MULTISCALE MODEL. SIMUL.
Vol. 7, No. 4, pp. 1898–1927

CONSENSUS FORMATION IN A TWO-TIME-SCALE MARKOVIAN SYSTEM*

VIKRAM KRISHNAMURTHY†, KEVIN TOPLEY†, AND GEORGE YIN‡

Abstract. This work analyzes distributed linear averaging within a connected network of sensors that each track the stationary distribution of an ergodic Markov chain with a slowly switching regime. Our approach is based on a two-time-scale stochastic approximation. A hyperparameter modeled as a Markov chain on a slower time-scale modulates the regime of each observed Markov chain. The average of all currently observed stationary distributions constitutes the average-consensus estimate to be reached by all sensors. Assuming the Markov chains do not share a common stationary distribution conditioned on their regime, then under the proposed linear averaging algorithm, the exchange graph conditions required for the sequence of sensor state values to converge weakly to the average-consensus are obtained. Estimation of a weighted average of *all* observed stationary distributions, not only the current ones, is proved feasible over a long-run time horizon, provided an additional communication condition holds. The sensor state values are also shown to converge weakly to solutions of a differential inclusion when the communication exchange graphs or observed Markov chains belong to a family of possible values, thus leading to a set-valued consensus formation. The rate of convergence of the consensus algorithm is studied by considering the scaled tracking errors when oriented about their steady-state for each regime of the hyperparameter. In addition, a Brownian bridge limit is obtained for a centered and scaled sequence of empirical measures. An adaptation rate is proposed as the minimum exponential rate of the sensor trajectories to the average-consensus estimate. Various optimization problems related to this adaptation rate are posed, as well as an approximate ratio that relates between any two sets of exchange graphs the adaptation rate, sensor scaled error, and absolute sum total averaging weights. Simulations illustrate our results and observation model.

Key words. two-time-scale, consensus formation, weak convergence

AMS subject classifications. 62L20, 62M05, 90C35

DOI. 10.1137/080743652

1. Introduction.

The coordination of a consensus within a multiagent system is a problem that has been considered in a variety of general settings and solved under a wide range of assumptions regarding the interagent communication capabilities. Algorithms that result in "consensus" have thus in general taken on a variety of forms; for some examples see [13, 43, 42, 18, 45, 19, 3, 15, 12, 41, 17, 7, 8, 37, 33]. Here we will focus on a specific consensus algorithm, also referred to as a consensus "protocol" [30, 31, 35, 2, 36], under which a network consensus is obtained by linear averaging of the data held at neighboring sensor nodes.

Linear averaging has, understandably, been a particular research focal point regarding consensus formation; see [10, 50, 21, 38, 5, 40]. In these and many other works on consensus, the multiagent system is generally defined as a collection (or "network") of *coupled* sensors (or "nodes") with no extensive memory base; hence the

*Received by the editors December 2, 2008; accepted for publication (in revised form) May 6, 2009; published electronically July 30, 2009.
http://www.siam.org/journals/mms/7-4/74365.html

†Department of Electrical Engineering and Department of Electrical and Electronic Engineering, University of British Columbia, Vancouver V6T 1Z4, Canada (vikramk@ece.ubc.ca, kevint@ece.ubc.ca). The research of these authors was supported by Defense Research Development Canada and the Natural Sciences and Engineering Research Council of Canada.

‡Department of Mathematics, Wayne State University, Detroit, MI 48202 (gyin@math.wayne.edu). The research of this author was supported in part by the National Science Foundation under DMS-0603287 and in part by the National Security Agency under grant MSPF-068-029.

linear averaging scheme is analyzed with respect to a group of decentralized and distributed agents. This particular network paradigm and consensus mechanism will be together referred to here as *decentralized distributed averaging* (DDA). Under a DDA algorithm, every sensor has a direct linear effect on the estimates of each of its neighbors. Thus the DDA algorithm is scalable and is robust under a variety of network communication conditions [23, 9, 30, 31]. This is in contrast to consensus protocols for specialized data fusion problems [43, 42, 18], which have complex algorithmic forms that cannot be simplified to mere averaging; see also [38, 21].

In precise terms, the DDA consensus algorithm is an iterative procedure wherein each sensor node linearly averages the information it holds with the information it receives from other sensors with whom it directly communicates. This computation is performed unanimously by all sensors; thus it is clear that DDA requires minimum data storage, labeling of data, or specialization of individual sensors, i.e., use of base nodes. For a network of n sensors, the DDA algorithm can be parameterized by a weighted digraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{W}\}$, where $\mathcal{V} = \{1, \dots, n\}$ denotes the set of sensor nodes, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of directed edges specifying which sensors are coupled (i.e., can transmit or receive information), and $\mathcal{W} \in \mathbf{R}^{n\times n}$ denotes the weights with which neighboring sensors average their held information. By definition the (i, j)th element of $\mathcal{W}$ is zero if $(i, j) \notin \mathcal{E}$; with abuse of notation we denote this $\mathcal{W} \in \mathcal{E}$.

