SIAM J. CONTROL OPTIM.
Vol. 48, No. 1, pp. 134–161

COORDINATION AND CONSENSUS OF NETWORKED AGENTS WITH NOISY MEASUREMENTS: STOCHASTIC ALGORITHMS AND ASYMPTOTIC BEHAVIOR*

MINYI HUANG† AND JONATHAN H. MANTON‡

Abstract. This paper considers the coordination and consensus of networked agents where each agent has noisy measurements of its neighbors' states. For consensus seeking, we propose stochastic approximation-type algorithms with a decreasing step size, and introduce the notions of mean square and strong consensus. Although the decreasing step size reduces the detrimental effect of the noise, it also reduces the ability of the algorithm to drive the individual states towards each other. The key technique is to ensure a trade-off for the decreasing rate of the step size. By following this strategy, we first develop a stochastic double array analysis in a two-agent model, which leads to both mean square and strong consensus, and extend the analysis to a class of well-studied symmetric models. Subsequently, we consider a general network topology, and introduce stochastic Lyapunov functions together with the so-called direction of invariance to establish mean square consensus. Finally, we apply the stochastic Lyapunov analysis to a leader following scenario.

Key words. multiagent systems, graphs, consensus problems, measurement noise, stochastic approximation, mean square convergence, almost sure convergence

AMS subject classifications. 93E03, 93E15, 94C15, 68R10

DOI. 10.1137/06067359X

1. Introduction.

The recent years have witnessed an enormous growth of research on the coordination and control of distributed multiagent systems, and specific topics appear in different forms such as swarming of honeybees, flocking of birds, migration of animals, synchronization of coupled oscillators, and formation of autonomous vehicles; see [48, 14, 17, 29, 43, 33] and the references therein. A common feature of these systems, which take diverse forms, is that the constituent agents need to maintain a certain coordination so as to cooperatively achieve a group objective, wherein the decision of individual agents is made with various constraints due to the distributed nature of the underlying system. The study of these multiagent models is crucial for understanding many complex phenomena related to animal behavior, and for designing distributed control systems.

For multiagent coordination, it is usually important to propagate shared information within the system by communication rules which may be supported by the specific interconnection structure between the agents. This is particularly important in cooperative control systems since they often operate in a dynamic environment, and the involved agents need to collectively acquire key information at the overall system level [38, 3]. In this context, of fundamental importance is the so-called consensus or agreement problem, where consensus means a condition where all the agents individually adjust their own value for an underlying quantity (e.g., a location as the destination of a robot team) so as to converge to a common value. For many prac-

*Received by the editors October 28, 2006; accepted for publication (in revised form) July 30, 2008; published electronically February 11, 2009. This work was partially supported by the Australian Research Council and the Australian National University.
http://www.siam.org/journals/sicon/48-1/67359.html
†Corresponding author. School of Mathematics and Statistics, Carleton University, Ottawa, ON K1S 5B6, Canada (mhuang@math.carleton.ca).
‡Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC, 3010, Australia (j.manton@ieee.org).

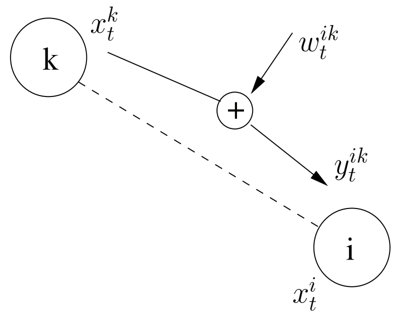

FIG. 1. *Measurement with additive noise* w_t^{ik}.

tical situations, the chief objective is to agree on the same state; the actual state is of secondary importance. In view of primarily being required to converge, one might suggest to simply set the agents' states to any fixed state. In reality, however, such a consensus protocol is trivial and less interesting; its more serious limitation is that this protocol is overly sensitive to small relative errors when the individual states initially have been very close to each other. For these reasons, in the literature, almost all consensus algorithms are constructed based on averaging rules, and this leads to good dynamic properties (such as good transient behavior and convergence) [23, 50, 6]. We mention that there has been a long history of research on consensus problems due to the broad connections of this subject with a wide range of disciplines including statistical decision theory, management science, distributed computing, ad hoc networks, biology [49, 20, 10, 18, 28, 26, 48], and the quickly developing area of multiagent control systems [3, 14, 17, 29, 33, 34, 43]. A comprehensive survey on consensus problems in multiagent coordination can be found in [38].

In the context of coordinating spatially distributed agents, a basic consensus model consists of a time-invariant network in which each agent updates its state by forming a convex combination of the states of its neighbors and itself [23, 6, 50], such that the iterates of all individual states converge to a common value. Starting from this formulation, many generalizations are possible. A variety of consensus algorithms has been developed to deal with delayed measurements [5, 31, 34], dynamic topologies [34], or unreliable (on/off) communication links (see the survey [38]). For convergence analysis, stochastic matrix analysis is an important tool [23], and in models with time-dependent communications, set-valued Lyapunov theory is useful [32].

In this paper, we are interested in consensus seeking in an uncertain environment where each agent can obtain only noisy measurements of the states of its neighbors; see Figure 1 for illustration. Such modeling reflects many practical properties in distributed networks. For instance, the interagent information exchange may involve the use of sensors, quantization [36, 37], and wireless fading channels, which makes it unlikely to have exact state exchange. We note that most previous research has used noise-free state iteration by assuming exact data exchange between the agents, with only a few exceptions (see, e.g., [51, 39, 9]). A least mean square optimization method was used in [51] to choose the constant coefficients in the averaging rule with additive noises so that the long term consensus error is minimized. In a continuous time consensus model [15], deterministic disturbances were included in the dynamics. In [9], multiplicative noises were introduced to model logarithmic quantization error. In [21, 42], convergence results were obtained for random graph based consensus problems, and [21] used an approach of stochastic stability. In earlier work [7, 46, 47], convergence of consensus problems was studied in a stochastic setting, but the interagent exchange of random messages was assumed to be error-free. In particular, Tsitsiklis, Bertsekas, and Athans [47] obtained consensus results via asynchronous stochastic gradient based algorithms for a group of agents minimizing their common cost function.

In models with noisy measurements, one may still construct an averaging rule with a constant coefficient matrix. However, the resulting evolution of the state vector dramatically differs from the noise-free case, leading to divergence. The reason is that the noise causes a steady drift of the agents' states during the iterates, which in turn prevents generating a stable group behavior.

To deal with the measurement noise, we propose a stochastic approximation-type algorithm with the key feature of a decreasing step size. The algorithm has a gradient descent interpretation. Our formulation differs from [47] since in the averaging rule of the latter, the exogenous term, which may be interpreted as a local noisy gradient of the agents' common cost, is assigned a controlled step size while the weights for the exact messages received from other agents are maintained to be above a constant level; such a separability structure enables the authors in [47] to obtain consensus with a sufficiently small constant step size for the gradient term, or with only an upper bound for the deceasing rate of the step size. In contrast, in our model the signal received from other agents is corrupted by additive noise (see Figure 1), and consequently in selecting the step size, it is critical to maintain a trade-off in attenuating the noise to prevent long term fluctuations and meanwhile ensuring a suitable stabilizing capability of the recursion so as to drive the individual states towards each other. To achieve this objective, the step size must be decreased neither too slowly, nor too quickly. It turns out, for proving mean square consensus via stochastic Lyapunov functions, that we may simply use the standard step size condition in traditional stochastic approximation algorithms. But in the stochastic double array analysis, some mild lower and upper bound conditions will be imposed on the step size.

We begin by analyzing a two-agent model. As it turns out, this simple model provides a rich structure for developing convergence analysis and motivates the solution to more general models. In this setup, the key technique is the stochastic double array analysis [45, 12]. Next, we extend the analysis to a class of symmetric models. In fact, many symmetric models have arisen in practical applications including platoons of vehicles, robot teams, unicycle pursuit models [30, 29], cooperative sensor network deployment for tracking [1] or sampling [25], and consensus problems [9]. Subsequently, to deal with a general network topology, we develop a stochastic Lyapunov analysis, and convergence is established under a connectivity condition for the associated undirected graph.

The paper is organized as follows. In section 2 we formulate the consensus problem in the setting of directed graphs and propose the consensus algorithm. Section 3 establishes convergence results in a two-agent model, and the analysis is extended to models with circulant symmetry in section 4. We develop stochastic Lyapunov analysis in section 5 and apply it to leader following in section 6. Section 7 presents numerical simulations, and section 8 concludes the paper.

2. Formulation of the stochastic consensus problem. We begin by considering directed graphs for modeling the spatial distribution of n agents. A directed graph (or digraph) $G = (\mathcal{N}, \mathcal{E})$ consists of a set of nodes $\mathcal{N} = \{1, 2, \dots, n\}$ and a set of edges $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$. An edge in G is denoted as an ordered pair (i, j), where

$i \neq j$ (so there is no edge between a node and itself) and i, j are called the initial and terminal node, respectively. A path (from i_1 to i_l) in G consists of a sequence of nodes $i_1, i_2, \ldots, i_l$, $l \geq 2$, such that $(i_k, i_{k+1}) \in \mathcal{E}$ for all $1 \leq k \leq l-1$. The digraph G is said to be strongly connected if for any two distinct nodes i and j, there exist a path from i to j and also a path from j to i.

For convenience of exposition, we often refer to node i as agent A_i. The two names, agent and node, will be used interchangeably. Agent A_k (resp., node k) is a neighbor of A_i (resp., node i) if $(k, i) \in \mathcal{E}$, where $k \neq i$. Denote the neighbors of node i by $\mathcal{N}_i \subset \mathcal{N}$. Note that any undirected graph[1] can be converted into a directed graph simply by splitting each edge in the former into two edges, one in each direction.

For agent A_i, let $x_t^i \in \mathbb{R}$ be its state at time $t \in \mathbb{Z}^+ = \{0, 1, 2, \ldots\}$. Denote the state vector $x_t = [x_t^1, \ldots, x_t^n]^T$. For each $i \in \mathcal{N}$, agent A_i receives noisy measurements of the states of its neighbors. Denote the resulting measurement by A_i of A_k's state by

$$y_t^{ik} = x_t^k + w_t^{ik}, \qquad t \in \mathbb{Z}^+, \quad k \in \mathcal{N}_i, \tag{1}$$

where $w_t^{ik} \in \mathbb{R}$ is the additive noise; see Figure 1. The underlying probability space is denoted by $(\Omega, \mathcal{F}, P)$. We shall call y_t^{ik} the observation of the state of A_k obtained by A_i, and we assume each A_i knows its own state x_t^i exactly. The additive noise w_t^{ik} in (1) reflects unreliable information exchange during interagent sensing and communication; see, e.g., [39, 2, 41] for related modeling.

(A1) The noises $\{w_t^{ik}, t \in \mathbb{Z}^+, i \in \mathcal{N}, k \in \mathcal{N}_i\}$ are independent and identically distributed (i.i.d.) with respect to the indices i, k, t and each w_t^{ik} has zero mean and variance $Q \geq 0$. The noises are independent of the initial state vector x_0 and $E|x_0|^2 < \infty$.

Condition (A1) means that the noises are i.i.d. with respect to both space (associated with neighboring agents) and time. We will begin with our analysis based on the above assumption for simplicity.

The state of each agent is updated by

$$x_{t+1}^i = (1 - a_t) x_t^i + \frac{a_t}{|\mathcal{N}_i|} \sum_{k \in \mathcal{N}_i} y_t^{ik}, \qquad t \in \mathbb{Z}^+, \tag{2}$$

where $i \in \mathcal{N}$ and $a_t \in [0, 1]$ is the step size. This gives a weighted averaging rule in that the right-hand side is a convex combination of the agent's state and its $|\mathcal{N}_i|$ observations, where $|S|$ denotes the cardinality of a set S. The objective for the consensus problem is to select the sequence $\{a_t, t \geq 0\}$ so that the n individual states x_t^i, $i \in \mathcal{N}$, converge to a common limit in a certain sense.

To get some insight into algorithm (2), we rewrite it in the form

$$x_{t+1}^i = x_t^i + a_t (m_t^i - x_t^i), \tag{3}$$

where $m_t^i = (1/|\mathcal{N}_i|) \sum_{k \in \mathcal{N}_i} y_t^{ik}$. The structure of (3) is very similar to the recursion used in classical stochastic approximation algorithms in that $m_t^i - x_t^i$ provides a correction term controlled by the step size a_t. Indeed, by introducing a suitable local potential function, $m_t^i - x_t^i$ may be interpreted as the noisy measurement of a scaled negative gradient of the local potential along the direction x_t^i. A more detailed discussion will be presented in section 5 when developing the stochastic Lyapunov

[1] The edge in an undirected graph is denoted as an unordered pair.

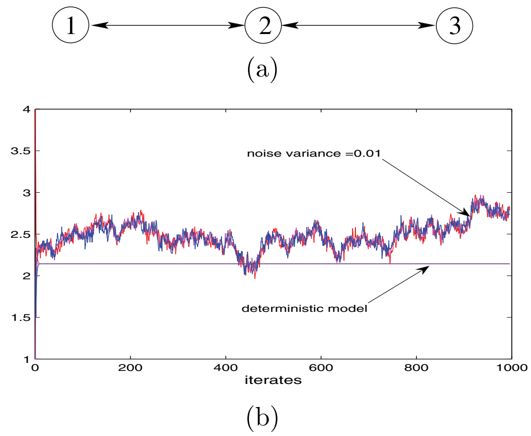

FIG. 2. (a) *The three nodes.* (b) *In the noise-free case, the states of the nodes quickly converge to the same constant level* ≈ 2.143. *Under Gaussian measurement noises with variance* $\sigma^2 = 0.01$, *the three state trajectories have large fluctuations.*

analysis. Due to the noise contained in $\{m_t^i, t \geq 0\}$, each state x_t^i will fluctuate randomly. These fluctuations will not die off if a_t does not converge to 0. For illustration, we introduce an example as follows.

