t}(I_{N-1} - a(s)\tilde{L}_{\mathcal{G}})\eta(T)\\ &+ \sum_{s=T}^{t}\Big(\prod_{r=s+1}^{t}(I_{N-1} - a(r)\tilde{L}_{\mathcal{G}})\Big)\\ &\times a(s)\psi_{(N-1)\times N}(I_N - J)\Sigma_{\mathcal{G}}W(s).\end{aligned} \tag{A.15}$$

Here, $\tilde{L}_{\mathcal{G}}$ is given in Lemma A.1. Define $f(T,t,\tilde{L}_{\mathcal{G}}) = \prod_{s=T}^{t}(I_{N-1} - a(s)\tilde{L}_{\mathcal{G}})$, and $g(s,t,\tilde{L}_{\mathcal{G}}) = \prod_{r=s+1}^{t}(I_{N-1} - a(r)\tilde{L}_{\mathcal{G}})$. By properties of the 2-norm, we have

$$\begin{aligned}E(\|\eta(t+1)\|^2) \le\ &\|f(T,t,\tilde{L}_{\mathcal{G}})\|^2 E(\|\eta(T)^2\|)\\ &+ C_0\|\sum_{s=T}^{t}a(s)g(s,t,\tilde{L}_{\mathcal{G}})\|^2,\end{aligned} \tag{A.16}$$

where $C_0 = \|\psi_{(N-1)\times N}(I_N - J)\Sigma_{\mathcal{G}}\|^2\sigma_W^2$. Then it will be sufficient to analyze the convergence rate of $\|f(T,t,\tilde{L}_{\mathcal{G}})\|^2$ and $\|\sum_{s=T}^{t}a(s)g(s,t,\tilde{L}_{\mathcal{G}})\|^2$, respectively.

Assume the Jordan normal form of $\tilde{L}_{\mathcal{G}}$ is $diag(J_{\lambda_1,n_1}, \ldots, J_{\lambda_q,n_q})$, where J_{λ_i,n_i} is an $n_i \times n_i$ Jordan block with the diagonal filled with λ_i and the superdiagonal composed by ones. Then we can assert that each element of the matrix function $f(T,t,\tilde{L}_{\mathcal{G}})$ and $\sum_{s=T}^{t}a(s)g(s,t,\tilde{L}_{\mathcal{G}})$ are finite linear combinations of $f^{(k)}_{\lambda_m}(T,t,\lambda_m)$, and $\sum_{s=T}^{t}a(s)g^{(k)}_{\lambda_m}(s,t,\lambda_m)$, respectively, $m = 2,\ldots,q$, $k = 0,1,\ldots,n_m-1$, $i,j \in \mathcal{V}$. Here, $f(T,t,\lambda) = \prod_{s=T}^{t}(1 - a(s)\lambda)$, and $g(s,t,\lambda) = \prod_{r=s+1}^{t}(1 - a(r)\lambda)$. It will be sufficient to analyze these items separately.

For any given λ_m, $m = 2,\ldots,q$, we have $\|1 - a(s)\lambda_m\| = \big(1 + a^2(s)|\lambda_m|^2 - 2a(s)Re(\lambda_m)\big)^{1/2}$. By Assumption (A2), there exists $T_2 > 0$ such that for any given $T \ge T_2$, we have $a(t) \le \min\{Re(\lambda_m)/|\lambda_m|^2, 3/(8Re(\lambda_m))\}$. Then we can conclude that

$$\|1 - a(s)\lambda_m\| \le \big(1 - a(s)Re(\lambda_m)\big)^{1/2}, \tag{A.17}$$

and

$$\|1 - a(s)\lambda_m\| \ge \big(1 - 2a(s)Re(\lambda_m)\big)^{1/2} \ge 1/2. \tag{A.18}$$