Distributed consensus-tracking algorithm. In this paper, a discrete-time DDA algorithm is considered to operate together with a linear stochastic approximation (SA) tracking algorithm based locally at each sensor $i \in \mathcal{V}$. Each SA tracking algorithm aims to estimate the stationary distribution $\pi^i \in \mathbf{R}^S$ of a fast S-state ergodic Markov chain $X^i \in \mathbf{R}^S$ that is observed privately by sensor node i. Motivated by applications in adaptive tracking in sensor networks, we assume the transition matrix of each Markov chain $\{X^i : i \in \mathcal{V}\}$ jump changes slowly with time. More specifically, we assume that the transition matrix of each Markov chain $\{X^i : i \in \mathcal{V}\}$ is conditioned on a hyperparameter θ taking values in a finite set $\mathcal{M}$. The dynamics of θ is provided by θ_k, a slowly varying m-state Markov chain with state-space $\mathcal{M} = \{\theta^1, \dots, \theta^m\}$, and transition matrix $P^\varepsilon = I + \varepsilon Q$, where $\varepsilon > 0$ is a small parameter and $Q = (q_{ij}) \in \mathbf{R}^{m\times m}$ is a generator of a continuous-time Markov chain. Since ε is small, θ is a slowly varying process and will jump infrequently among different states.

For each $\theta \in \mathcal{M}$ and $i \in \mathcal{V}$, X^i_k is an S-state Markov chain with a transition matrix modulated by the slow Markov chain θ_k; thus we describe the sequence $\{\theta_k, X^i_k : i \in \mathcal{V},\ k \in \mathbb{N}\}$ as a switched Markovian system. As a result, each stationary distribution $\pi^i(\theta)$ is θ-dependent. When θ_k switches its values from one state to another within $\mathcal{M}$, $\pi^i(\theta_k)$ switches accordingly, and so it is necessary for the sensor network to track these time-varying distributions.

Let us give an intuitive explanation of consensus formation. Given the averaging weights $\mathcal{W}$ and graph edge set $\mathcal{E}$ of the sensor network described above, each node $i \in \mathcal{V}$ computes via the distributed consensus tracking algorithm (iteration (1.1) below) a state-value, denoted $s^i \in \mathbf{R}^S$. We will prove that a suitably scaled continuous-time version of the state-value s^i converges weakly to a linear combination $\tilde{\pi}^i(\theta(t)) = \sum_{l=1}^n \psi^i_l \pi^l(\theta(t)) \in \mathbf{R}^S$ of the stationary distributions $\{\pi^i(\theta(t)) : i \in \mathcal{V}\}$. Here the coefficients ψ^i_l are determined by the choices of $\mathcal{W}$ and $\mathcal{E}$. Consensus formation deals with the assumptions on the information exchange between sensors required to obtain a specific structure of $\tilde{\pi}^i(\theta(t))$ in the above expression. The sensor network achieves a *consensus* if for any given $l \in \mathcal{V}$, $\psi^i_l = \psi^j_l$ for all $i, j \in \mathcal{V}$. The *average-consensus* occurs if $\psi^i_l = \frac{1}{n}$ for all i, l. Due to the dependence of π^i on θ, the values of $\tilde{\pi}^i(\theta)$ stochastically switch among a finite set of values, each corresponding to a particular $\theta \in \mathcal{M}$. The average-consensus in this setting is thus a random process. To supplement the weak convergence results, we derive conditions on the averaging weights required for each sensor to obtain the stochastic average-consensus $\bar{\pi}(\theta(t))$.

Assuming both the DDA and local tracking algorithms use a constant step-size μ at each discrete-time iteration $k \in \mathbb{N}$, it will be seen that in order for all sensors to track the average distribution $\bar{\pi}(\theta)$ each sensor must distributively average *both* its observed signal, denoted X_k^i, *and* its state-value s_k^i. Define the *observation* exchange graph $\mathcal{G}^o = \{\mathcal{V}, \mathcal{E}^o, \mathcal{W}^o\}$ and *state-value* exchange graph $\mathcal{G}^v = \{\mathcal{V}, \mathcal{E}^v, \mathcal{W}^v\}$ that, respectively, determine how each type of data, X^i (observation) and s^i (state), is distributively averaged. The distributed consensus-tracking algorithm we then consider is as follows:

(1.1)
$$\begin{aligned} s_{k+1} &= (I - \mu\mathcal{D}^v + \mu\mathcal{W}^v - \mu\mathcal{D}^o)s_k + \mu\mathcal{W}^o X_k \ , \qquad && s(0) = X(0), \quad \mathcal{D}^p = \mathrm{diag}(\mathcal{W}^p \mathbb{1}), \quad p \in \{o, v\} \\ &= s_k - \mu H s_k + \mu \mathcal{W}^o X_k \ , && H = \mathcal{D}^o + \mathcal{D}^v - \mathcal{W}^v, \end{aligned}$$

where $s_k = [s_k^1, \dots, s_k^n]' \in \mathbf{R}^{Sn\times 1}$, $X_k = [X_k^1, \dots, X_k^n]' \in \mathbf{R}^{Sn\times 1}$, the term $\mathbb{1}$ denotes a column vector of ones with appropriate dimension, and we let $\mathrm{diag}(r)$ be the $n \times n$ diagonal matrix with elements corresponding to those of the vector $r \in \mathbf{R}^n$. We will henceforth let each element of the weight matrices $\{\mathcal{W}^v, \mathcal{W}^o\}$ represent an $S \times S$ identity matrix scaled by the respective element; that is, we let $\mathcal{W} = \mathcal{W} \bigotimes I_{S\times S}$, where $\bigotimes$ denotes the Kronecker product [22]. As a final note, throughout the paper we let subscript k indicate dependence on a discrete-time in $\mathbb{N}$ (e.g., θ_k), whereas dependence on continuous-time $t \geq 0$ will be denoted (t) (e.g., $\theta(t)$). The single exception to this rule is the initial data, which is always denoted (0) for both discrete-time and continuous-time systems; see (1.1).

The above algorithm can be viewed as comprising two DDA algorithms together with a family of local stochastic approximations; each DDA algorithm averages with the local state-value either the state-values at neighboring nodes (determined by $\mathcal{G}^v$) or the observations at neighboring nodes (determined by $\mathcal{G}^o$). For a sufficiently small step-size μ, the iteration (1.1) implies that at each discrete-time $k \in \mathbb{N}$, the state-value s_k^i computed at any sensor $i \in \mathcal{V}$ is updated linearly in three ways as follows (the superscripts * and ** below are intermediate variables for illustration only and will not be used subsequently):

1. by its own local observation X_k^i,

$$(s_k^i)^* = s_k^i + \mu \mathcal{W}_{ii}^o (X_k^i - s_k^i); \tag{1.2}$$

2. by the local observation X_k^j of each neighboring sensor $j \in \{j : (i,j) \in \mathcal{E}^o\}$,

$$(s_k^i)^{**} = (s_k^i)^* + \mu \sum_{j\neq i} \mathcal{W}_{ij}^o (X_k^j - (s_k^i)^*);$$

3. by the state-value s_k^j of each neighboring sensor $j \in \{j : (i,j) \in \mathcal{E}^v\}$,

$$s_{k+1}^i = (s_k^i)^{**} + \mu \sum_{j\neq i} \mathcal{W}_{ij}^v (s_k^j - (s_k^i)^{**}).$$

As μ vanishes, these three nested steps may be written as the single iteration (1.1) because in this limit all second and third order terms in μ become negligible compared to the first order terms that are retained in (1.1). We also note that due to the similarity between steps 1 and 2, each row $i \in \mathcal{V}$ of the exchange graph $\mathcal{G}^o$ may either specify which sensors *communicate* their observed information X^j to sensor i, or equivalently which Markov chains X^j sensor i can actually *observe*, assuming that the same SA algorithm (1.2) with update weights $\mathcal{W}^o_{ij}$ are applied to each Markov chain X^j.

Context. The tracking algorithm (1.1) was first introduced in [35] as the distributed consensus filter. However, in [35] it is assumed (1) $\mathcal{W}^o = \mathcal{W}^v + I$, (2) the diagonal of $\mathcal{W}^v$ is zero, and (3) each nonzero averaging weight is unity. Furthermore, the observation variables $\{X^1_k, \dots, X^n_k\}$ are reduced to a single time-varying scalar observed in independent and identically distributed Gaussian noise by all sensors. In contrast to [35], recall that we assume a Markov-modulated dynamic model where each sensor $i \in \mathcal{V}$ observes the Markov chain X^i_k. Our observation model involves a two-time-scale formulation with parameters μ and ε. The step-size μ is used in the recursive tracking and averaging algorithms, whereas the step-size ε represents the transition rate of the Markov chain θ.

To analyze the continuous-time limit of (1.1), it is assumed in [35] that the observed scalar has a uniformly bounded derivative as the step-size μ approaches zero. Here, we assume $\varepsilon = O(\mu)$, indicating that (1.1) has a tracking rate on the same time-scale as the Markov modulating process. Our main result shows that under suitable conditions on the network exchange graphs $\{\mathcal{G}^v, \mathcal{G}^o\}$, the assumption $\varepsilon = O(\mu)$ implies the limit of a piecewise constant continuous-time interpolation of the sequence of each sensor's estimates s^i converges weakly to the solution of a Markov modulated ODE as $\mu \to 0$. Considering also the sequence of sensor tracking errors, we demonstrate under similar conditions that an interpolation of the normalized errors converges weakly to a switching diffusion. Under additional weight constraints it then follows as a special case of the above results that the estimates s^i converge weakly to the average-consensus $\bar{\pi}(\theta)$.