EXAMPLE 1. *Consider a strongly connected digraph with* $\mathcal{N} = \{1, 2, 3\}$, *as in Figure* 2(a), *where* $\mathcal{N}_1 = \{2\}$, $\mathcal{N}_2 = \{1, 3\}$, *and* $\mathcal{N}_3 = \{2\}$. *We follow the measurement model* (1), *and the states are updated by* $x_{t+1}^1 = (x_t^1 + y_t^{12})/2$, $x_{t+1}^2 = (x_t^2 + y_t^{21} + y_t^{23})/3$, *and* $x_{t+1}^3 = (x_2^3 + y_t^{32})/2$, $t \geq 0$. *The i.i.d. Gaussian noises* w_t^{ik} *satisfy* (A1) *with variance* $\sigma^2 = 0.01$.

The simulation for Example 1 takes the initial condition $[x_0^1, x_0^2, x_0^3] = [4, 1, 2]$. For the noise-free case, we change the state update rule in Example 1 by replacing each y_t^{ik} by x_t^k, which results in a standard rule in the literature; see, e.g., [23]. Figure 2(b) shows that measurement noises cause a dramatic loss of convergence. In fact, by recasting to the form (2), the algorithm in Example 1 essentially takes the step size $a^{(i)} = |\mathcal{N}_i|/(|\mathcal{N}_i| + 1)$ for node i to give equal weights $1/(|\mathcal{N}_i| + 1)$ to $|\mathcal{N}_i| + 1$ nodes; for instance, we may rewrite $x_{t+1}^2 = (x_t^2 + y_t^{21} + y_t^{23})/3$ as $x_{t+1}^2 = x_t^2 + a^{(2)}[(y_t^{21} + y_t^{23})/2 - x_t^2]$, where $a^{(2)} = 2/3$.

With the aim of obtaining a stable behavior for the agents, we make the following assumption.

(A2) (i) The sequence $\{a_t, t \geq 0\}$ satisfies $a_t \in [0, 1]$, and (ii) there exists $T_0 \geq 1$ such that

$$\alpha t^{-\gamma} \leq a_t \leq \beta t^{-\gamma} \tag{4}$$

for all $t \geq T_0$, where $\gamma \in (0.5, 1]$ and $0 < \alpha \leq \beta < \infty$.

By requiring $a_t > \alpha t^{-\gamma}$ for $t \geq T_0$ with a suitable T_0, we may take large values for α while still ensuring $a_t \in [0, 1]$, $t \geq T_0$. This offers more flexibility in selecting the step size sequence. Here $\{a_t, t < T_0\}$ may be chosen freely as long as $a_t \in [0, 1]$; this resulting algorithm gives a convex combination at all times in the averaging rule as in conventional consensus algorithms. The parameters $T_0, \alpha, \beta, \gamma$ will be treated as fixed constants associated with $\{a_t, t \geq 0\}$. Note that (A2) implies the following weaker condition.

(A2′) (i) The sequence $\{a_t, t \geq 0\}$ satisfies $a_t \in [0, 1]$, and (ii) $\sum_{t=0}^{\infty} a_t = \infty$, $\sum_{t=0}^{\infty} a_t^2 < \infty$.

Notice that (A2′)(ii) is a typical condition used in classical stochastic approximation theory [11, 24]. In the subsequent sections, the double array analysis will be developed based on the slightly stronger assumption (A2) while (A2′) will be used for the stochastic Lyapunov analysis.

The vanishing rate of $\{a_t, t \geq 0\}$ is crucial for consensus. When $a_t \to 0$ in (2), the signal x_t^k (contained in y_t^{ik}), as the state of A_k, is attenuated together with the noise. Hence, a_t cannot decrease too fast since, otherwise, the agents may prematurely converge to different individual limits.

Since the averaging rule (2) can be considered a stochastic approximation algorithm [27, 4], we may apply the standard method of analysis to it; namely, we can average out the noise component in (2) to derive an associated ordinary differential equation (ODE) system

$$\frac{dx^i(s)}{ds} = (1/|\mathcal{N}_i|) \sum_{k \in \mathcal{N}_i} x^k(s) - x^i(s), \quad s \geq 0, \quad i \in \mathcal{N}. \tag{5}$$

The important feature of the ODE system (5) is that it has an equilibrium set as a linear subspace of $\mathbb{R}^n$, instead of a singleton. This indicates more uncertain asymptotic behavior in the state evolution of the stochastic consensus algorithm due to the lack of a *single* equilibrium point generating the attracting effect, and is in contrast to typical stochastic approximation algorithms where the associated ODE usually has a single equilibrium, at least locally.

We introduce some definitions to characterize the asymptotic behavior of the agents.

Definition 2 (weak consensus). *The agents are said to reach weak consensus if* $E|x_t^i|^2 < \infty$, $t \geq 0$, $i \in \mathcal{N}$, *and* $\lim_{t\to\infty} E|x_t^i - x_t^j|^2 = 0$ *for all distinct* $i, j \in \mathcal{N}$.

Definition 3 (mean square consensus). *The agents are said to reach mean square consensus if* $E|x_t^i|^2 < \infty$, $t \geq 0$, $i \in \mathcal{N}$, *and there exists a random variable* x^* *such that* $\lim_{t\to\infty} E|x_t^i - x^*|^2 = 0$ *for all* $i \in \mathcal{N}$.

Definition 4 (strong consensus). *The agents are said to reach strong consensus if there exists a random variable* x^* *such that with probability* 1 *(w.p.1) and for all* $i \in \mathcal{N}$, $\lim_{t\to\infty} x_t^i = x^*$.

It is obvious that mean square consensus implies weak consensus. If a sequence converges w.p.1, we also say it converges almost surely (a.s.). Note that for both mean square and strong consensus, the states x_t^i, $i \in \mathcal{N}$, must converge to a common limit, which may depend on the initial states, the noise sequence, and the consensus algorithm itself.

2.1. The generalization to vector states. It is straightforward to generalize the results of this paper to the case of vector individual states $\mathbf{x}_t^k \in \mathbb{R}^d$, where $d > 1$, and (1)–(2) may be extended to the vector case by taking a vector noise term. For the vector version of (2), we see that the d components in $\mathbf{x}_t^k$ are decoupled during iteration and may be treated separately. Throughout this paper, we consider only scalar individual states.

3. Convergence in a two-agent model. We begin by analyzing a two-agent model, which will provide interesting insight into understanding consensus seeking in a noisy environment. The techniques developed for such a system will provide motivation for analyzing more general models. The rich structure associated with this seemingly simple model well justifies a careful investigation.

3.1. Mean square consensus. Let (1)–(2) be applied by the two agents where $\mathcal{N} = \{1,2\}$. In the subsequent analysis, a key step is to examine the evolution of the difference $\xi_t = x_t^1 - x_t^2$ between the two states. We notice the relation

$$\xi_{t+1} = (1-2a_t)\xi_t + a_t v_t, \qquad t \geq 0, \tag{6}$$

where $v_t = w_t^{12} - w_t^{21}$. By inequality (4), we may find an integer $T_1 > T_0$ such that

$$1 - 2\alpha t^{-\gamma} \geq 1 - 2a_t > 0 \qquad \text{for all} \quad t \geq T_1. \tag{7}$$

In the estimate below, we start with T_1 as the initial time. It follows from (6) that

$$\xi_{t+1} = \prod_{i=T_1}^{t}(1-\bar{a}_i)\xi_{T_1} + \sum_{k=T_1}^{t}\left[\prod_{i=k+1}^{t}(1-\bar{a}_i)\right]a_k v_k, \qquad t \geq T_1, \tag{8}$$

where $\bar{a}_t = 2a_t$. Define

$$\Pi_{l,k} = \prod_{i=k+1}^{l}(1-\bar{a}_i)a_k, \tag{9}$$

where $l > k \geq T_1$. By convention, $\Pi_{k,k} = a_k$.

Lemma 5. *Let $\Pi_{l,k}$ be defined by* (9) *with $k \leq l$ and assume* (A2).

(i) *If $\gamma = 1$, we have*

$$\Pi_{l,k} \leq \exp\left\{-2\alpha\sum_{t=k+1}^{l}t^{-1}\right\}\frac{\beta}{k} \leq \frac{\beta(k+1)^{2\alpha}}{k(l+1)^{2\alpha}}. \tag{10}$$

(ii) *If $1/2 < \gamma < 1$, we have*

$$\Pi_{l,k} \leq \exp\left\{\frac{-2\alpha}{1-\gamma}[(l+1)^{1-\gamma} - (k+1)^{1-\gamma}]\right\}\frac{\beta}{k^\gamma}. \tag{11}$$

Proof. First, for the case $k < l$, it is obvious that

$$\Pi_{l,k} \leq \left(1 - \frac{2\alpha}{l^\gamma}\right)\cdots\left(1 - \frac{2\alpha}{(k+1)^\gamma}\right)\frac{\beta}{k^\gamma}. \tag{12}$$

By the fact $\ln(1-x) < -x$ for all $x \in (0,1)$, it follows that

$$\left(1 - \frac{2\alpha}{l^\gamma}\right)\cdots\left(1 - \frac{2\alpha}{(k+1)^\gamma}\right) \leq \exp\left\{-2\alpha\sum_{t=k+1}^{l}t^{-\gamma}\right\}. \tag{13}$$

By (12)–(13), we get (i) and (ii) for $k < l$. Clearly, (i) and (ii) hold for $k = l$. ∎

Let $\{c(t), t \geq t_0\}$ and $\{h(t), t \geq t_0\}$ be two sequences of real numbers indexed by integers $t \geq t_0$, and $h(t) > 0$ for all $t \geq t_0$. Denote $c(t) = O(h(t))$ (resp., $c(t) = o(h(t))$) if $\overline{\lim}_{t\to\infty}|c(t)|/h(t) \leq C_d < \infty$ (resp., $\lim_{t\to\infty}|c(t)|/h(t) = 0$). Here C_d is called a dominance constant in the relation $c(t) = O(h(t))$. In practice, it is desirable to take a value for C_d as small as possible.

LEMMA 6. *Under* (A2), *we have the upper bound estimate* (i) *if* $\gamma = 1$,

$$\sum_{k=T_1}^{t} \Pi_{t,k}^2 = \begin{cases} O(t^{-4\alpha}) & \text{if } \ 0 < \alpha < 1/4, \\ O(t^{-1}\ln t) & \text{if } \ \alpha = 1/4, \\ O(t^{-1}) & \text{if } \ \alpha > 1/4, \end{cases} \tag{14}$$

where T_1 *is specified in* (7), *and* (ii) *if* $1/2 < \gamma < 1$,

$$\sum_{k=T_1}^{t} \Pi_{t,k}^2 = O(t^{-\gamma}). \tag{15}$$

Proof. See the appendix. □

Remark. We give some discussions on estimating the dominance constant C_d for Lemma 6. For (14), when $\alpha \neq 1/4$ but is close to 1/4 from the left (resp., right), our estimation method shows that we need to take a large C_d associated with $O(t^{-4\alpha})$ (resp., t^{-1}). For the case $\alpha = 1/4$ in (14), we may take $C_d = \beta^2$. For (15), we take $C_d = 4\alpha$, regardless of the value of $\gamma \in (1/2, 1)$.

COROLLARY 7. *Let* $\{\tilde{a}_t, t \geq 1\}$ *be a sequence such that* (i) $\tilde{a}_t \in [0,1]$ *and* (ii) *there exists* $\gamma_0 \in (0, 1/2)$ *such that* $\tilde{\alpha} t^{-\gamma_0} \leq \tilde{a}_t \leq \tilde{\beta} t^{-\gamma_0}$, *where* $\tilde{\alpha} > 0$. *Denote* $\tilde{\Pi}_{l,k} = \prod_{i=k+1}^{l} (1 - \tilde{a}_i)\tilde{a}_k$, $l \geq k \geq 1$. *Then for any fixed* $\tilde{T}_1 \geq 1$, $\sum_{k=\tilde{T}_1}^{t} \tilde{\Pi}_{t,k}^2 = O(t^{-\gamma_0})$.

Proof. First, (11) is still valid after replacing γ (resp., $\Pi_{l,k}$) by γ_0 (resp., $\tilde{\Pi}_{l,k}$). The argument in proving (15) can be repeated when γ is replaced by γ_0, which leads to the corollary. □

THEOREM 8. *Suppose* (A1)–(A2) *hold for the system of two agents, and* x_t^1, x_t^2 *are updated according to algorithm* (2). *Then there exists a random variable* x^* *such that* $\lim_{t\to\infty} E|x_t^i - x^*|^2 = 0$ *for* $i = 1, 2$, *which implies mean square consensus.*

Proof. First, denote $z_t = (x_t^1 + x_t^2)/2$ and $\tilde{w}_t = (w_t^{12} + w_t^{21})/2$ for $t \geq 0$. It is easy to check that

$$z_{t+1} = z_t + a_t \tilde{w}_t, \qquad t \geq 0, \tag{16}$$

which leads to $z_{t+1} = z_0 + \sum_{k=0}^{t} a_k \tilde{w}_k$. Since $\sum_{t=0}^{\infty} a_t^2 < \infty$, there exists a random variable z^* such that $\lim_{t\to\infty} E|z_t - z^*|^2 = 0$.