By (A.17) and the fact that $1 - x \le e^{-x}$ for $x \ge 0$, it follows that for any given $T \ge \max\{T_0, T_2\}$,

$$\begin{aligned}\|f(T,t,\lambda_m)\| \le\ &\prod_{s=T}^{t}\big(1 - a(s)Re(\lambda_m)\big)^{1/2}\\ \le\ &e^{-(Re(\lambda_m)/2)\sum_{s=T}^{t}a(s)}.\end{aligned} \tag{A.19}$$

For $f^{(1)}_{\lambda_m}(T,t,\lambda_m)$, by (A.18), we have

$$\begin{aligned}\|f^{(1)}_{\lambda_m}(T,t,\lambda_m)\| &\le \|f(T,t,\lambda_m)\|\sum_{s_1=T}^{t}\frac{a(s_1)}{\|1 - a(s_1)\lambda_m\|}\\ &\le 2\|f(T,t,\lambda_m)\|\sum_{s_1=T}^{t}a(s_1),\end{aligned}$$

$$\begin{aligned}&\|f^{(2)}_{\lambda_m}(T,t,\lambda_m)\|\\ &= \Big\|\sum_{s_1=T}^{t}\Big(a(s_1)\sum_{s_2=T,s_2\ne s_1}^{t}\Big(a(s_2)\prod_{s=T,s\ne s_1,s_2}^{t}(1 - a(s)\lambda_m)\Big)\Big)\Big\|\\ &\le \|f(T,t,\lambda_m)\|\sum_{s_1=T}^{t}\Big(\frac{a(s_1)}{\|1 - a(s_1)\lambda_m\|}\sum_{s_2=T,s_2\ne s_1}^{t}\frac{a(s_2)}{\|1 - a(s_2)\lambda_m\|}\Big)\\ &\le 2^2\|f(T,t,\lambda_m)\|(\sum_{s_1=T}^{t}a(s_1))^2.\end{aligned}$$

Similarly, for $k = 3,\ldots,n_m-1$, we have

$$\|f^{(k)}_{\lambda_m}(T,t,\lambda_m)\| \le 2^k\|f(T,t,\lambda_m)\|\Big(\sum_{s_1=T}^{t}a(s_1)\Big)^k. \tag{A.20}$$

Noting that for any given $\epsilon > 0$ and $k = 0,1,\ldots,n_m-1$, there exists a constant $C(k,\epsilon) > 0$ such that

$$\Big(\sum_{s_1=T}^{t}a(s_1)\Big)^k \le C(k,\epsilon)e^{\epsilon\sum_{s_1=T}^{t}a(s_1)}. \tag{A.21}$$

By Eqs. (A.19), (A.20) and (A.21), it follows that for any given $\lambda \in (0,\lambda_2^*)$, and $m = 2,\ldots,q$, $k = 0,1,\ldots,n_m-1$, we have

$$\|f^{(k)}_{\lambda_m}(T,t,\lambda_m)\| = O(e^{-(\lambda/2)\sum_{s=T}^{t}a(s)}). \tag{A.22}$$

Similarly, we have

$$\|g^{(k)}_{\lambda_m}(s,t,\lambda_m)\| = O(e^{-(\lambda/2)\sum_{r=s+1}^{t}a(r)}). \tag{A.23}$$

In the remainder of this proof, without loss of generality, we assume that $T \ge \max\{T_1, T_2\}$, with T_1 is given in Lemma A.3.

If $\gamma \in (0.5, 1)$, by Lemma A.2, it follows that $e^{-(\lambda/2)\sum_{s=T}^{t}a(s)} = o(t^{-1})$. Then by Eqs. (A.22)–(A.23) and Lemma A.3, we have $E(\|\delta(t)\|^2) = O(t^{-\gamma})$.

(ii) If $\gamma = 1$, when $\lambda_2^*\alpha > 1$, we can always take ϵ and $\lambda = \lambda_2^* - \epsilon$ satisfying that $\lambda\beta > \lambda\alpha > 1$. By Lemma A.2, we have $e^{-\lambda\sum_{s=T}^{t}a(s)} = O(t^{-1/2})$. Combining Lemma A.4 and Eqs. (A.22)–(A.23), we can conclude that $E(\|\delta(t)\|^2) = O(t^{-1})$. If $\lambda_2^*\alpha \le 1$, then for any given λ less than λ_2^*, the conclusion $E(\|\delta(t)\|^2) = O(t^{-\lambda\alpha})$ follows by Lemmas A.2 and A.4.