Within this general framework we also address two modifications of the consensus-tracking algorithm (1.1). One modification assumes that rather than a single chain X^i each sensor i observes a family of Markov chains $X^i_k \in \mathbb{X}^i_k$, thus implying the sensor state-values converge weakly to solutions of a switched differential inclusion. The second modification considers sensor estimation of the cumulative distribution function (CDF) $\Pi^i(\theta)$, $i \in \mathcal{V}$, $\theta \in \mathcal{M} = \{\theta^1, \dots, \theta^m\}$, by using empirical measures. In this case the sensor scaled tracking error converges weakly to a switched Brownian bridge. We deal with each of these modifications separately in sections 2 and 3.

Related work. We briefly review here the related literature on consensus formation. The distributed consensus-tracking algorithm (1.1) can be viewed as the combination and extension of two subalgorithms: a local sensor adaptive SA tracking algorithm,

$$s^i_{k+1} = s^i_k + \mu(X^i_k - s^i_k), \quad s^i(0) = X^i(0), \quad 0 < \mu \ll 1,$$

operating in parallel with the well-known (DDA) Laplacian consensus algorithm,

$$s_{k+1} = (I - \mu\mathcal{L})s_k, \quad \mathcal{L} = \mathcal{D}^v - \mathcal{W}^v. \tag{1.3}$$

The consensus algorithm (1.3) has been explored in works such as [31, 9] regarding its ability to achieve an average-consensus of sensor (or "node") initial state-values $s(0) = [s^1(0), \dots, s^n(0)]$, that is, $\bar{s}(0) = \frac{1}{n}\sum_{i=1}^n s^i(0)$. It is clear the average-consensus $\bar{s}(0)$ is constant at all times; thus the algorithm (1.3) by itself can result in a *static* consensus that need not be driven by some external parameters, such as the observed Markov chains X_k considered here (see also [50] and the references therein).

As proposed in [35], we extend the Laplacian consensus algorithm to include not only averaging of the sensor state-values s^i but also of the sensor observations X_k^i. This extension is in fact necessary for DDA to ensure average-consensus when considered within the generality of the current setting, specifically when each sensor observes and estimates a unique parameter, as we assume here (see (4.1) below). Unlike certain works on distributed averaging such as [35, 21, 5], we allow $\{\mathcal{G}^v, \mathcal{G}^o\}$ to be both directed and weighted, such as those considered, for instance, in [31, 49, 2]. By directed it is meant $(i,j) \in \mathcal{E} \nRightarrow (j,i) \in \mathcal{E}$, and by weighted it is meant $\mathcal{W}$ need not have every element belong to the set $\{0,1\}$. We also note that if $(i,j) \in \mathcal{E}$ implies $\mathcal{W}_{ij} \neq 0$, as we may assume here, then knowledge of $\{\mathcal{W}^v, \mathcal{W}^o\}$ is sufficient to completely define $\{\mathcal{G}^v, \mathcal{G}^o\}$.

The Laplacian consensus algorithm (1.3) derives its name due to its asymptotic equivalence with the gradient system $\dot{s}(t) = -\nabla\Phi_{\mathcal{G}}(s(0))$ associated with the Laplacian potential

$$\Phi_{\mathcal{G}}(s(0)) = \frac{1}{2}\sum_{i,j=1}^{n} \mathcal{W}_{ij}^v (s^j(0) - s^i(0))^2 = \frac{1}{2} s(0)'\mathcal{L}s(0),$$

where $\mathcal{L} = \operatorname{diag}(\mathcal{W}^v\mathbb{1}) - \mathcal{W}^v$ is defined as the Laplacian matrix of an undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}^v, \mathcal{W}^v\}$ and each node $i \in \mathcal{V}$ has an initial state-value $s^i(0) \in \mathbf{R}$ [31]. Many works [38, 30, 28, 21, 46] consider the Laplacian consensus dynamics (1.3) in isolation from any tracking model; such works are thus concerned with a static, rather than time-varying, consensus formation. The general theme addressed in these works regards the network graph conditions under which (1.3) solves the static average-consensus problem, that is, having all sensors reach $\bar{s}(0)$, given that $\mathcal{E}^v$ has time-varying or stochastic properties. These concerns are closely related to the issues of time-delayed or asynchronous communication within sensor networks and other multiagent systems; we refer the reader to works such as [44, 2, 20, 11, 7, 39, 31, 28, 29] for detailed consideration.

In contrast, [10] assumes the communication edge set $\mathcal{E}^v$ is fixed, undirected, and *connected* (that is, there exists at least one node from which all other nodes can be reached by traversing the edges in $\mathcal{E}^v$). Under these conditions it is shown in the cited work that the matrix $W = I - \mu\mathcal{L}$ satisfying

$$\text{(1.4)} \qquad \begin{array}{ll} \text{minimize} & \| W - \frac{1}{n}\mathbb{1}\mathbb{1}' \|_2 \\ \text{subject to} & W \in \mathcal{E}^v,\ W\mathbb{1} = \mathbb{1},\quad \mathbb{1}'W = \mathbb{1}' \end{array}$$

will imply the maximum asymptotic per-step convergence factor $r_{\text{step}}(W)$ of the sensor state-value estimates s_k to the average-consensus value $\bar{s}(0)$, where

$$r_{\text{step}}(W) = \sup_{s_k \neq \bar{s}(0)} \frac{\| s_{k+1} - \bar{s}(0) \|_2}{\| s_k - \bar{s}(0) \|_2}.$$