Now we estimate $\xi_t = x_t^1 - x_t^2$. We see that

$$E\xi_{t+1}^2 \leq 2 \left(E\xi_{T_1}^2 \prod_{k=T_1}^{t} |1 - 2a_k|^2 + \sup_{k \geq T_1} Ev_k^2 \times \sum_{k=T_1}^{t} \Pi_{t,k}^2 \right).$$

By Lemma 6, $\lim_{t\to\infty} E\xi_{t+1}^2 = 0$. Then mean square consensus follows easily. □

The i.i.d. noise assumption in Theorem 8 may be relaxed to independent noises with zero mean and uniformly bounded variances.

We use this two-agent model to illustrate the importance of a trade-off in the decreasing rate of a_t. To avoid triviality, assume the noise variance $Q > 0$ in (A1).

First, let $\gamma_0 \in (0, 1/2)$ and $a_0 = 0$, $a_t = t^{-\gamma_0}$ for $t \geq 1$, which decreases more slowly than in (4). By (16), it follows that $\lim_{t\to\infty} E|z_t|^2 = \infty$. Let ξ_t be given by (6). By Corollary 7, we can show $\lim_{t\to\infty} \xi_t^2 = 0$. So we conclude that this too-slowly-decreasing step size causes divergence of x_t^1 and x_t^2 due to inadequately attenuated noise, although they reach weak consensus since $\lim_{t\to\infty} \xi_t^2 = 0$.

Next, we take $\gamma_1 > 1$ and $a_0 = 0$, $a_t = t^{-\gamma_1}$ for $t \geq 1$, which decreases faster than in (4). Then there exists a random variable z^* such that $\lim_{t\to\infty} E|z_t - z^*|^2 = 0$.

Furthermore, by the fact $\prod_{i=2}^{\infty}(1-2\bar{a}_i) > 0$, we obtain from (8) that there exists a random variable ξ^* such that $\lim_{t\to\infty} E|\xi_t - \xi^*|^2 = 0$ and $E|\xi^*|^2 > 0$. So x_t^1 and x_t^2 both converge in mean square. But the state gap ξ_t cannot be asymptotically eliminated due to the excessive loss of the stabilizing capability, associated with the homogenous part of (6), when a_t decreases too quickly.

3.2. Strong consensus. So far we have shown that the two states converge in mean square to the same limit. It is well known that in classical stochastic approximation theory [11, 24], similarly structured algorithms have sample path convergence properties under reasonable conditions. It is tempting to analyze sample path behavior in this consensus context. The analysis below moves towards this objective. The following lemma is instrumental.

LEMMA 9 (see [45]). *Let $\{w, w_t, t \ge 1\}$ be i.i.d. real-valued random variables with zero mean, and $\{a_{ki}, 1 \le i \le l_k \uparrow \infty, k \ge 1\}$ a double array of constants. Assume* (i) $\max_{1\le i\le l_k} |a_{ki}| = O((l_k^{1/p}\log k)^{-1})$, $0 < p \le 2$, *and* $\log l_k = o(\log^2 k)$, *and* (ii) $E|w|^p < \infty$. *Then* $\lim_{k\to\infty} \sum_{i=1}^{l_k} a_{ki} w_i = 0$ *a.s.*

This lemma is an immediate consequence of Theorem 4 and Corollary 3 in [45, pp. 331 and 340], which deal with the sum of random variables with weights in a double array.

Now we need to estimate the magnitude of the individual terms $\Pi_{t,k}$. Note that for each $t > T_1$, $\Pi_{t,k}$ is defined for k starting from T_1 up to t. Hereafter, for notational brevity, we make a convention about notation by setting $\Pi_{t,k} \equiv 0$ for $1 \le k < T_1$ when $t \ge T_1$, and $\Pi_{t,k} \equiv 0$ for $1 \le k \le t$ when $1 \le t < T_1$. After this extension, all the entries $\Pi_{t,k}$ constitute a triangular array.

LEMMA 10. *For case* (i) *with* $\gamma = 1$, *under* (A2) *we have*

$$\sup_{1\le k\le t} \Pi_{t,k} = \begin{cases} O(t^{-2\alpha}) & \text{if} \quad 0 < \alpha < 1/2, \\ O(t^{-1}) & \text{if} \quad \alpha \ge 1/2, \end{cases} \tag{17}$$

and for case (ii) *with* $1/2 < \gamma < 1$, *we have* $\sup_{1\le k\le t} \Pi_{t,k} = O(t^{-\gamma})$.

Proof. By use of (10), it is easy to obtain the bound for case (i). Now we give the proof for case (ii). By Lemma 5(ii), it follows that

$$\Pi_{t,k} \le e^{-\delta(t+1)^{1-\gamma}} e^{\delta(k+1)^{1-\gamma}} \frac{\beta}{k^\gamma} \le e^{-\delta(t+1)^{1-\gamma}} \max_{1\le k\le t} e^{\delta(k+1)^{1-\gamma}} \frac{\beta}{k^\gamma},$$

where $\delta = 2\alpha/(1-\gamma)$. Denote the function $f(s) = e^{\delta(s+1)^{1-\gamma}}(\beta/s^\gamma)$, where the real number $s \in [1,\infty)$. By calculating the derivative $f'(s)$, it can be shown that for all $s \ge s_0 = [1 + \frac{\gamma}{\delta(1-\gamma)}]^{1/(1-\gamma)}$, $f'(s) > 0$. Hence there exists $c_0 > 0$ independent of t such that

$$\max_{1\le k\le t} e^{\delta(k+1)^{1-\gamma}} \frac{\beta}{k^\gamma} \le \max_{s\in[1,t]} f(s) \le c_0 \vee \left(e^{\delta(t+1)^{1-\gamma}} \frac{\beta}{t^\gamma} \right),$$

which implies that $\sup_{1\le k\le t} \Pi_{t,k} = O(t^{-\gamma})$. This completes the proof. □

THEOREM 11. *Assume all conditions in Theorem* 8 *hold and, in addition,* $\alpha > 1/4$ *in the case* $\gamma = 1$. *Then we have* (a) z_t *converges a.s.,* (b) $\lim_{t\to\infty} \xi_t = 0$ *a.s., and* (c) *the two sequences* $\{x_t^1, t \ge 0\}$ *and* $\{x_t^2, t \ge 0\}$ *converge to the same limit a.s., which implies strong consensus.*

Proof. Recall that $z_{t+1} = z_0 + \sum_{k=0}^{t} a_k \tilde{w}_k$ for $t \ge 0$, where $\tilde{w}_t = (w_t^{12} + w_t^{21})/2$. Since $\{\tilde{w}_k, k \ge 0\}$ is a sequence of independent random variables with zero mean

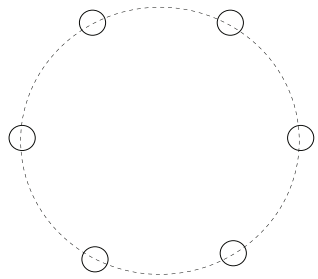

FIG. 3. *A symmetric ring network where each agent has two neighbors.*

and satisfies $\sum_{k=0}^{\infty} E|a_k\tilde{w}_k|^2 \le \sup_k E|\tilde{w}_k|^2(\sum_{k=0}^{\infty} a_k^2) < \infty$, by the Khintchine–Kolmogorov convergence theorem [13], $\sum_{k=0}^{\infty} a_k\tilde{w}_k$ converges a.s. Hence assertion (a) holds.

Now we prove (b). By Lemma 10 we have

$$\sup_{1\le k\le t} \Pi_{t,k} = O((t^{1/2}\log t)^{-1}), \tag{18}$$

whenever $\alpha > 0$ (resp., $\alpha > 1/4$) in the case $1/2 < \gamma < 1$ (resp., $\gamma = 1$). To apply Lemma 9, we take $l_k = k$ and $p = 2$, which combined with (18) yields $\lim_{t\to\infty}\sum_{k=1}^{t}\Pi_{t,k}v_k = 0$, a.s. Hence $\lim_{t\to\infty}\xi_t = 0$ a.s. by (8), and (b) follows. Assertion (c) immediately follows from (a) and (b). $\square$

The requirement $\alpha > 1/4$, associated with $\gamma = 1$, is a mild condition, and from an algorithmic point of view, it is not an essential restriction since in applications $\{a_t, t \ge 0\}$ is a sequence to be designed. In fact, by a slightly more complicated procedure, the restriction $\alpha > 1/4$ can be removed; see the more recent work [22].

4. Models with symmetric structures. We continue to consider models where the neighboring relation for the n agents displays a certain symmetry. A simple example is shown in Figure 3 with ring-coupled agents each having two neighbors.

We specify the associated digraph as follows. First, the n nodes are listed by the order $1, 2, \ldots, n$. The ith node has a neighbor set $\mathcal{N}_i$ listed as $(\alpha_1^i, \ldots, \alpha_K^i)$ as a subset of $\{1, \ldots, n\}$. The constant $K \ge 1$ denotes the number of neighbors, which is the same for all agents. Then the $(i+1)$th node's neighbors are given by $(\alpha_1^i+1, \ldots, \alpha_K^i+1)$. In other words, by incrementing each α_k^i (associated with A_i) by one, where $1 \le k \le K$, we obtain the neighbor set for node $i+1$, and after a total of n steps, we retrieve node i and its neighbors $\mathcal{N}_i$. In fact, the underlying digraph may be realized by arranging the n nodes sequentially on a ring and adding the edges accordingly. For this reason, we term the fulfillment of the above incrementing rule as the circulant invariance property of the digraph. In this section, if an index (e.g., $\alpha_k^i + 1$) for a node or agent exceeds n, we identify it as an integer between 1 and n by taking mod(n).

Notice that the above symmetry assumption does not ensure the strong connectivity of the digraph. For illustration, consider a digraph with the set of nodes $\mathcal{N} = \mathcal{S}_1 \cup \mathcal{S}_2$, where $\mathcal{S}_1 = \{1, 3, 5\}$ and $\mathcal{S}_2 = \{2, 4, 6\}$. All nodes inside each $\mathcal{S}_i$, $i = 1, 2$, are neighbors to each other, but there exists no edge between two nodes with one in $\mathcal{S}_1$ and the other in $\mathcal{S}_2$. This digraph has the circulant invariance property without connectivity. Throughout this section, we make the following assumption.

(A3) The digraph $G = (\mathcal{N}, \mathcal{E})$ has the circulant invariance property and strong connectivity.

Define the centroid of the state configuration $(x_t^1, \dots, x_t^n)$ as $z_t = (1/n)\sum_{i=1}^n x_t^i$. Under (A3), it is easy to show that z_t satisfies

$$z_{t+1} = z_t + (a_t/(nK)) \sum_{i\in\mathcal{N}} \sum_{k\in\mathcal{N}_i} w_t^{ik}, \qquad t \geq 0. \tag{19}$$

LEMMA 12. *Under* (A1)–(A3), *the sequence* $\{z_t, t \geq 0\}$ *converges in mean square and a.s.*

Proof. The lemma may be proved by the same method as in analyzing $\{z_t, t \geq 0\}$ in Theorems 8 and 11 for the two-agent model, and the details are omitted. ∎

We further denote the difference between x_t^{i+1} and x_t^i by

$$\xi_t^i = x_t^{i+1} - x_t^i, \qquad 1 \leq i \leq n. \tag{20}$$

Note that i and $i+1$ are two consecutively labelled agents, unnecessarily being neighbors to each other. By our convention, x_t^{n+1} is identified as x_t^1. Thus $\xi_t^n = x_t^1 - x_t^n$. The variables ξ_t^i, $1 \leq i \leq n$, are not linearly independent. Recall that $|\mathcal{N}_i| = K$ for all $i \in \mathcal{N}$. Specializing algorithm (2) to the model of this section, we have

$$x_{t+1}^i = (1-a_t)x_t^i + (a_t/K) \sum_{k\in\mathcal{N}_i} (x_t^k + w_t^{ik}) \tag{21}$$

for each $i \in \mathcal{N}$, and

$$\begin{aligned} x_{t+1}^{i+1} &= (1-a_t)x_t^{i+1} + (a_t/K) \sum_{k\in\mathcal{N}_{i+1}} (x_t^k + w_t^{i+1,k}) \\ &= (1-a_t)x_t^{i+1} + (a_t/K) \sum_{k\in\mathcal{N}_i} (x_t^{k+1} + w_t^{i+1,k+1}), \end{aligned} \tag{22}$$

where we obtain (22) by use of the circulant invariance of the neighboring relation.

By subtracting both sides of (22) by (21), we get the dynamics

$$\xi_{t+1}^i = (1-a_t)\xi_t^i + (a_t/K)\sum_{k\in\mathcal{N}_i} \xi_t^k + (a_t/K)\tilde{w}_t^i, \qquad i \in \mathcal{N}, \tag{23}$$

where

$$\tilde{w}_t^i = \sum_{k\in\mathcal{N}_i} \tilde{w}_t^{i,k}, \qquad \tilde{w}_t^{i,k} = w_t^{i+1,k+1} - w_t^{i,k} \tag{24}$$

with $k \in \mathcal{N}_i$ for $\tilde{w}_t^{i,k}$.

LEMMA 13. *Let* ξ_t^i *and* $\tilde{w}_t^i$ *be defined by* (20) *and* (24), *respectively. Under* (A3) *we have the zero-sum property:* $\sum_{i\in\mathcal{N}} \xi_t^i = 0$ *and* $\sum_{i\in\mathcal{N}} \tilde{w}_t^i = 0$ *for all* $t \geq 0$.