Proof of Theorem 3.2. By Assumption (A1b), we know that $\widehat{L}_{\mathcal{G}}$ can be diagonalize [15] and has N real eigenvalues $0 = \lambda_1(\widehat{L}_{\mathcal{G}}) \le \lambda_2(\widehat{L}_{\mathcal{G}}) \le \cdots \le \lambda_N(\widehat{L}_{\mathcal{G}})$. Denote $\Xi = [\xi_1, \xi_2, \ldots, \xi_N]$, where ξ_i is the unit eigenvector of $\widehat{L}_{\mathcal{G}}$ associated with $\lambda_i(\widehat{L}_{\mathcal{G}})$. Note that $\pi^T = (1/N)\mathbf{1}_{1\times N}$, hence $\xi_1 = 1/(\sqrt{N})\mathbf{1}_{N\times 1}$, and $\Xi^T\widehat{L}_{\mathcal{G}}\Xi = \Lambda_{\widehat{L}_{\mathcal{G}}} = diag(0, \lambda_2(\widehat{L}_{\mathcal{G}}), \ldots, \lambda_N(\widehat{L}_{\mathcal{G}}))$. Define $\tilde{\delta}(t) = \Xi^T\delta(t)$. Denote $\Xi^T L_{\mathcal{G}}\Xi = \begin{pmatrix} 0 & \mathbf{0}_{1\times(N-1)} \\ \mathbf{0}_{(N-1)\times 1} & \tilde{L}_{\mathcal{G}}\end{pmatrix}$, $\eta(t) = (\tilde{\delta}_2(t), \ldots, \tilde{\delta}_N(t))^T$, and $\Xi_{N\times(N-1)} = [\xi_2, \ldots, \xi_N]$. Thus, we have

$$\begin{aligned}E(\|\delta(t+1)\|^2) =\ &E(\eta^T(t)(I_{N-1} - 2a(t)\Lambda_{\widehat{L}_{\mathcal{G}}} + a^2(t)\tilde{L}_{\mathcal{G}}^T\tilde{L}_{\mathcal{G}})\eta(t))\\ &+ E(\|\Xi^T_{N\times(N-1)}(I_N - J)\Sigma_{\mathcal{G}}W(t)\|^2)a^2(t).\end{aligned} \tag{A.24}$$

Note that $\Sigma_{\mathcal{G}}W(t) = (\sum_{i=1}^{N}a_{1i}\omega_{i1}, \ldots, \sum_{i=1}^{N}a_{Ni}\omega_{iN})$, $\Xi^T_{N\times(N-1)}\Xi_{N\times(N-1)} = I_{N-1}$, and $\Xi_{N\times(N-1)}\Xi^T_{N\times(N-1)} = I_N - J_N$. By Assumption (A3b), we have

$$\begin{aligned}&E(\|\Xi^T_{N\times(N-1)}(I_N - J)\Sigma_{\mathcal{G}}W(t)\|^2)\\ &= \sum_{i,j=1}^{N}a_{ij}^2E(\omega_{ji}^2(t))(N-1)/N.\end{aligned} \tag{A.25}$$

By Assumption (A2), there exists $T_3 > 0$ such that $a(t)\lambda_2(\widehat{L}_{\mathcal{G}}) - a^2(t)\|\widetilde{L}_{\mathcal{G}}\|^2 > 0, \forall t \geq T_3$. In the remainder of this proof, without loss of generality, we assume that $T \geq \max\{T_1, T_2, T_3\}$, with T_1 and T_2 are given, respectively, in Lemma A.3 and the proof of Theorem 3.1.