In (1.4) we note that $||\cdot||_2$ indicates spectral norm; thus (1.4) is a convex optimization problem and in [10] is cast as a semidefinite program. The relation of [10] to the present work is noted since the same constraints of (1.4) are sufficient to ensure sensor estimation of the average-consensus $\bar{\pi}(\theta)$, and thus the weights W that solve (1.4) may also be argued as the optimal weights for fast convergence to $\bar{\pi}(\theta)$; see section 4.

The consensus-tracking framework of our paper is also similar to recent works [34, 35, 32, 1, 39]. These papers deploy (1.3) as part of a distributed Kalman filter or in conjunction with local sensor estimation of a time-varying parameter; thus these works also deal with a time-varying consensus value. For example, assuming that each sensor node i observes in continuous-time the m-dimensional signal data pair $\{z_i(t), \dot{z}_i(t)\}$, the following result is proved in [32]: For any sequence of weighting matrices $W_i(t)$ the dynamical system

$$\dot{x}_i = W_i^{-1}(x_j - x_i) + \dot{z}_i + W_i^{-1}\dot{W}_i(z_i - x_i), \quad x_i(0) = z_i(0), \tag{1.5}$$

ensures that every sensor estimate x^i tracks the weighted-average consensus

$$\lim_{t\to\infty} x_i(t) = \left(\sum_{i=1}^{n} W_i(\infty)\right)^{-1} (W_i(\infty)\ z_i(\infty)),\ i = 1, \ldots, n,$$

with zero steady-state error, provided the Laplace transforms of both $z_i(t)$ and $W_i(t)$ each have all poles in the left-hand plane with at most one pole at zero. In the switched Markovian observation model considered here, we do not assume knowledge of $\dot{z}_i(t)$ since this would imply direct observation of the parameter $\{\theta_k\}$. In our framework each sensor observes θ only indirectly through changes in the approximated stationary distribution of the observed Markov chains X_k; thus an algorithm such as (1.5) is not applicable.

In light of extensive research conducted regarding multiagent coordination, there are several practical applications of the DDA algorithm in sensor networks, for instance the synchronization of node clocks [36] or distributed load balancing [14]. The problem of synchronizing coupled oscillators by means of DDA was discussed in [16], whereas in [4] the authors consider DDA as a mechanism to ensure a team of UAVs (unmanned air vehicles) will approach a steady-state wherein each UAV surveys an equal portion of a one-dimensional perimeter. Conversely, [47] models a set of mobile agents in the plane and develops a distributed control law for stable flocking behavior. Similarly, [24] applies the decentralized network paradigm of coupled agents to explain the consensus behavior of a group of self-driven particles during phase transition as was previously considered in [48].

Outline. Section 2 formulates the observed Markovian system and fundamental exchange graph requirements and then presents the resulting weak convergence of the sensor state-value estimates s^i under (1.1). The scaled sensor tracking error is discussed in section 3, with particular focus on sensor estimation of the stationary CDF $\Pi^i(\theta)$ associated with $\pi^i(\theta)$. Discussion and rationale of the network graph conditions required for average-consensus are detailed in section 4, as well as a discussion regarding the factors concerning the sensor adaption rates. Numerical simulations are presented in section 5. The remaining proofs are contained in section 6.

2. Asymptotic consensus dynamics.

In this section, we show that (1.1) is sufficient for each sensor s_k^i for $i \in \mathcal{V}$ to track a linear combination of the stationary distributions $\{\pi^1(\theta), \ldots, \pi^n(\theta)\}$. Let X_k^i be an S-state Markov chain with state-space $\{e_1, \ldots, e_S\}$, where each e_i is an $S \times 1$ standard unit vector. Each X_k^i is θ-dependent for $\theta \in \mathcal{M} = \{\theta^1, \ldots, \theta^m\}$, a finite set such that the transition matrix of X_k^i conditioned on θ is given by $A^i(\theta) = (a_{lj}^i(\theta))$, where

$$a_{lj}^i(\theta) = P(X_{k+1}^i = e_j | X_k^i = e_l, \theta_k = \theta).$$

To proceed, we pose the following conditions.

(A) 1. For each $\theta \in \mathcal{M}$, the transition matrix $A^i(\theta)$ is irreducible and aperiodic for each $i \in \mathcal{V}$.
2. Parameterize the transition probability matrix of θ as

$$P^\varepsilon = I + \varepsilon Q,$$

where ε is a small parameter satisfying $0 < \varepsilon \ll 1$ and Q is the generator of a continuous-time finite-state Markov chain.
3. The process θ_k is slow in the sense that $\varepsilon = O(\mu)$. For simplicity, we take $\varepsilon = \mu$ henceforth.