Proof. The first equality holds by the definition of ξ_t^i, $1 \leq i \leq n$. We now prove the second equality:

$$\begin{aligned} \sum_{i\in\mathcal{N}} \tilde{w}_t^i &= \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{N}_i} w_t^{i+1,k+1} - \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{N}_i} w_t^{i,k} \\ &= \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{N}_i} w_t^{i,k} - \sum_{i\in\mathcal{N}}\sum_{k\in\mathcal{N}_i} w_t^{i,k} \\ &= 0, \end{aligned} \tag{25}$$

where we get (25) by the circulant invariance property. $\square$

Before further analysis, we introduce the $n \times n$ stochastic matrix

$$M(a) = I + aM^c, \qquad a \in [0,1]. \tag{26}$$

The circulant matrix M^c is given in the form

$$M^c = \begin{bmatrix} -1 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & -1 & c_1 & \ddots & \vdots \\ c_{n-2} & c_{n-1} & -1 & \ddots & c_2 \\ \vdots & \ddots & \ddots & \ddots & c_1 \\ c_1 & \cdots & c_{n-2} & c_{n-1} & -1 \end{bmatrix},$$

where $M^c_{ii} = -1$ for $1 \le i \le n$, and for $2 \le k \le n$,

$$M^c_{1k} = c_{k-1} = \begin{cases} 1/K & \text{if } k \in \mathcal{N}_1, \\ 0 & \text{otherwise.} \end{cases}$$

Since M^c is a circulant matrix [16], it is well defined after the first row is determined. In fact, both M^c and $M(a)$ are circulant matrices.

PROPOSITION 14. *Under* (A3), *$M(a)$ is doubly stochastic for any $a \in [0,1]$; i.e., both $M(a)$ and $[M(a)]^T$ are stochastic matrices, and $M(a)$ is irreducible for $a \in (0,1]$.*

Proof. All row and column sums in $M(a)$ are equal to one. Hence $M(a)$ is doubly stochastic. Since G is strongly connected, $M(a)$ is irreducible for $a > 0$. $\square$

Define $\xi_t = [\xi_t^1, \ldots, \xi_t^n]^T$ and $\tilde{w}_t = [\tilde{w}_t^1, \ldots, \tilde{w}_t^n]^T$. We can check that ξ_t satisfies

$$\xi_{t+1} = M(a_t)\xi_t + (a_t/K)\tilde{w}_t, \quad t \ge 0. \tag{27}$$

The following lemma plays an essential role for the stability analysis of (27).

LEMMA 15. *Assume* (A2)–(A3) *hold, and the real vector $\theta = [\theta_1, \ldots, \theta_n]^T$ has a zero column sum, i.e., $\sum_{i=1}^n \theta_i = 0$. Then for all $t \ge k \ge 0$, we have*

(i) *The column sum of $M(a_t)\ldots M(a_k)\theta$ is zero, i.e., $\sum_{i=1}^n M^\theta_{t,k}(i) = 0$, where we denote $M^\theta_{t,k} = [M^\theta_{t,k}(1), \ldots, M^\theta_{t,k}(n)]^T = M(a_t)\ldots M(a_k)\theta$.*

(ii) *There exist constants $\delta^* \in (0,1)$ and $T_2 > 0$, both independent of θ, such that*

$$|M(a_t)\ldots M(a_k)\theta| \le |(1-\delta^* a_t)\ldots(1-\delta^* a_k)\theta|$$

for all $t \ge k \ge T_2$, where T_2 is chosen such that $a_t \le 1/2$ for all $t \ge T_2$.

Proof. The matrix $M(a_k)$, $k \ge 0$, is doubly stochastic by Proposition 14. Then θ having a zero column sum implies $M(a_k)\theta$ has a zero column sum. Repeating this argument, we obtain part (i).

We now prove (ii). First, let $\omega_n = e^{2\pi \mathbf{i}/n}$, where $\mathbf{i} = \sqrt{-1}$ is the imaginary unit, and denote

$$F_n = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega_n & \omega_n^2 & \cdots & \omega_n^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \omega_n^{n-1} & \omega_n^{2(n-1)} & \cdots & \omega_n^{(n-1)(n-1)} \end{pmatrix},$$

which is the so-called Fourier matrix of order n and satisfies $F_n^* F_n = I$, where F_n^* is the conjugate transpose of F_n. For $a \in [0,1]$, we introduce the polynomial

$$\varphi(a,z) = (1-a) + a(c_1 z + c_2 z^2 + \cdots + c_{n-1} z^n).$$

By well-known results for circulant matrices [16, 8], the n eigenvalues $\{\lambda_{1,t}, \ldots, \lambda_{n,t}\}$ of $M(a_t)$ are given by $\lambda_{k,t} = \varphi(a_t, \omega_n^{k-1})$ for $1 \le k \le n$. Obviously, $\lambda_{1,t} = 1$. Furthermore, $M(a_t)$ may be diagonalized in the form $M(a_t) = F_n^* \times \mathrm{diag}(\lambda_{1,t}, \ldots, \lambda_{n,t}) \times F_n$.

It is easy to verify that

$$\begin{aligned} M(a_t)\ldots M(a_k) &= F_n^* \times \Pi_{j=k}^t \mathrm{diag}(\lambda_{1,j}, \ldots, \lambda_{n,j}) \times F_n \\ &= F_n^* \times \Pi_{j=k}^t \mathrm{diag}(0, \lambda_{2,j}, \ldots, \lambda_{n,j}) \times F_n + (1/n) 1_n 1_n^T. \end{aligned}$$

Since $1_n 1_n^T \theta = 0$ for any θ with a zero column sum, we have

$$M(a_t)\ldots M(a_k)\theta = F_n^* \times \Pi_{j=k}^t \mathrm{diag}(0, \lambda_{2,j}, \ldots, \lambda_{n,j}) \times F_n \theta. \tag{28}$$

Notice that we may write $\varphi(a, w_n^{k-1}) = 1 + c_{k,1}a + \mathbf{i} c_{k,2} a$ for $2 \le k \le n$, where $c_{k,1}$ and $c_{k,2}$ are constants independent of a. For $0 < a < 1$, the matrix $M(a)$ is irreducible and aperiodic,[2] and hence for $2 \le k \le n$, $|\varphi(a, w_n^{k-1})| < \lambda_{1,t} = 1$; the reader is referred to [40] for additional details on spectral theory of stochastic matrices. Then we necessarily have $c_{k,1} < 0$, and in addition, for $0 < a < 1$,

$$|\varphi(a, \omega_n^{k-1})|^2 = (1 + c_{k,1}a)^2 + c_{k,2}^2 a^2 < 1, \quad 2 \le k \le n. \tag{29}$$

By taking $a \uparrow 1$ in (29), we get $-2 \le c_{k,1} < 0$, $|c_{k,2}| \le 1$, and $c_{k,1}^2 + c_{k,2}^2 \le -2c_{k,1}$ for $2 \le k \le n$. Hence it follows that, for $2 \le k \le n$,

$$\begin{aligned} |\lambda_{k,t}|^2 &= (1 + c_{k,1}a_t)^2 + c_{k,2}^2 a_t^2 \\ &\le 1 + 2c_{k,1}a_t - 2c_{k,1}a_t^2 \\ &= (1 + c_{k,1}a_t/2)^2 + c_{k,1}a_t - c_{k,1}^2 a_t^2/4 - 2c_{k,1}a_t^2. \end{aligned}$$

Since $-2 \le c_{k,1} < 0$, we have $c_{k,1}a_t - c_{k,1}^2 a_t^2/4 - 2c_{k,1}a_t^2 = |c_{k,1}| a_t (c_{k,1}a_t/4 + 2a_t - 1) \le 0$ for all $a_t \le 1/2$. Hence for all $t \ge T_2$ such that $a_t \le 1/2$, we have

$$|\lambda_{k,t}| = |\varphi(a_t, \omega_n^{k-1})| \le 1 + c_{k,1}a_t/2, \tag{30}$$

where $2 \le k \le n$. Denote $\delta^* = \inf_{2 \le k \le n} (1/2)|c_{k,1}| > 0$. Then it follows that

$$\Pi_{j=l}^t |\lambda_{k,j}| < \Pi_{j=l}^t (1 - \delta^* a_j) \tag{31}$$

for $2 \le k \le n$, where $t \ge l \ge T_2$. Hence we obtain

$$\begin{aligned} |M(a_t)\ldots M(a_k)\theta|^2 &= \theta^T F_n^* [\Pi_{j=k}^t \mathrm{diag}(0, \lambda_{2,j}, \ldots, \lambda_{n,j})]^* F_n F_n^* \\ &\qquad \times [\Pi_{j=k}^t \mathrm{diag}(0, \lambda_{2,j}, \ldots, \lambda_{n,j})] F_n \theta \\ &\le \Pi_{j=k}^t (1 - \delta^* a_j)^2 |\theta|^2. \end{aligned}$$

This completes the proof. $\square$

[2] When $a < 1$, the n diagonal entries of $M(a)$ are all positive, which ensures aperiodicity of $M(a)$.

COROLLARY 16. *Let* θ, T_2 *and* δ^* *be given as in Lemma* 15 *and denote* $M(t,k) = M(a_t)\dots M(a_k)$ *for* $t > k \geq T_2$. *Then* $M^o(t,k) = F_n^*[\Pi_{j=k}^t \operatorname{diag}(0, \lambda_{2,j}, \dots, \lambda_{n,j})]F_n$ *is a real matrix satisfying*

$$M(t,k)\theta = M^o(t,k)\theta. \tag{32}$$

Moreover, $|M^o(t,k)|_\infty \leq C\Pi_{j=k}^t(1-\delta^* a_j)$ *for some* $C > 0$ *independent of* t, k. *The infinity norm* $|\cdot|_\infty$ *denotes the largest absolute value of the elements in the matrix.*

Proof. Obviously $M^o(t,k)$ is a real matrix since $M^o(t,k) = M(a_t)\dots M(a_k) - (1/n)1_n 1_n^T$, and (32) follows from (28). The estimate for $|M^o(t,k)|_\infty$ follows from (31). ☐

THEOREM 17. *Assume* (A1)–(A3). *Then algorithm* (2) *achieves* (i) *mean square consensus, and* (ii) *strong consensus for* (a) $\gamma \in (1/2, 1)$ *associated with any* $\alpha > 0$ *in* (A2)*, and* (b) $\gamma = 1$ *provided that* $\alpha > 1/(2\delta^*)$.

Proof. The theorem is proved using the same procedure as in the two-agent case. For $\{\xi_t, t \geq 0\}$, we first write the recursion of ξ_t by (27) with the initial time $t = T_1 \vee T_2$ and show its mean square convergence by Lemma 13 and Lemma 15(ii). For proving almost sure convergence of ξ_t, we use Lemma 13, Corollary 16, and Lemma 9 to carry out the double array analysis, where we need to take $\alpha > 1/(2\delta^*)$ for the case $\gamma = 1$.

These combined with Lemma 12 lead to the mean square and almost sure convergence of the n sequences $\{x_t^i, t \geq 0\}$, $i \in \mathcal{N}$, to the same limit. ☐

For deterministic models, if the coefficient matrix in the consensus algorithm is doubly stochastic, the sum of the individual states remains a constant during the iterates. Moreover, if the algorithms achieve consensus, the state of each agent converges to the initial state average, giving the so-called average-consensus [34, 51]. In our model, due to the noise, the limit is a random variable differing from the initial state average although $M(a_t)$ is a doubly stochastic matrix. We have the following performance estimate which illustrates the effect of the noise.

PROPOSITION 18. *Under* (A1)–(A3)*, the state iterates in* (2) *satisfy*

$$E|\lim_{t\to\infty} x_t^i - \operatorname{ave}(x_0)| = \lim_{t\to\infty} E|x_t^i - \operatorname{ave}(x_0)|^2 = O(Q) \qquad \text{for all } i \in \mathcal{N}, \tag{33}$$

where $\operatorname{ave}(x_0) = (1/n)\sum_{k=1}^n x_0^k$ *is the initial state average and* Q *is the variance of the i.i.d. noises.*

Proof. This follows from the mean square consensus result in Theorem 17 and the relation (19). ☐

As the noise variance tends to zero, (33) indicates that the mean square error between $\lim_{t\to\infty} x_t^i$ and $\operatorname{ave}(x_0)$ converges to zero. This is consistent with the corresponding average-consensus results in deterministic models.

5. Consensus seeking on connected undirected graphs. In this section we consider more general network topologies but require that all links are bidirectional; i.e., we restrict our attention to undirected graphs.

Let the location of the n agents be associated with an undirected graph (to be simply called a graph) $G = (\mathcal{N}, \mathcal{E})$ consisting of a set of nodes $\mathcal{N} = \{1, 2, \dots, n\}$ and a set of edges $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$. We denote each edge as an unordered pair (i,j), where $i \neq j$. A path in G consists of a sequence of nodes $i_1, i_2, \dots, i_l$, $l \geq 2$, such that $(i_k, i_{k+1}) \in \mathcal{E}$ for all $1 \leq k \leq l-1$. The graph G is said to be connected if there exists a path between any two distinct nodes. The agent A_k (resp., node k) is a neighbor of A_i (resp., node i) if $(k,i) \in \mathcal{E}$, where $k \neq i$. Denote the neighbors of node i by $\mathcal{N}_i \subset \mathcal{N}$. We make the following assumption.

(A4) The undirected graph G is connected.