Thus

$$\begin{aligned} E(\|\delta(t+1)\|^2) \leq & \ (1-\lambda_2(\widehat{L}_{\mathcal{G}})a(t))E\|\delta(t)\|^2 + C_1 a^2(t). \\ \leq & \ e^{-\lambda_2(\widehat{L}_{\mathcal{G}})\sum_{s=T}^{t} a(s)}E(\|\delta(T)\|^2) \\ & + C_1 \sum_{s=T}^{t} a^2(s) e^{-\lambda_2(\widehat{L}_{\mathcal{G}})\sum_{r=s+1}^{t} a(r)}. \end{aligned}$$

If $\gamma \in (0.5, 1)$, then (5) follows from Lemmas A.2 and A.3. If $\gamma = 1$, the result follows by Lemmas A.2 and A.4.

Proof of Theorem 3.3. Note that for the undirected graph, its Laplacian matrix $L_{\mathcal{G}}$ is diagonalizable. We denote its N real eigenvalues by $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. Denote $\Gamma = [\gamma_1, \gamma_2, \ldots, \gamma_N]$, where γ_i is the unit eigenvector of $L_{\mathcal{G}}$ associated with λ_i. Clearly, we have $\gamma_1 = (1/\sqrt{N})\mathbf{1}_{N\times 1}$. Denote $\tilde{\delta}(t) = \Gamma^{-1}\delta(t)$ and $\eta(t) = (\tilde{\delta}_2(t), \ldots, \tilde{\delta}_N(t))^T$. It is easy to check that $E(\|\delta(t+1)\|^2) \leq (1-\lambda_2 a(t))^2 E(\|\delta(t)\|^2) + C_1 a^2(t)$. With the similar analysis in the proof of Theorem 3.2, the results in Theorem 3.3 can be obtained.

Proof of Theorem 3.4. For any given $T > 0$, by Eqs. (A.24) and (A.25), we have

$$\begin{aligned} & E(\|\delta(t+1)\|^2) \\ \geq & \prod_{s=T}^{t}(1-2\lambda_N(\widehat{L}_{\mathcal{G}})a(s))E(\|\delta(T)\|^2) \\ & + \tilde{C}_1 \sum_{s=T}^{t}\Big(\prod_{r=s+1}^{t}(1-2\lambda_N(\widehat{L}_{\mathcal{G}})a(r))\Big)a^2(s). \end{aligned}$$

For any given $\epsilon > 0$, by Assumption (A2), we can conclude that there exists $T_4 > 0$ such that $a(t)\lambda_N(\widehat{L}_{\mathcal{G}}) \leq (\ln(1+\epsilon))/(1+\epsilon), \forall t \geq T_4$. In the remainder of this proof, without loss of generality, we assume that $T \geq \max\{T_1, T_2, T_4\}$, with T_1 and T_2 are given, respectively, in Lemma A.3 and the proof of Theorem 3.1. In view of $1-x \geq e^{-(1+\epsilon)x}$ for all $x \in (0, (\ln(1+\epsilon))/(1+\epsilon))$, we have $E(\|\delta(t+1)\|^2) \geq \tilde{C}_1\sum_{s=T}^{t} a^2(s)e^{-2(1+\epsilon)\lambda_N(\widehat{L}_{\mathcal{G}})\sum_{r=s+1}^{t} a(r)}$. By Lemma A.3, we have $\liminf_{t\to\infty} t^{\gamma}E(\|\delta(t)\|^2) \geq \tilde{C}_1\alpha^2/(2(1+\epsilon)\lambda_N(\widehat{L}_{\mathcal{G}})\beta)$. And the result in Eq. (13) follows by letting $\epsilon \to 0$. Eq. (14) can be obtained similarly by Lemma A.5.