(B) All eigenvalues of the matrix $H = (\mathcal{L} + \mathcal{D}^o)$ have positive real parts or non-negative real parts. We denote these exchange graph conditions as $0 \prec H$ and $0 \preccurlyeq H$, respectively.

Condition (A) specifies our observation model as a two-time-scale Markovian system. Condition (B) is a constraint on the network exchange graphs $\{\mathcal{G}^o, \mathcal{G}^v\}$ and ensures bounded stability of (1.1) in the limit as $\mu \to 0$. We note that $0 \prec H$ implies $-H$ is a Hurwitz matrix; under this or $0 \preccurlyeq H$ we later show in section 4 that further weight conditions can ensure an average-consensus is obtained.

THEOREM 2.1. *Assume conditions* (A)–(B) *and suppose* $s(0)$ *is independent of* μ. *Define the continuous-time interpolated sequences of iterates*

$$s^\mu(t) = s_k, \quad \theta^\mu(t) = \theta_k \quad \textit{for} \quad t \in [k\mu, (k+1)\mu).$$

Then as $\mu \to 0$, $(s^\mu(\cdot), \theta^\mu(\cdot))$ *converges weakly to* $(s(\cdot), \theta(\cdot))$ *such that* $\theta(\cdot)$ *is a continuous-time Markov chain with generator* Q *and* $s(\cdot)$ *satisfies*

$$\frac{ds(t)}{dt} = -Hs(t) + \mathcal{W}^0\pi(\theta(t)),\ t \geq 0,\ s(0) = X(0). \tag{2.1}$$

Note that if $s(0) = s^\mu(0)$, we require that $s^\mu(0)$ converge to $s(0)$ weakly. However, for simplicity, we choose $s(0)$ to be independent of μ. The above theorem implies the sensor iterates resulting from (1.1) converge weakly to a Markovian switched ODE. This is in contrast to the "standard" analysis of SA algorithms [27], where the limiting process is a deterministic ODE. The proof of the theorem uses the martingale problem formulation of Stroock and Varadhan; see also [51]. If a consensus is obtained by all sensors, that is, $s^i(t) = s^j(t)$ for all $i, j \in \mathcal{V}$, then this consensus is a stochastic process dictated by θ. Again this is in contrast to works concerned with a static consensus formation (i.e., [31, 38, 21, 30, 9]), as well as others wherein the average-consensus estimate is a linear combination of time-varying signals unanimously observed by *all* sensors [35] or with observed rates of change [32]. In comparison, the consensus estimate we obtain is an average of piecewise fixed finite-state Markov chain stationary distributions, where it may be assumed each Markov chain is observed by only one sensor.

Long-time horizon. Theorem 2.1 states a convergence result for small μ and large k such that μk remains bounded. However, if the consensus-tracking algorithm (1.1) is in operation for a long time, we would like to establish its behavior for a large-time horizon. We thus next consider the case when μ is small and k is large such that $\mu k \to \infty$. This is essentially a stability result corresponding to the limit of the switching ODE (2.1) as $t \to \infty$. For this result we assume that Q is irreducible, or equivalently the associated continuous-time Markov chain $\theta(t)$ is irreducible. Recall that the irreducibility means that the system of equations

$$\begin{cases} \nu' Q = 0, \\ \mathbb{1}'\nu = \sum_{j=1}^{m} \nu_j = 1 \end{cases} \tag{2.2}$$

has a unique solution satisfying $\nu_j > 0$ for all $j = 1, \ldots, m$. The vector $\nu = (\nu_1, \ldots, \nu_m)' \in \mathbf{R}^m$ is then the stationary distribution of θ. Since θ is a finite-state Markov chain, we note that if $0 \prec H$, then

$$\sum_{i=1}^{m} \int_0^T \exp(-H(T-u))\mathcal{W}^0 \pi(\theta^i) I_{\{\theta(u)=\theta^i\}} du$$

converges with probability 1 (w.p.1) as $T \to \infty$ due to the exponential dominance of the term $\exp(-H(t-u))$. A closer scrutiny permits an expression of this limit in closed form. Denote

$$s_* = H^{-1} \sum_{i=1}^{m} \mathcal{W}^o \pi(\theta^i)\nu_i.$$

THEOREM 2.2. *Assume the conditions of Theorem* 2.1 *with the modifications that Q is irreducible and $-H$ is Hurwitz. Then for any sequence $\{t_\mu\}$ satisfying $t_\mu \to \infty$ as $\mu \to 0$, $s^\mu(\cdot + t_\mu)$ converges weakly to s_*. Moreover, for any $0 < T < \infty$, $\sup_{|t|\le T} |s^\mu(t+t_\mu) - s_*| \to 0$ in probability.*

In Theorem 2.2, the generator Q is assumed irreducible. We may also consider the case when Q has an absorbing state and thus the remaining $m-1$ states are transient. This implies there is a zero row in Q. In this case, although the Markov chain is not irreducible, the limit probability distribution still exists. This distribution is a vector with one component equal to unity (corresponding to the absorbing state) and the remaining components equal to zero. As a result, the techniques used in Theorem 2.2 can still be applied. We state the following result but omit its proof for brevity.