5.1. The measurement model and stochastic approximation. The formulation in section 2 is adapted to the undirected graph $G = (\mathcal{N}, \mathcal{E})$ as follows. For each $i \in \mathcal{N}$, we denote the measurement by agent A_i of agent A_k's state by

$$y_t^{ik} = x_t^k + w_t^{ik}, \qquad t \in \mathbb{Z}^+, \quad k \in \mathcal{N}_i, \tag{34}$$

where w_t^{ik} is the additive noise. Write the state vector $x_t = [x_t^1, \dots, x_t^n]^T$. We introduce the following assumption which is slightly weaker for the noise condition than (A1).

(A1′) The noises $\{w_t^{ik}, t \in \mathbb{Z}^+, i \in \mathcal{N}, k \in \mathcal{N}_i\}$ are independent with respect to the indices i, k, t and also independent of x_0, and each w_t^{ik} has zero mean and variance $Q_t^{i,k}$. In addition, $E|x_0|^2 < \infty$ and $\sup_{t\ge 0, i\in\mathcal{N}} \sup_{k\in\mathcal{N}_i} Q_t^{ik} < \infty$.

We use the state updating rule

$$x_{t+1}^i = (1-a_t)x_t^i + \frac{a_t}{|\mathcal{N}_i|}\sum_{k\in\mathcal{N}_i} y_t^{ik}, \qquad t \in \mathbb{Z}^+, \tag{35}$$

where $i \in \mathcal{N}$ and $a_t \in [0,1]$, and we have the relation

$$x_{t+1}^i = x_t^i + a_t(m_t^i - x_t^i), \tag{36}$$

where $m_t^i = (1/|\mathcal{N}_i|)\sum_{k\in\mathcal{N}_i} y_t^{ik}$.

5.2. Stochastic Lyapunov functions. The specification of the stochastic Lyapunov function makes use of the relative positions of the agents. For agent A_i, we define its local potential as

$$P_i(t) = (1/2)\sum_{j\in\mathcal{N}_i} |x_t^i - x_t^j|^2, \qquad t \ge 0.$$

Accordingly, the total potential and total mean potential are given by

$$P_{\mathcal{N}}(t) = \sum_{i\in\mathcal{N}} P_i(t), \qquad V(t) = E\sum_{i\in\mathcal{N}} P_i(t), \quad t \ge 0.$$

It is easy to show that $m_t^i - x_t^i$ in (36) may be decomposed into the form

$$m_t^i - x_t^i = -\frac{1}{|\mathcal{N}_i|}\frac{\partial P_i(t)}{\partial x_t^i} + \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i} w_t^{ij}. \tag{37}$$

This means the state of each agent is updated along the descent direction of the local potential subject to an additive noise, and justifies a stochastic approximation interpretation of algorithm (35). This interpretation is also applicable to digraphs.

Under (A4), it is easy to show that $P_{\mathcal{N}}(t) = 0$ if and only if $x_t^1 = \cdots = x_t^n$. For our convergence analysis, we will use $P_{\mathcal{N}}(t)$ as a stochastic Lyapunov function. We introduce the graph Laplacian for G as a symmetric matrix $L = (a_{ij})_{1\le i,j\le n}$, where

$$a_{ij} = \begin{cases} d_i & \text{if } j = i, \\ -1 & \text{if } j \in \mathcal{N}_i, \\ 0 & \text{otherwise,} \end{cases} \tag{38}$$

and $d_i = |\mathcal{N}_i|$ is the degree (i.e., the number of neighbors) of node i. Denote $1_n = [1, 1, \dots, 1]^T \in \mathbb{R}^n$. Since G is connected, $\text{rank}(L) = n - 1$ and the null space of L

is span$\{1_n\}$ [19, 35]. We have the following relation in terms of the graph Laplacian [19]:

$$P_{\mathcal{N}}(t) = (1/2)\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}_i}|x_t^i - x_t^j|^2 = x_t^T L x_t, \qquad t \geq 0.$$

By (35), we have the state updating rule

$$x_{t+1}^i = (1-a_t)x_t^i + (a_t/|\mathcal{N}_i|)\sum_{j\in\mathcal{N}_i} x_t^j + (a_t/|\mathcal{N}_i|)\sum_{j\in\mathcal{N}_i} w_t^{ij}. \tag{39}$$

Denote

$$\tilde{w}_t^i = (1/|\mathcal{N}_i|)\sum_{j\in\mathcal{N}_i} w_t^{ij}, \qquad \tilde{w}_t = [\tilde{w}_t^1, \ldots, \tilde{w}_t^n]^T. \tag{40}$$

With $d_i = |\mathcal{N}_i|$, we further introduce the matrix $\hat{L} = (\hat{a}_{ij})_{1\leq i,j\leq n}$, where

$$\hat{a}_{ij} = \begin{cases} 1 & \text{if } j = i, \\ -d_i^{-1} & \text{if } j \in \mathcal{N}_i, \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

Define the diagonal matrix $D_{\mathcal{N}} = \mathrm{diag}(d_1^{-1}, \ldots, d_n^{-1})$. Note that $\hat{L} = D_{\mathcal{N}} L$.

LEMMA 19. *For* $t \geq 0$ *and* $\{x_t, t \geq 0\}$ *generated by* (34)–(35), *we have*

$$\begin{aligned} P_{\mathcal{N}}(t+1) = P_{\mathcal{N}}(t) &- 2a_t x_t^T L D_{\mathcal{N}} L x_t + a_t^2 x_t^T L D_{\mathcal{N}} L D_{\mathcal{N}} L x_t \\ &+ 2a_t x_t^T L\tilde{w}_t - 2a_t^2 x_t^T L D_{\mathcal{N}} L \tilde{w}_t + a_t^2 \tilde{w}_t^T L \tilde{w}_t. \end{aligned} \tag{42}$$

Proof. By (39), we get the vector equation

$$x_{t+1} = x_t - a_t \hat{L} x_t + a_t \tilde{w}_t, \qquad t \geq 0, \tag{43}$$

which leads to the recursion of the total potential as follows:

$$\begin{aligned} P_{\mathcal{N}}(t+1) &= x_{t+1}^T L x_{t+1} \\ &= x_t^T L x_t - 2a_t x_t^T L D_{\mathcal{N}} L x_t + a_t^2 x_t^T L D_{\mathcal{N}} L D_{\mathcal{N}} L x_t \\ &\quad + 2a_t x_t^T L \tilde{w}_t - 2a_t^2 x_t^T L D_{\mathcal{N}} L \tilde{w}_t + a_t^2 \tilde{w}_t^T L \tilde{w}_t, \end{aligned}$$

and the lemma follows. ∎

In the subsequent proofs, we use $A \Rightarrow B$ as the abbreviation for "A implies B," and $A \Leftrightarrow B$ for "A is equivalent to B."

LEMMA 20. *Under* (A4), *we have the following assertions:*

(i) *The null spaces of* L, $LD_{\mathcal{N}}L$, *and* $LD_{\mathcal{N}}LD_{\mathcal{N}}L$ *are each given by* span$\{1_n\}$.

(ii) *There exist* $c_1 > 0$ *and* $c_2 > 0$ *such that* $LD_{\mathcal{N}}L \geq c_1 L$ *and* $LD_{\mathcal{N}}LD_{\mathcal{N}}L \leq c_2 L$.

(iii) *In addition, we assume* (A1′)–(A2′) *and let* T_c *be such that* $1-2a_tc_1+a_t^2c_2 \geq 0$ *for all* $t \geq T_c$. *Then for all* $t \geq T_c$, *we have*

$$V(t+1) \leq (1 - 2a_t c_1 + a_t^2 c_2)V(t) + O(a_t^2). \tag{44}$$

Proof. See the appendix. ∎

THEOREM 21. *Under* (A1′)–(A2′) *and* (A4), *algorithm* (35) *achieves weak consensus.*

Proof. For T_c given in Lemma 20(iii), we select $\hat{T}_c \geq T_c$ to ensure $a_t \leq c_1 c_2^{-1}$. Hence $1 - c_1 a_t \geq 1 - 2c_1 a_t + c_2 a_t^2 \geq 0$ for all $t \geq \hat{T}_c$, and we find a fixed constant $C > 0$ such that

$$V(t+1) \leq (1 - c_1 a_t)V(t) + Ca_t^2$$

for all $t \geq \hat{T}_c$; this leads to

$$V(t+1) \leq V(\hat{T}_c) \prod_{j=\hat{T}_c}^{t} (1 - c_1 a_j) + C \sum_{k=\hat{T}_c}^{t} \prod_{j=k+1}^{t} (1 - c_1 a_j) a_k^2, \tag{45}$$

where $\prod_{j=t+1}^{t}(1 - c_1 a_j) \triangleq 1$. Under (A2′), elementary estimates using (45) yield

$$\lim_{t\to\infty} V(t) = 0. \tag{46}$$

It immediately follows that

$$\lim_{t\to\infty} E|x_t^i - x_t^k|^2 = 0, \qquad i \in \mathcal{N},\ k \in \mathcal{N}_i. \tag{47}$$

Since G is connected, there exists a path between any pair of distinct nodes i and k. By repeatedly applying (47) to all pairs of neighboring nodes along that path, we can show that $\lim_{t\to\infty} E|x_t^i - x_t^k|^2 = 0$ for any $i, k \in \mathcal{N}$. ∎

COROLLARY 22. *In Theorem* 21*, we assume all other assumptions but replace* (A2′)(ii) *by the condition* (H)*: There exists $T_0 > 0$ such that for $t \geq T_0$, $\alpha_0 t^{-\gamma_0} \leq a_t \leq \beta_0 t^{-\gamma_0}$ holds for some $0 < \alpha_0 < \beta_0 < \infty$ and $\gamma_0 \in (0, 1/2]$. Then algorithm* (35) *still achieves weak consensus.*

Proof. For (45), we have $\prod_{j=k+1}^{t}(1 - c_1 a_j)a_k^2 \leq \prod_{j=k+1}^{t}(1 - c_1 a_j/2)^2 a_k^2$. We apply Corollary 7 to show that (46) still holds. This completes the proof. ∎

Remark. Notice that under (H), $\sum_{t=0}^{\infty} a_t^2 = \infty$. The conditions of Corollary 22 in general do not ensure mean square consensus.

5.3. The direction of invariance. Theorem 21 shows the difference between the states of any two agents converges to zero in mean square. However, this alone does not mean that they will converge to a common limit. The asymptotic vanishing of the stochastic Lyapunov function indicates only that the state vector x_t will approach the subspace span$\{1_n\}$. To obtain mean square consensus results, we need some additional estimation. The strategy is to show that the oscillation of the sequence $\{x_t, t \geq 0\}$ along the direction 1_n will gradually die off. This is achieved by proving the existence of a vector η which is not orthogonal to 1_n and such that the linear combination $\eta^T x_t$ of the components in x_t converges. For convenience, η will be chosen to satisfy the additional requirement that $\eta^T x_{t+1}$ depends not on the whole of x_t but only on $\eta^T x_t$; this will greatly facilitate the associated calculation.

DEFINITION 23. *Let $x_t = [x_t^1, \ldots, x_t^n]^T$ be generated by* (35)*. If $\eta = [\eta_1, \ldots, \eta_n]^T$ is a real-valued vector of unit length, i.e., $|\eta|^2 = \sum_{i=1}^{n} \eta_i^2 = 1$, and satisfies*

$$\eta^T x_{t+1} = \eta^T x_t + a_t \eta^T \tilde{w}_t, \qquad t \geq 0, \tag{48}$$

for any initial condition x_0 and any step size sequence $a_t \in [0, 1]$, where $\tilde{w}_t$ is given in (40)*, then η is called a direction of invariance associated with* (35).

The directions of invariance associated with the consensus algorithm (35) are easily characterized in terms of the degrees of the nodes of the underlying graph.

THEOREM 24. *We have the following assertions:*

(i) *There exists a real-valued vector* $\eta = (\eta_1, \ldots, \eta_n)^T$ *of unit length satisfying* $\eta^T \hat{L} = 0$, *where* $\hat{L}$ *is defined by* (41).

(ii) *If* $|\eta| = 1$, *then* η *is a direction of invariance for* (35) *if and only if* $\eta^T \hat{L} = 0$.

(iii) *Under* (A4), *the direction of invariance for* (35) *has the representation* $\eta = c[d_1, \ldots, d_n]^T$, *where* $c = \pm(\sum_{i=1}^n d_i^2)^{1/2}$ *and* $d_i = |\mathcal{N}_i|$ *is the degree of node* i.

Proof. It is easy to prove (i) since $\hat{L}$ does not have full rank, and η is in fact the left eigenvector of $\hat{L}$ associated with the eigenvalue 0.

We now show (ii). The condition $\eta^T \hat{L} = 0$ combined with (43) implies

$$\eta^T x_{t+1} = \eta^T x_t + a_t \eta^T \tilde{w}_t.$$

The sufficiency part of (ii) follows easily. Conversely, if the unit length vector η satisfies (48) for all initial states x_0^i and the step size a_t as specified in Definition 23, then we necessarily have $\eta^T \hat{L} = 0$. So the necessity part of (ii) holds.

We continue to prove (iii) under (A4). By (ii) and the definition of $\hat{L}$, η with $|\eta| = 1$ is a direction of invariance if and only if $\eta^T D_{\mathcal{N}} L = 0$, which in turn is equivalent to $L D_{\mathcal{N}} \eta = 0$. By (A4) and Lemma 20, we have $D_{\mathcal{N}} \eta = c 1_n$, where $c \neq 0$ is a constant to be determined. This gives $\eta = c[d_1, \ldots, d_n]^T$, where c is determined by the condition $|\eta| = 1$. The direction of invariance is unique up to sign. ◻

If η is a direction of invariance, then Theorem 24 shows under (A4) that all elements of η have the same sign. Therefore, η is not orthogonal to 1_n, and the requirement stated at the beginning of this section is met. Geometrically, the notion of the direction of invariance means under (35) and zero noise conditions, the projection (i.e., $(\eta^T x_t)\eta$) of x_t in $\mathbb{R}^n$ along the direction η would remain a constant vector regardless of the value of $a_t \in [0, 1]$ used in the iterates.