Proof of Theorem 3.5. We will continue to adopt the notations in the proof of Theorem 3.3. By the properties of 2-norm, we have

$$\|\delta(t+1)\| \leq \|\Gamma\|\|\tilde{\delta}(t+1)\| \leq \sqrt{N}\max_{1\leq i\leq N}\|\tilde{\delta}_i(t+1)\|. \tag{A.26}$$

It is easy to check that for any given $T > 0$, the closed-form of the solution is

$$\begin{aligned} \tilde{\delta}_1(t+1) =& \tilde{\delta}_1(0) = \frac{1}{\sqrt{N}}\mathbf{1}_{1\times N}(I_N - J)X(0) = 0, \\ \tilde{\delta}_i(t+1) =& \prod_{s=T}^{t}(1-\lambda_i a(s))\tilde{\delta}_i(T) \\ & + \gamma_i^T(I_N - J)\Sigma_{\mathcal{G}}\sum_{s=T}^{t}\Big(\prod_{r=s+1}^{t}(1-\lambda_i a(r))\Big)a(s)W(s). \end{aligned} \tag{A.27}$$

In the remainder of this proof, we assume $T \geq \max\{T_1, T_2, T_5\}$ with T_1, T_2 and T_5 are given, respectively, in Lemma A.3, the proof of Theorem 3.1, and Lemma A.6.

For the first item on the right hand side of (A.27), in view of Lemma A.2 and $1-x \leq e^{-x}$ when $x > 0$, we have $\prod_{s=T}^{t}(1-\lambda_i a(s)) \leq e^{-\lambda_i\sum_{s=T}^{t} a(s)} = o(t^{-1/2})$.

For the second item, noting that γ_i is orthogonal, it follows that

$$\begin{aligned} \|\gamma_i^T(I_N - J)\Sigma_{\mathcal{G}}\| \leq & \ (tr(\Sigma_{\mathcal{G}}^T(I_N - J)\Sigma_{\mathcal{G}}))^{1/2} \\ = & \ \Big(\frac{N-1}{N}\sum_{i,j} a_{ij}^2\Big)^{1/2}. \end{aligned} \tag{A.28}$$

By the properties of 2-norm, we have

$$\begin{aligned} & \Big\|\sum_{s=T}^{t}\Big(\prod_{r=s+1}^{t}(1-\lambda_i a(r))\Big)a(s)W(s)\Big\| \\ \leq & \ \sqrt{|\mathcal{E}|}\max_{(j,i)\in\mathcal{E}}\Big\|\sum_{s=T}^{t}\Big(\prod_{r=s+1}^{t}(1-\lambda_i a(r))\Big)a(s)\omega_{ji}(s)\Big\|. \end{aligned} \tag{A.29}$$

Hereinafter, we will mainly focus on the estimation of the convergence rate for the stochastic part $\sum_{s=T}^{t}\Big(\prod_{r=s+1}^{t}(1-\lambda_i a(r))\Big)$ $a(s)\omega_{ji}(s)$. By Lemma A.6 and the law of iterated logarithm of the martingale difference sequence [27], we have

$$\limsup_{t\to\infty}\frac{\big|\sum_{s=T}^{t}\Big(\prod_{r=T}^{s}(1-\lambda_i a(r))^{-1}\Big)a(s)\omega_{ji}(s)\big|}{(2S_t^{\lambda_i}E(\omega_{ji}^2(t))\log\log(S_t^{\lambda_i}E(\omega_{ji}^2(t))))^{\frac{1}{2}}} = 1 \ a.s. \tag{A.30}$$

Here, $S_t^{\lambda_i} = \sum_{s=T}^{t} a^2(s)\prod_{r=T}^{s}(1-\lambda_i a(r))^{-2}$. In view of $1-x \leq e^{-x}$ when $x > 0$, we have

$$\prod_{r=T}^{t}(1-\lambda_i a(r))^2 S_t^{\lambda_i} \leq \sum_{s=T}^{t} a^2(s)e^{-2\lambda_i\sum_{r=s+1}^{t} a(r)}. \tag{A.31}$$