THEOREM 2.3. *Assume the conditions of Theorem* 2.2 *with the modification that the Markov chain θ has an absorbing state. Without loss of generality, denote this state by θ^1. Then for any sequence $\{t_\mu\}$ satisfying $t_\mu \to \infty$ as $\mu \to 0$, $s^\mu(\cdot + t_\mu)$ converges weakly to s_a, where*

$$s_a = H^{-1}\mathcal{W}^0 \pi(\theta^1).$$

Moreover, for any $0 < T < \infty$, $\sup_{|t|\le T} |s^\mu(t+t_\mu) - s_a| \to 0$ in probability.

Set-valued consensus formation. In Theorem 2.1, it is assumed that for each θ, the observed Markov chains $\{X_k^i : i \in \mathcal{V}\}$ will each have a respective transition matrix $A^i(\theta)$ that is irreducible and aperiodic. This is an ergodicity condition that may be generalized to the case when the observed stationary distributions are not unique, but instead each belong to a set $\mathcal{A}^i(\theta)$ of transition matrices with corresponding set of stationary distributions $\Gamma_\pi^i(\theta)$. By the same DDA algorithm (1.1), then each sensor can reach a set-valued average-consensus regarding the collection of sets $\Gamma(\theta) = [\Gamma_\pi^1(\theta), \ldots, \Gamma_\pi^n(\theta)]$. As each observed Markov chain X_k^i will have a stationary distribution belonging to exactly one element in the set $\Gamma_\pi^i(\theta)$, then under (1.1) the collection of values that the estimate s^i will converge weakly to is given by

$$\mathcal{P}^i(\theta) = \bigcup_{\pi^j \in \Gamma_\pi^j(\theta)} \left\{ \sum_{j=1}^{n} \psi_{ij}\pi^j(\theta) \right\}, \tag{2.3}$$

where $\Lambda = (\psi_{ij})$ is a constant matrix later defined in (6.12). The set-valued average-consensus $\bar{\mathcal{P}}(\theta)$ may likewise be expressed by

$$\bar{\mathcal{P}}(\theta) = \bigcup_{\pi^i \in \Gamma_\pi^i(\theta)} \left\{ \frac{1}{n} \sum_{i=1}^{n} \pi^i(\theta) \right\}. \tag{2.4}$$

Due to the linearity of (1.1), the assumption of set-valued transition matrices $\mathcal{A}^i(\theta)$ is in fact equivalent to considering set-valued exchange graphs $\{\mathcal{G}^v, \mathcal{G}^o\}$, since different exchange graphs result in each sensor's estimation of a different linear combination of the time-varying singletons, $\{\pi^1(\theta_k), \ldots, \pi^n(\theta_k)\}$. Although some of these combinations may imply the formation of a consensus (i.e., $\lim_{t\to\infty} s^i(t) = s^j(t)$ for all $i, j \in \mathcal{V}$ and each $\theta \in \mathcal{M}$), this is certainly not in general true, and thus this again leads to consideration of consensus formation not to a fixed point but rather to a set such as (2.3) or (2.4). We next consider the case of set-valued transition matrices $\mathcal{A}^i(\theta)$ and demonstrate that, in the limit as μ vanishes, the switching ODE (2.1) is replaced by a switching differential inclusion.

THEOREM 2.4. *Assume the conditions of Theorem* 2.1 *with the following modification. In lieu of the irreducibility and aperiodicity assumption of $A(\theta)$ in* (A), *assume for each $\theta \in \mathcal{M}$ there is an invariant measure $\pi \in \Gamma_\pi(\theta)$ such that*

$$\operatorname{dist}\left(\frac{1}{n} \sum_{k=\ell}^{n+\ell-1} E_\ell X_k(\theta), \Gamma_\pi(\theta) \right) \to 0 \quad \text{in probability as} \quad n \to \infty,$$

where $\operatorname{dist}(x, B) = \inf_{y\in B} |x - y|$ *is the usual distance function. The conclusion of Theorem* 2.1 *is then modified as $(s^\mu(\cdot), \theta^\mu(\cdot))$ converges weakly to $(s(\cdot), \theta(\cdot))$ such that*

$$\frac{ds(t)}{dt} \in -Hs(t) + \mathcal{W}^o \Gamma_\pi(\theta(t)). \tag{2.5}$$

We note that the conditional set of estimates reached by each sensor may be viewed as the result of uncertainty regarding the observed Markovian transition matrices, or in other words uncertainty in the signal dynamics. Equivalently, the conditional set of estimates can be seen as the result of uncertainty in the network exchange graphs, and hence we use set-valued parameters rather than singletons.

3. Scaled tracking error and cumulative distribution function (CDF) estimation. Since the network consensus converges to a stochastic process, it is important to examine the asymptotic convergence rate of the algorithm (1.1). We proceed here with the study of the scaled tracking errors of the sensor state-values s_k. Then we consider consensus formation when sensors estimate the empirical CDF instead of the probability mass function.