5.4. Mean square consensus. Now we are in a position to establish mean square consensus.

LEMMA 25. *Assume* (A1′)–(A2′) *and* (A4), *and let* $\{x_t, t \geq 0\}$ *be given by* (35), $\eta_0 = [d_1, \ldots, d_n]^T$, *where* $d_i = |\mathcal{N}_i|$. *Then there exists a random variable* y^* *such that* $\lim_{t\to\infty} E|\eta_0 x_t - y^*|^2 = 0$.

Proof. By Theorem 24, $\eta_0/|\eta_0|$ is a direction of invariance. Hence, we have

$$\eta_0^T x_{t+1} = \eta_0^T x_0 + a_0 \eta_0^T \tilde{w}_0 + \cdots + a_t \eta_0^T \tilde{w}_t.$$

By (A1′) and (A2′), it follows that $\eta_0^T x_t$ converges in mean square. ◻

The weak consensus result combined with the convergence of $\eta_0^T x_t$ ensures that x_t itself converges.

THEOREM 26. *Under* (A1′)–(A2′) *and* (A4), *algorithm* (35) *achieves mean square consensus.*

Proof. By Theorem 21, we have weak consensus, i.e.,

$$\lim_{t\to\infty} E|x_t^i - x_t^k|^2 = 0 \qquad \text{for all } i, k \in \mathcal{N}. \tag{49}$$

On the other hand, by Lemma 25, as $t \to \infty$,

$$\eta_0^T x_t = \eta_0^T [x_t^1 - x_t^1, \ldots, x_t^n - x_t^1]^T + \eta_0^T [x_t^1, \ldots, x_t^1]^T$$

converges in mean square, which combined with (49) implies x_t^1 converges in mean square. By (49) again, the mean square consensus result follows. ◻

6. Leader following and convergence. Now we apply the stochastic Lyapunov function approach to the scenario of leader following [23, 44]. Suppose there are n agents located in the digraph $G_d = (\mathcal{N}, \mathcal{E})$, and without loss of generality, denote the leader by agent A_1. We denote by $\mathcal{N}_F = \mathcal{N}\backslash\{1\}$ the set of follower agents. For $i \in \mathcal{N}$, denote the individual states by x_t^i, $t \in \mathbb{Z}^+$. The leader A_1 does not receive measurements from other agents; to capture this feature in G_d, there is no edge reaching A_1 from other agents. The initial state of A_1 is chosen randomly, after which the state remains constant. That is, $x_t^1 \equiv \vartheta$, where ϑ is a random variable, which is unknown to any other agent A_i, $i \in \mathcal{N}_F$.

For node $i \in \mathcal{N}_F$, its measurement is given as

$$y_t^{i,k} = x_t^k + w_t^{i,k}, \qquad t \in \mathbb{Z}^+, \quad k \in \mathcal{N}_i,$$

where $w_t^{i,k}$ is the additive noise. For $i \in \mathcal{N}_F$, the state is updated by

$$x_{t+1}^i = (1 - a_t)x_t^i + \frac{a_t}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} y_t^{ij}. \tag{50}$$

We adapt (A1′) to the graph $G_d = (\mathcal{N}, \mathcal{E})$ in an obvious manner. But it should be kept in mind that in this leader following model the noise term w_t^{ik} is defined only for $i \in \mathcal{N}_F$ since the leader has no neighbor. Also, $x_0^1 \equiv \vartheta$ since A_1 is the leader, and under (A1′), we have $E|\vartheta|^2 < \infty$.

To make the problem nontrivial, we use the following underlying assumption.

(A5) In $G_d = (\mathcal{N}, \mathcal{E})$, node 1 is the neighbor of at least one node in $\mathcal{N}_F$.

Now, based on the digraph $G_d = (\mathcal{N}, \mathcal{E})$, we set each $(i, j) \in \mathcal{E}$ as an unordered pair and this procedure induces an undirected graph $G_u = (\mathcal{N}, \mathcal{E}^u)$ with its associated graph Laplacian L^u. We decompose L^u into the form

$$L^u = \begin{bmatrix} L_1^u \\ L_{n-1}^u \end{bmatrix},$$

where L_1^u is the first row in L^u.

In order to develop the stochastic Lyapunov analysis, we need some restrictions on the set of nodes $\mathcal{N}_F$ and the associated edges. Let $(\mathcal{N}_F, \mathcal{E}_F)$ denote the directed subgraph of $(\mathcal{N}, \mathcal{E})$ obtained by removing node 1 and all edges containing 1 as the initial node. We introduce the following assumption.

(A6) An ordered pair $(i, j) \in \mathcal{E}_F$ implies the ordered pair (j, i) is also in $\mathcal{E}_F$.

Remark. (A5)–(A6) imply that at least one follower can receive information from the leader while the information exchange among the followers is bidirectional.

In analogy to the construction of G_u, we induce from the digraph $(\mathcal{N}_F, \mathcal{E}_F)$ an undirected graph, denoted by $G_{Fu} = (\mathcal{N}_F, \mathcal{E}_F^u)$. We introduce the following assumption.

(A7) The undirected graph $G_{Fu} = (\mathcal{N}_F, \mathcal{E}_F^u)$ is connected.

PROPOSITION 27. *Under* (A5)–(A7), *the undirected graph* $G_u = (\mathcal{N}, \mathcal{E}^u)$ *is connected and* $\text{rank}(L^u) = \text{rank}(L_{n-1}^u) = n - 1$.

Proof. It is obvious that G_u is connected. Hence $\text{rank}(L^u) = n - 1$. Since $1_n^T L^u = 0$, it follows that L_1^u is a linear combination of the rows in L_{n-1}^u, which implies $\text{rank}(L_{n-1}^u) = n - 1$. ∎

Denote $x_{\vartheta,t} = [\vartheta, x_t^2, \ldots, x_t^n]^T$, $\tilde{w}_i = (1/|N_i|) \sum_{k \in \mathcal{N}_i} w_t^{i,k}$ for $i \geq 2$, and $\tilde{w}_t = [0, \tilde{w}_t^2, \ldots, \tilde{w}_t^n]^T$. Write $D_0 = \text{diag}(0, d_2^{-1}, d_n^{-1})$. By writing (50) in the vector form, we get the following lemma.

LEMMA 28. *We have the recursion for the state vector*

$$x_{\vartheta,t+1} = x_{\vartheta,t} - a_t D_0 L^u x_{\vartheta,t} + a_t \tilde{w}_t, \qquad t \geq 0,$$

where $x_{\vartheta,t}$ is generated by algorithm (50).

THEOREM 29. *Under* (A1′)–(A2′), (A5)–(A7)*, and algorithm* (50)*, we have*

$$\lim_{t\to\infty} E|x_t^i - \vartheta|^2 = 0 \tag{51}$$

for all $i \in \mathcal{N}_F$, where ϑ is the fixed random variable as the state for the leader.

Proof. Step 1. Define the stochastic Lyapunov function $P_{\vartheta,\mathcal{N}}(t) = x_{\vartheta,t}^T L^u x_{\vartheta,t}$, where $L^u \geq 0$, and denote $V(t) = EP_{\vartheta,\mathcal{N}}(t)$, $t \geq 0$. By Lemma 28, it is easy to show

$$\begin{aligned} V(t+1) = V(t) &- 2a_t E[x_{\vartheta,t}^T L^u D_0 L^u x_{\vartheta,t}] \\ &+ a_t^2 E[x_{\vartheta,t}^T L^u D_0 L^u D_0 L^u x_{\vartheta,t}] + O(a_t^2). \end{aligned} \tag{52}$$

Let $y_\theta = [\theta, y_2, \ldots, y_n]^T$, where θ denotes a fixed real number. First, by $\mathrm{rank}(L^u) = n-1$, we can show that $L^u y_\theta = 0 \Leftrightarrow y_\theta = \theta 1_n^T$. Obviously $L^u y_\theta = 0 \Rightarrow L^u D_0 L^u y_\theta = 0 \Rightarrow L^u D_0 L^u D_0 L^u y_\theta = 0$. On the other hand, letting $L^u = [(L^u)^{1/2}]^2$, where $(L^u)^{1/2} \geq 0$, we have $L^u D_0 L^u D_0 L^u y_\theta = 0 \Rightarrow (L^u)^{1/2} D_0 L^u y_\theta = 0 \Rightarrow L^u D_0 L^u y_\theta = 0 \Rightarrow \mathrm{diag}(0, d_2^{-1/2}, \ldots, d_n^{-1/2}) L^u y_\theta = 0 \Leftrightarrow \mathrm{diag}(d_2^{-1/2}, \ldots, d_n^{-1/2}) L_{n-1}^u y_\theta = 0 \Leftrightarrow L_{n-1}^u y_\theta = 0 \Leftrightarrow L^u y_\theta = 0$ since $\mathrm{rank}(L_{n-1}^u) = \mathrm{rank}(L^u) = n-1$ by Proposition 27. Now we conclude that $\theta 1_n$ is the unique point where each of $y_\theta^T L^u y_\theta$, $y_\theta^T L^u D_0 L^u y_\theta$, and $y_\theta^T L^u D_0 L^u D_0 L^u y_\theta$ attains its minimum 0.

Step 2. Letting $y^{(n-1)} = [y_2, \ldots, y_n]^T$, we introduce three positive semidefinite quadratic forms in terms of $y^{(n-1)}$: $Q_1(y^{(n-1)}) = y_\theta^T L^u y_\theta$, $Q_2(y^{(n-1)}) = y_\theta^T L^u D_0 L^u y_\theta$, and $Q_3(y^{(n-1)}) = y_\theta^T L^u D_0 L^u D_0 L^u y_\theta$. Let $z = y^{(n-1)} - \theta 1_{n-1}$, and we may write

$$0 \leq Q_1(y^{(n-1)}) = z^T M_1 z + v^T z + c,$$

where M_1 is an $(n-1) \times (n-1)$ symmetric matrix, $v \in \mathbb{R}^{n-1}$, and $c \in \mathbb{R}$. Clearly $z^T M_1 z + v^T z + c = 0 \Leftrightarrow z = 0$ since $Q_1(y^{(n-1)}) = 0 \Leftrightarrow y_\theta = \theta 1_n$ by Step 1; by elementary linear algebra and a contradictory argument we can show $c = 0$, $v^T = 0$, and $M_1 > 0$. Hence, $Q_1(y^{(n-1)}) = z^T M_1 z$. Since M_1 is constructed based on the second order coefficient of $y^{(n-1)}$ in $y_\theta^T L^u y_\theta$, we see that M_1 is independent of θ. Similarly, we can find matrices $M_2 > 0$ and $M_3 > 0$, both independent of θ, such that

$$Q_2(y^{(n-1)}) = z^T M_2 z, \qquad Q_3(y^{(n-1)}) = z^T M_3 z,$$

where $z = y^{(n-1)} - \theta 1_{n-1}$. We denote the smallest and largest eigenvalue of M_i, respectively, by $\lambda_{i,\min} > 0$ and $\lambda_{i,\max} > 0$ for $i = 1, 2, 3$. Now we have

$$Q_2(y^{(n-1)}) = z^T M_2 z \geq \lambda_{2,\min} \lambda_{1,\max}^{-1} z^T M_1 z = \lambda_{2,\min} \lambda_{1,\max}^{-1} Q_1(y^{(n-1)}), \tag{53}$$

$$Q_3(y^{(n-1)}) = z^T M_3 z \leq \lambda_{3,\max} \lambda_{1,\min}^{-1} z^T M_1 z = \lambda_{3,\max} \lambda_{1,\min}^{-1} Q_1(y^{(n-1)}). \tag{54}$$

Step 3. Now it follows from (52) and (53)–(54) that

$$V(t+1) \leq (1 - 2\tau_1 a_t + \tau_2 a_t^2) V(t) + O(a_t^2), \tag{55}$$

where $\tau_1 = \lambda_{2,\min} \lambda_{1,\max}^{-1}$ and $\tau_2 = \lambda_{3,\max} \lambda_{1,\min}^{-1}$. Consequently, by use of product estimates as in (45), we can show $\lim_{t\to\infty} V(t) = 0$. Since the first entry in $x_{\vartheta,t}$

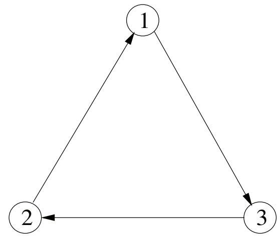

FIG. 4. *A digraph with* 3 *nodes.*

is ϑ and the associated undirected graph $G_u = (\mathcal{N}, \mathcal{E}^u)$ is connected, by the same argument as in proving weak consensus in Theorem 21, we can obtain (51). ∎

Remark. In Theorem 29, if (A2′)(ii) is replaced by the condition (H): $\alpha_0 t^{-\gamma_0} \le a_t \le \beta_0 t^{-\gamma_0}$ for $t \ge T_0$, where $\alpha_0 > 0$ and $\gamma_0 \in (0, 1/2]$ (see Corollary 22), then Theorem 29 still holds. This may be proved by combining the proving argument for Corollary 22 with (55) to get $\lim_{t\to\infty} V(t) = 0$.