By (A.30)–(A.31) and Lemma A.3, we have

$$\begin{aligned} & \limsup_{t\to\infty} t^{\frac{\gamma}{2}}(E(\omega_{ji}^2(t))\log\log(S_t^{\lambda_i}E(\omega_{ji}^2(t))))^{-\frac{1}{2}} \\ & \quad \times \Big\|\sum_{s=T}^{t}\Big(\prod_{r=s+1}^{t}(1-\lambda_i a(r))\Big)a(s)\omega_{ji}(s)\Big\| \\ \leq & \ \limsup_{t\to\infty}\Big(t^{\gamma}\prod_{r=T}^{t}(1-\lambda_i a(r))^2 S_t^{\lambda_i}\Big)^{\frac{1}{2}} \\ & \quad \times \frac{\big|\sum_{s=T}^{t}\Big(\prod_{r=T}^{s}(1-\lambda_i a(r))^{-1}\Big)a(s)\omega_{ji}(s)\big|}{(S_t^{\lambda_i}E(\omega_{ji}^2(t))\log\log(S_t^{\lambda_i}E(\omega_{ji}^2(t))))^{\frac{1}{2}}} \\ \leq & \ \beta/\sqrt{\lambda_i\alpha} \ a.s. \end{aligned} \tag{A.32}$$

Combining (A.28), (A.29), and (A.32), we have

$$\begin{aligned} & \limsup_{t\to\infty} t^{\frac{\gamma}{2}}(E(\omega_{ji}^2(t))\log\log(S_t^{\lambda_i}E(\omega_{ji}^2(t))))^{-\frac{1}{2}}\|\tilde{\delta}_i(t+1)\| \\ & \leq \beta\sqrt{\frac{|\mathcal{E}|(N-1)\sum_{i,j} a_{ij}^2}{\lambda_i\alpha N}} \ a.s. \end{aligned} \tag{A.33}$$

Note that $S_t^{\lambda_2} \leq S_t^{\lambda_i} \leq S_t^{\lambda_N}$ for $\lambda_2 \leq \lambda_i \leq \lambda_N$. Thus we have

$$\begin{aligned} & \limsup_{t\to\infty} t^{\frac{\gamma}{2}}(\Upsilon\log\log(S_t^{\lambda_N}\Upsilon))^{-\frac{1}{2}}\|\tilde{\delta}_i(t+1)\| \\ & \quad \leq \beta\sqrt{\frac{|\mathcal{E}|(N-1)\sum_{i,j} a_{ij}^2}{\lambda_2\alpha N}} \ a.s. \end{aligned} \tag{A.34}$$

And the conclusion in (i) follows by (A.26) and (A.34).

If $\gamma = 1$, by following the same step of the case $\gamma \in (0.5, 1)$, the conclusion in (ii) can be obtained similarly by Lemmas A.2, A.4 and A.6.

References

[1] T.C. Aysal, M.J. Coates, M.G. Rabbat, Distributed average consensus with dithered quantization, IEEE Trans. Signal Process. 56 (10) (2008) 4905–4918.

[2] M. Huang, J.H. Manton, Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior, SIAM J. Control Optim. 48 (1) (2009) 134–161.

[3] T. Li, J.-F. Zhang, Mean square average-consensus under measurement noises and fixed topologies: necessary and sufficient conditions, Automatica 45 (8) (2009) 1929–1936.

[4] S. Kar, J.M.F. Moura, Distributed consensus algorithms in sensor networks: quantized data and random link failures, IEEE Trans. Signal Process. 58 (3) (2010) 1383–1400.

[5] S. Liu, L. Xie, H. Zhang, Distributed consensus for multi-agent systems with delays and noises in transmission channels, Automatica 47 (2011) 920–934.

[6] Y. Cao, W. Yu, W. Ren, G. Chen, An overview of recent progress in the study of distributed multi-agent coordination, IEEE Trans. Ind. Inf. 9 (1) (2013) 427–438.

[7] L. Cheng, Z.-G. Hou, M. Tan, A mean square consensus protocol for linear multi-agent systems with communication noises and fixed topologies, IEEE Trans. Automat. Control 59 (1) (2014) 261–267.