7. Numerical studies.

7.1. Simulations with a symmetric digraph. The digraph is shown in Figure 4, where $\mathcal{N}_1 = \{2\}$, $\mathcal{N}_2 = \{3\}$, and $\mathcal{N}_3 = \{1\}$. The initial condition for $x_t = [x_t^1, x_t^2, x_t^3]$ is $[4, 3, 1]$ at $t = 0$, and the i.i.d. Gaussian measurement noises have variance $\sigma^2 = 0.01$. Figure 5 shows the simulation with equal weights to an agent's neighbors and itself (as in Example 1) in the averaging rule ($x_{t+1}^1 = (x_t^1 + y_t^{12})/2$, etc.), without obtaining consensus. Figure 6 shows the convergence of algorithm (2) with the step size sequence $\{a_t = (t + 5)^{-0.85}, t \ge 0\}$.

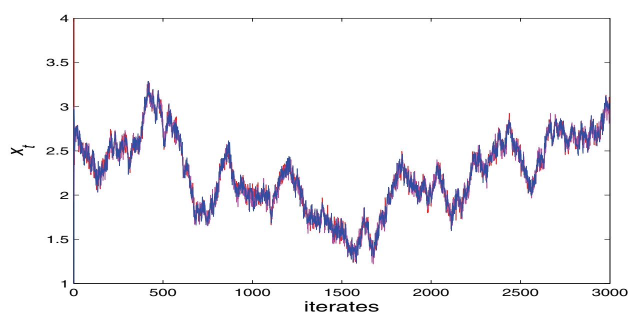

FIG. 5. *Equal weights are used for each agent's state and observation.*

7.2. Simulations with an undirected graph. The undirected graph is shown in Figure 7 with $\mathcal{N} = \{1, 2, 3, 4\}$ and $\mathcal{E} = \{(1, 2), (2, 3), (2, 4)\}$. The initial condition is $x_t|_{t=0} = [5, 1, 3, 2]^T$, and the i.i.d. Gaussian noises have variance $\sigma^2 = 0.01$. The simulation of the averaging rule with equal weights is given in Figure 8; hence we have $x_{t+1}^1 = (x_t^1 + y_t^{12})/2$ and $x_{t+1}^2 = x_t^2/4 + (y_t^{21} + y_t^{23} + y_t^{24})/4$, etc., where $t \ge 0$. It is seen that the 4 state trajectories in Figure 8 move towards each other rather quickly at the beginning, but they maintain long term fluctuations as the state iteration continues. The stochastic algorithm (35) is used in Figure 9, where $a_t = (t + 5)^{-0.85}$, $t \ge 0$. Figure 9 shows the 4 trajectories all converge to the same constant level.

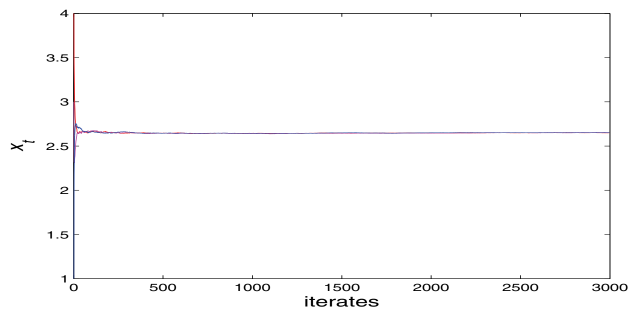

FIG. 6. *The* 3*-agent example using decreasing step size* $a_t = (t + 5)^{-0.85}$.

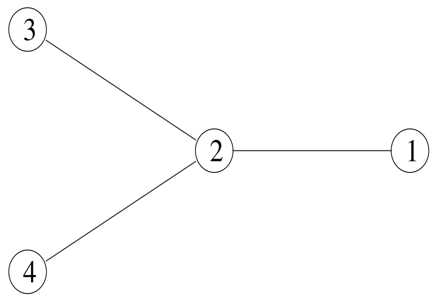

FIG. 7. *The undirected graph with* 4 *nodes.*

7.3. The leader following model. We adapt the undirected graph in Figure 7 to the leader following situation as follows. We set node 1 as the leader (without a neighbor) and $\mathcal{N}_2 = \{1, 3, 4\}$, $\mathcal{N}_3 = \{2\}$, $\mathcal{N}_4 = \{2\}$. We take $x_0^1 \equiv 4$ and the initial condition is given as $x_t|_{t=0} = [4, 2, 1, 3]^T$. Figure 10 shows the simulation with equal weights for each follower agent and its neighbors. We see that all three states of the follower agents move into a neighborhood of the constant level 4 and oscillate around that value. Compared with Figure 8, the trajectories of the followers in Figure 10 have a far smaller fluctuation. The reason is that in the leader following case, the total potential attains its minimum only at the leader's state rather than at all points in span$\{1_n\}$, which results in more regular behavior for the agents. In Figure 11 we show the simulation of algorithm (50) with $a_t = (t + 5)^{-0.65}$, $t \geq 0$, which exhibits a satisfactory convergent behavior.

8. Concluding remarks. We consider consensus problems for networked agents with noisy measurements. First, the double array analysis is developed to analyze mean square and almost sure convergence. Next, stochastic Lyapunov functions are introduced to prove mean square consensus with the aid of the so-called direction of invariance, and this approach is further applied to leader following. We note that the methods developed in this paper may be extended to deal with general digraphs, and the second order moment condition for the noise may be relaxed when applying the stochastic double array analysis; see the recent work [22] for details. For future work, it is of interest to develop stochastic algorithms in models with dynamic topologies and asynchronous state updates, and in particular, extend the double array analysis to networks with switching topologies.

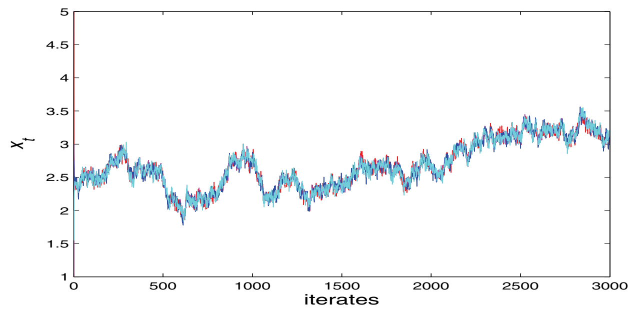

FIG. 8. *The 4-agent example using equal weights for each agent's state and observations.*

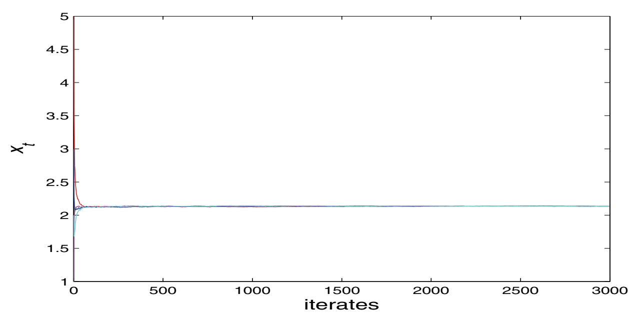

FIG. 9. *The 4-agent example using a decreasing step size $a_t = (t+5)^{-0.85}$.*

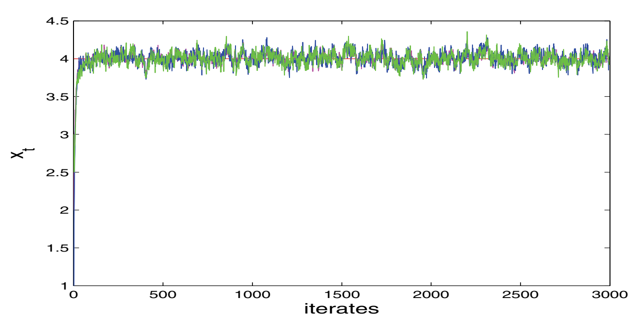

FIG. 10. *Leader following using equal weights for each follower agent's state and observations.*

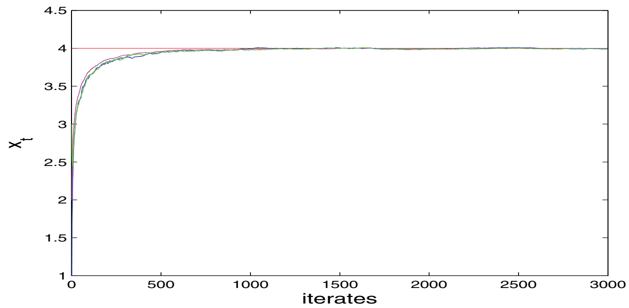

FIG. 11. *Leader following using a decreasing step size* $a_t = (t+5)^{-0.65}$.

Appendix.

Proof of Lemma 6. For case (i), by (10) we have

$$\sum_{k=T_1}^{t} \Pi_{t,k}^2 \le \sum_{k=T_1}^{t} \frac{\beta^2 (k+1)^{4\alpha}}{k^2 (t+1)^{4\alpha}}.$$

The desired upper bound is obtained from elementary estimates by considering three scenarios for α as in (14).

We continue with the estimate for case (ii). Let $\delta = 2\alpha/(1-\gamma) > 0$ and define

$$S_t = \sum_{k=1}^{t} k^{-2\gamma} e^{2\delta(k+1)^{1-\gamma}}, \qquad H_t = t^{-\gamma} e^{2\delta(t+1)^{1-\gamma}}, \quad t \ge 1.$$

Clearly there exists a sufficiently large $t_0 > 0$ such that H_t is strictly increasing for $t \ge t_0$. In addition, both S_t and H_t diverge to infinity. If we can show that for $t > t_0$,

$$0 < R_t = \frac{S_t - S_{t-1}}{H_t - H_{t-1}} \to R^*, \qquad \text{as } t \to \infty, \tag{A.1}$$

for some $R^* > 0$, then it is straightforward to show that $S_t = O(H_t)$. To show the existence of a limit in (A.1), we write

$$R_t = \frac{t^{-2\gamma} e^{2\delta(t+1)^{1-\gamma}}}{t^{-\gamma} e^{2\delta(t+1)^{1-\gamma}} - (t-1)^{-\gamma} e^{2\delta t^{1-\gamma}}}. \tag{A.2}$$

We have

$$\begin{aligned}
& t^{-\gamma} e^{2\delta(t+1)^{1-\gamma}} - (t-1)^{-\gamma} e^{2\delta t^{1-\gamma}} \\
&= t^{-\gamma} e^{2\delta(t+1)^{1-\gamma}} - t^{-\gamma} e^{2\delta[(t+1)^{1-\gamma} + t^{1-\gamma} - (t+1)^{1-\gamma}]} (1-t^{-1})^{-\gamma} \\
&= t^{-\gamma} e^{2\delta(t+1)^{1-\gamma}} [1 - e^{2\delta[t^{1-\gamma} - (t+1)^{1-\gamma}]} (1-t^{-1})^{-\gamma}] \\
&= t^{-\gamma} e^{2\delta(t+1)^{1-\gamma}} \left[1 - e^{-2\delta[(1-\gamma)t^{-\gamma} + o(t^{-\gamma})]}\right] [1 + \gamma t^{-1} + o(t^{-1})] \\
&= t^{-\gamma} e^{2\delta(t+1)^{1-\gamma}} [2\delta(1-\gamma)t^{-\gamma} + o(t^{-\gamma})][1 + \gamma t^{-1} + o(t^{-1})] \\
&= 2\delta(1-\gamma) t^{-2\gamma} e^{2\delta(t+1)^{1-\gamma}} [1 + o(1)] \\
&= 4\alpha t^{-2\gamma} e^{2\delta(t+1)^{1-\gamma}} [1 + o(1)]. \qquad \text{(A.3)}
\end{aligned}$$

By combining (A.2) and (A.3), it follows that $\lim_{t\to\infty} R_t = 4\alpha > 0$, and hence $S_t = O(H_t)$. Subsequently, we have

$$\begin{aligned}\sum_{k=T_1}^{t} \Pi_{t,k}^2 &= O\left(e^{-2\delta(t+1)^{1-\gamma}} \sum_{k=1}^{t} k^{-2\gamma} e^{2\delta(k+1)^{1-\gamma}}\right)\\ &= O(e^{-2\delta(t+1)^{1-\gamma}} H_t)\\ &= O(t^{-\gamma}),\end{aligned}$$

which completes the proof for case (ii), and the lemma follows. ∎

Proof of Lemma 20. (i) First, it is a well-known fact [19, 35] that when G is connected, the null space of L is $\text{span}\{1_n\}$. Since $L \geq 0$, there exists a positive semidefinite matrix, denoted as $L^{1/2}$ such that $L = (L^{1/2})^2$. We also write $D_{\mathcal{N}}^{1/2} = \text{diag}(d_1^{-1/2}, \ldots, d_n^{-1/2})$ which gives $D_{\mathcal{N}} = (D_{\mathcal{N}}^{1/2})^2$. For $x \in \mathbb{R}^n$, we have $Lx = 0 \Rightarrow LD_{\mathcal{N}}Lx = 0 \Rightarrow LD_{\mathcal{N}}LD_{\mathcal{N}}Lx = 0$.

On the other hand, we have

$$\begin{aligned}LD_{\mathcal{N}}LD_{\mathcal{N}}Lx = 0 &\Rightarrow x^T LD_{\mathcal{N}}LD_{\mathcal{N}}Lx = 0\\ &\Leftrightarrow |L^{1/2}D_{\mathcal{N}}Lx|^2 = 0 \Leftrightarrow L^{1/2}D_{\mathcal{N}}Lx = 0\\ &\Rightarrow LD_{\mathcal{N}}Lx = 0 \Rightarrow x^T LD_{\mathcal{N}}Lx = 0\\ &\Leftrightarrow D_{\mathcal{N}}^{1/2}Lx = 0 \Leftrightarrow Lx = 0.\end{aligned}$$

Hence, it follows that $Lx = 0 \Leftrightarrow LD_{\mathcal{N}}Lx = 0 \Leftrightarrow LD_{\mathcal{N}}LD_{\mathcal{N}}Lx = 0$, and assertion (i) follows. Hence the matrices L, $LD_{\mathcal{N}}L$, and $LD_{\mathcal{N}}LD_{\mathcal{N}}L$ each have a rank of $n-1$.

(ii) We begin by proving the first part. Let $0 = \lambda_1,\ 0 < \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n$ and $0 = \hat{\lambda}_1,\ 0 < \hat{\lambda}_2 \leq \hat{\lambda}_3 \leq \cdots \leq \hat{\lambda}_n$, respectively, denote the eigenvalues of L and $LD_{\mathcal{N}}L$. Let $\Phi = (\alpha_1, \ldots, \alpha_n)$ and $\hat{\Phi} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_n)$ be two orthogonal matrices (i.e., $\Phi^T\Phi = I$, and $\hat{\Phi}^T\hat{\Phi} = I$) such that

$$L\Phi = \Phi\,\text{diag}(\lambda_1, \ldots, \lambda_n), \qquad LD_{\mathcal{N}}L\hat{\Phi} = \hat{\Phi}\,\text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_n).$$

In view of $\lambda_1 = \hat{\lambda}_1 = 0$, we get $L\alpha_1 = LD_{\mathcal{N}}L\hat{\alpha}_1 = 0$. By (i), we necessarily have either $\alpha_1 = \hat{\alpha}_1$ or $\alpha_1 = -\hat{\alpha}_1$. In fact, we may take $\alpha_1 = \hat{\alpha}_1 = \pm(1/\sqrt{n}) \cdot 1_n$. Consequently, it is easy to show that $\text{span}\{\alpha_2, \ldots, \alpha_n\} = \text{span}\{\hat{\alpha}_2, \ldots, \hat{\alpha}_n\}$, which is the orthogonal complement of $\text{span}\{1_n\}$ in $\mathbb{R}^n$.

Take any $x \in \mathbb{R}^n$. We may write $x = \sum_{i=1}^n y_i\alpha_i$, $x = \sum_{i=1}^n \hat{y}_i\hat{\alpha}_i$, where $y = (y_1, \ldots, y_n)$, $\hat{y} = (\hat{y}_1, \ldots, \hat{y}_n)$ are uniquely determined and satisfy $\sum_{i=1}^n y_i^2 = \sum_{i=1}^n \hat{y}_i^2 = |x|^2$. Recalling that we have taken $\alpha_1 = \hat{\alpha}_1 \neq 0$, it necessarily follows that $y_1 = \hat{y}_1$ since, otherwise, $(y_1 - \hat{y}_1)\alpha_1 \in \text{span}\{\alpha_2, \ldots, \alpha_n\}$ with $y_1 - \hat{y}_1 \neq 0$, which is impossible. Hence we get

$$\sum_{i=2}^n y_i^2 = \sum_{i=2}^n \hat{y}_i^2. \tag{A.4}$$

For $x \in \mathbb{R}^n$, since $\lambda_1 = \hat{\lambda}_1 = 0$ we have the estimate

$$x^T LD_{\mathcal{N}}Lx = \hat{y}^T\hat{\Phi}^T LD_{\mathcal{N}}L\hat{\Phi}\hat{y} = \sum_{i=2}^n \hat{\lambda}_i\hat{y}_i^2 \geq \hat{\lambda}_2 \sum_{i=2}^n \hat{y}_i^2.$$

On the other hand, we have $x^TLx \leq \lambda_n \sum_{i=2}^n y_i^2 = \lambda_n \sum_{i=2}^n \hat{y}_i^2$, where the equality follows from (A.4). Hence it follows that $x^TLD_{\mathcal{N}}Lx \geq \hat{\lambda}_2\lambda_n^{-1}x^TLx$, and therefore, the first part of (ii) is proved by taking $c_1 = \hat{\lambda}_2\lambda_n^{-1} > 0$.

We denote the eigenvalues of $LD_{\mathcal{N}}LD_{\mathcal{N}}L$ by $0 = \tilde{\lambda}_1,\ 0 < \tilde{\lambda}_2 \leq \tilde{\lambda}_3 \leq \cdots \leq \tilde{\lambda}_n$. Following a very similar argument, we can show that for any $x \in \mathbb{R}^n$,

$$x^TLD_{\mathcal{N}}LD_{\mathcal{N}}Lx \leq \tilde{\lambda}_n\lambda_2^{-1}x^TLx,$$

which implies the second part with $c_2 = \tilde{\lambda}_n\lambda_2^{-1} > 0$.

(iii) We obtain (44) by taking expectation on both sides of (42) and using (ii). ∎

Acknowledgments. The authors would like to thank the reviewers and the Associated Editors for very useful comments and suggestions. This paper was submitted when both authors were with the Research School of Information Sciences and Engineering, The Australia National University.

REFERENCES

[1] S. Aranda, S. Martinez, and F. Bullo, *On optimal sensor placement and motion coordination for target tracking*, in Proceedings of the IEEE International Conference on Robotics and Automation, Barcelona, Spain, 2005, pp. 4544–4549.

[2] P. Barooah and J. P. Hespanha, *Estimation on graphs from relative measurements: Distributed algorithms and fundamental limits*, IEEE Control Syst. Mag., 27 (2007), pp. 57–74.

[3] R. W. Beard and V. Stepanyan, *Synchronization of information in distributed multiple vehicle coordination control*, in Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, HI, 2003, pp. 2029–2034.

[4] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, New York, 1990.

[5] P.-A. Bliman and G. Ferrari-Trecate, *Average consensus problems in networks of agents with delayed communications*, in Proceedings of the 44th IEEE Conference on Decision and Control and 2005 European Control Conference, Seville, Spain, 2005, pp. 7066–7071.

[6] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, *Convergence in multiagent coordination, consensus, and flocking*, in Proceedings of the 44th IEEE Conference on Decision and Control and 2005 European Control Conference, Seville, Spain, 2005, pp. 2996–3001.

[7] V. Borkar and P. Varaiya, *Asymptotic agreement in distributed estimation*, IEEE Trans. Automat. Control, 27 (1982), pp. 650–655.

[8] A. Böttcher and S. M. Grudsky, *Spectral Properties of Banded Toeplitz Matrices*, SIAM, Philadelphia, 2005.

[9] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri, *Communication constraints in coordinated consensus problems*, in Proceedings of the American Control Conference, Minneapolis, MN, 2006, pp. 4189–4194.

[10] S. Chatterjee and E. Seneta, *Towards consensus: Some convergence theorems on repeated averaging*, J. Appl. Probab., 14 (1977), pp. 89–97.

[11] H.-F. Chen, *Stochastic Approximation and Its Applications*, Kluwer, Boston, MA, 2002.

[12] Y. S. Chow and T. L. Lai, *Limiting behavior of weighted sums of independent random variables*, Ann. Probab., 1 (1973), pp. 810–824.

[13] Y. S. Chow and H. Teicher, *Probability Theory: Independence, Interchangeability, Martingales*, Springer-Verlag, New York, 1978.

[14] J. Cortes, S. Martinez, and F. Bullo, *Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions*, IEEE Trans. Automat. Control, 51 (2006), pp. 1289–1298.

[15] G. A. de Castro and F. Paganini, *Convex synthesis of controllers for consensus*, in Proceedings of the American Control Conference, Boston, MA, 2004, pp. 4933–4938.

[16] P. J. Davis, *Circulant Matrices*, John Wiley, New York, 1979.

[17] J. P. Desai, V. Kumar, and J. P. Ostrowski, *Control of changes in formation for a team of mobile robots*, in Proceedings of the IEEE International Conference on Robotics and Automation, Detroit, MI, 1999, pp. 1556–1561.

[18] M. J. Fischer, *The consensus problem in unreliable distributed systems (a brief survey)*, in Foundations of Computation Theory, M. Karpinsky ed., Lecture Notes in Comput. Sci. 158, Springer-Verlag, New York, 1983, pp. 127–140.
[19] C. Godsil and G. Royle, *Algebraic Graph Theory*, Springer-Verlag, New York, 2001.
[20] M. H. de Groot, *Reaching a consensus*, J. Amer. Statist. Assoc., 69 (1974), pp. 118-121.
[21] Y. Hatano and M. Mesbahi, *Agreement over random networks*, IEEE Trans. Automat. Control, 50 (2005), pp. 1867–1872.
[22] M. Huang and J. H. Manton, *Stochastic approximation for consensus seeking: Mean square and almost sure convergence*, in Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA, 2007, pp. 306–311.
[23] A. Jadbabaie, J. Lin, and A. S. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Trans. Automat. Control, 48 (2003), pp. 988–1000.
[24] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, New York, 1997.
[25] N. E. Leonard, D. A. Paley, F. Lekien, R. Sepulchre, D. M. Fratantoni, and R. E. Davis, *Collective motion, sensor networks and ocean sampling*, in Proc. IEEE, 95 (2007), pp. 48–74.
[26] Y. Liu and Y. R. Yang, *Reputation propagation and agreement in mobile ad hoc networks*, in Proceedings of the IEEE Wireless Communications and Networking Conference, New Orleans, LA, 2003, pp. 1510–1515.
[27] L. Ljung, *Analysis of recursive stochastic algorithms*, IEEE Trans. Automat. Control, 22 (1977), pp. 551–575.
[28] N. Lynch, *Distributed Algorithms*, Morgan Kaufmann, San Mateo, CA, 1996.
[29] J. A. Marshall and M. E. Broucke, *On invariance of cyclic group symmetries in multiagent formations*, in Proceedings of the 44th IEEE Conference on Decision and Control and 2005 European Control Conference, Seville, Spain, 2005, pp. 746–751.
[30] J. A. Marshall, M. E. Broucke, and B. A. Francis, *Formations of vehicles in cyclic pursuit*, IEEE Trans. Automat. Control, 49 (2004), pp. 1963–1974.
[31] L. Moreau, *Stability of continuous-time distributed consensus algorithms*, in Proceedings of the 43rd IEEE Conference on Decision and Control, Paradise Island, Bahamas, 2004, pp. 3998–4003.
[32] L. Moreau, *Stability of multiagent systems with time-dependent communication links*, IEEE Trans. Automat. Control, 50 (2005), pp. 169–182.
[33] R. Olfati-Saber, *Flocking for multi-agent dynamic systems: Algorithms and theory*, IEEE Trans. Automat. Control, 51 (2006), pp. 401–420.
[34] R. Olfati-Saber and R. M. Murray, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Trans. Automat. Control, 49 (2004), pp. 1520–1533.
[35] R. Olfati-Saber, J. A. Fax, and R. M. Murray, *Consensus and cooperation in networked multi-agent systems*, Proc. IEEE, 95 (2007), pp. 215–233.
[36] A. V. Oppenheim and R. W. Schafter, *Discrete-Time Signal Processing*, Prentice–Hall, Englewood Cliffs, NJ, 1989.
[37] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice–Hall, Upper Saddle River, NJ, 1996.
[38] W. Ren, R. W. Beard, and E. M. Atkins, *A survey of consensus problems in multi-agent coordination*, in Proceedings of the American Control Conference, Portland, OR, 2005, pp. 1859–1864.
[39] W. Ren, R. W. Beard, and D. B. Kingston, *Multi-agent Kalman consensus with relative uncertainty*, in Proceedings of the American Control Conference, Portland, OR, 2005, pp. 1865–1870.
[40] E. Seneta, *Nonnegative Matrices and Markov Chains*, 2nd ed., Springer-Verlag, New York, 1981.
[41] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, *Consensus in ad hoc WSNs with noisy links—Part* I: *Distributed estimation of deterministic signals*, IEEE Trans. Signal Process., 56 (2008), pp. 350–364.
[42] A. Tahbaz-Salehi and A. Jadbabaie, *On consensus in random networks*, in Proceedings of the 44th Annual Allerton Conference on Communication, Control and Computing, Monticello, IL, 2006, pp. 1315–1321.
[43] H. G. Tanner, G. J. Pappas, and V. Kumar, *Input-to-state stability on formation graphs*, in Proceedings of the 41st IEEE Conference on Decision and Control, Las Vegas, NV, 2002, pp. 2439–2444.
[44] H. G. Tanner, G. J. Pappas, and V. Kumar, *Leader-to-formation stability*, IEEE Trans. Robotics Automat., 20 (2004), pp. 443–455.

[45] H. Teicher, *Almost certain convergence in double arrays*, Z. Wahrsch. Verw. Gebiete, 69 (1985), pp. 331–345.
[46] J. N. Tsitsiklis and M. Athans, *Convergence and asymptotic agreement in distributed decision problems*, IEEE Trans. Automat. Control, 29 (1984), pp. 42–50.
[47] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Trans. Automat. Control, 31 (1986), pp. 803–812.
[48] P. K. Visscher, *How self-organization evolves*, Nature, 421 (2003), pp. 799–800.
[49] R. L. Winkler, *The consensus of subjective probability distributions*, Management Science, 15 (1968), pp. B61–B75.
[50] L. Xiao and S. Boyd, *Fast linear iterations for distributed averaging*, Systems Control Lett., 53 (2004), pp. 65–78.
[51] L. Xiao, S. Boyd, and S.-J. Kim, *Distributed average consensus with least-mean-square deviation*, J. Parallel Distrib. Comput., 67 (2007), pp. 33–46.