[8] T. Li, J.-F. Zhang, Consensus conditions of multi-agent systems with time-varying topologies and stochastic communication noises, IEEE Trans. Automat. Control 55 (9) (2010) 2043–2057.

[9] M. Huang, J.H. Manton, Stochastic consensus seeking with noisy and directed inter-agent communication: fixed and randomly varying topologies, IEEE Trans. Automat. Control 55 (1) (2010) 235–241.

[10] M. Huang, Stochastic approximation for consensus: a new approach via ergodic backward products, IEEE Trans. Automat. Control 57 (12) (2012) 2994–3008.

[11] M. Huang, T. Li, J. Zhang, Stochastic approximation based consensus dynamics over Markovian networks, SIAM J. Control Optim. 53 (6) (2015) 3339–3363.

[12] S. Dasarathan, C. Tepedelenlioglu, M.K. Banavar, A. Spanias, Robust consensus in the presence of impulsive channel noise, IEEE Trans. Signal Process. 8 (6) (2015) 2118–2129.

[13] J. Xu, H. Zhang, L. Xie, Stochastic approximation approach for consensus and convergence rate analysis of multiagent system, IEEE Trans. Automat. Control 57 (12) (2012) 3163–3168.

[14] N. Amelina, A. Fradkov, Y. Jiang, D.J. Vergados, Approximate consensus in stochastic networks with application to load balancing, IEEE Trans. Inform. Theory 4 (61) (2015) 1739–1752.

[15] R. Olfati-Saber, R.M. Murray, Consensus problems in networks of agents with switching topology and time-delays, IEEE Trans. Automat. Control 49 (9) (2004) 1520–1533.

[16] A. Olshevsky, J.N. Tsitsiklis, Convergence speed in distributed consensus and averaging, SIAM J. Control Optim. 48 (1) (2009) 33–55.

[17] L. Xiao, S. Boyd, Fast linear iterations for distributed averaging, Systems Control Lett. 53 (1) (2004) 65–78.

[18] T.C. Aysal, B.N. Oreshkin, M.J. Coates, Accelerated distributed average consensus via localized node state prediction, IEEE Trans. Signal Process. 57 (4) (2009) 1563–1576.

[19] R.L.G. Cavalcante, B. Mulgrew, Adaptive filter algorithms for accelerated discrete-time consensus, IEEE Trans. Signal Process. 58 (3) (2010) 1049–1058.

[20] A. Olshevsky, J.N. Tsitsiklis, Degree fluctuations and the convergence time of consensus algorithms, IEEE Trans. Automat. Control 58 (10) (2013) 2626–2631.

[21] A. Olshevsky, Linear time average consensus on fixed graphs, in: Proceedings of the 5th IFAC Workshop on Distributed Estimation and Control in Networked Systems, University of Pennsylvania, Philadelphia, USA, 2015, September 10–11.

[22] S. Kar, J.M.F. Moura, Distributed consensus algorithms in sensor networks with imperfect communication: link failures and channel noise, IEEE Trans. Signal Process. 57 (1) (2009) 355–369.

[23] G. Yin, Y. Sun, L.Y. Wang, Asymptotic properties of consensus-type algorithms for networked systems with regime-switching topologies, Automatica 7 (47) (2011) 1366–1378.

[24] G. Yin, Q. Yuan, L.Y. Wang, Asynchronous stochastic approximation algorithms for networked systems: regime-switching topologies and multiscale structure, Multiscale Model. Simul. 11 (3) (2013) 813–839.

[25] H. Tang, T. Li, Continuous-time stochastic consensus: stochastic approximation and Kalman-Bucy filtering based protocols, Automatica 61 (2015) 146–155.

[26] G. Yin, L.Y. Wang, Y. Sun, D. Casbeer, R. Holsapple, D. Kingston, Asymptotic optimality for consensus-type stochastic approximation algorithms using iterate averaging, J. Control Theory Appl. 11 (1) (2013) 1–9.

[27] H. Chen, L. Guo, Identification and Stochastic Adaptive Control, Boston, MA: Birkhauser, 1991.