We show in section 6 that (2.1) implies the sensor estimates $s(t)$ will converge weakly to a stochastically switching steady-state $\Lambda\pi(\theta(t))$ for some fixed *equilibrium matrix* $\Lambda \in \mathbf{R}^{Sn\times Sn}$ as expressed in (6.12). We then define the scaled tracking error

$$v_k = \frac{s_k - \Lambda E(\pi(\theta_k))}{\sqrt{\mu}}, \tag{3.1}$$

where s_k is the vector of state-value estimates obtained from (1.1).

Next, assume that there exists a constant $k_\mu > 0$ sufficiently large such that $\{v_k : k \geq k_\mu\}$ is tight, which implies that the normalized error process scaled by

$1/\sqrt{\mu}$ does not blow up as $k \to \infty$ and $\mu \to 0$. Let us explain this assumption. The requirement $k \geq k_\mu$ is owing to the effect of initial conditions; the sequence of scaled errors needs to have some time to settle down. Recall that $\{v_k : k \geq k_\mu\}$ is tight if for any $\delta > 0$ there is a K_δ such that for all $n \geq \max(k_\mu, K_\delta)$, $P(|v_k| \geq K_\delta) \leq \delta$. To verify this condition, by virtue of Chebyshev's inequality, it suffices to have a suitable Liapunov function $\widetilde{U}(\cdot)$ such that $E\widetilde{U}(v_k)$ is bounded. If the random noise is uncorrelated, this boundedness can be obtained easily. If correlated noise is involved, tightness is established using the perturbed Liapunov function approach [51] (see also [27, Chapter 10] for detailed discussion and references). In this approach, small perturbations are added to the Liapunov function to deal with the correlation, and the perturbations are constructed so as to result in the needed cancellations. We omit the verbatim proof and refer the reader to the aforementioned references for further details.

THEOREM 3.1. *Assume the conditions of Theorem* 2.1. *Then the interpolated sequence of iterates $v^\mu(\cdot)$ defined by $v^\mu(t) = v_k$ for $t \in [(k-k_\mu)\mu, ((k+1)-k_\mu)\mu)$ converges weakly to a solution $v(\cdot)$ of the switching diffusion*

$$dv(t) = -Hv(t)dt + (\mathcal{W}^o \Sigma(\theta(t)) \mathcal{W}^{o\ \prime})^{1/2} dw, \tag{3.2}$$

where $w(\cdot)$ is a standard Brownian motion and for a fixed θ, $\Sigma(\theta)$ is the covariance given by

$$\Sigma(\theta) = \lim_{n\to\infty} \frac{1}{n} E \sum_{k=\ell}^{n+\ell-1} \sum_{j=\ell}^{n+\ell-1} (X_k(\theta) - EX_k(\theta))(X_j(\theta) - EX_j(\theta))'. \tag{3.3}$$

From (3.2) it is clear the sensor averaging weights $\mathcal{W}^o$ have a direct effect on the sensor's tracking error; in particular we see that any scaling of the averaging weights $\mathcal{W}^o$ implies the same scaling of the sensor diffusion process. We further detail this relation in section 4.

Average-consensus on the cumulative distribution function (CDF). So far we have assumed each sensor $j \in \mathcal{V}$ observes the state of a fast Markov chain X^j and tracks by (1.2) the associated stationary probability mass function $\pi^j(\theta(t))$. We now consider, for a given $j \in \mathcal{V}$, the approximation of the CDF of X_k^j by means of empirical measures. For each $\theta \in \mathcal{M}$, the CDF associated with X^j is denoted by $\Pi^j(\theta, x)$ for any $x \in \mathbf{R}^S$. For any $j \in \mathcal{V}$, $0 < T < \infty$, and any $x \in \mathbf{R}^S$, define the empirical measure

$$\eta_k = \frac{1}{k} \sum_{k_1=0}^{k-1} I_{\{X_{k_1}^j \leq x\}}, \ 0 \leq k \leq \frac{T}{\varepsilon}.$$

Note that $y \leq x$, with $x = (x^\iota) \in \mathbf{R}^S$ and $y = (y^\iota) \in \mathbf{R}^S$, is understood to hold componentwise (i.e., $y^\iota \leq x^\iota$ for $\iota = 1, \ldots, S$). The sequence η_k may be written recursively as

$$\eta_{k+1} = \eta_k - \frac{1}{k+1}\eta_k + \frac{1}{k+1} I_{\{X_k^j \leq x\}}.$$

So for sufficiently large k, the empirical CDF can be estimated as

$$\eta_{k+1} = \eta_k + \mu I_{\{X_k^j \leq x\}} \tag{3.4}$$

for arbitrarily small $\mu > 0$. Define the continuous-time interpolated process $\eta^\mu(t) = \eta_k$ for $t \in [\mu k, \mu k + \mu)$. In the above, for simplicity, we have suppressed the j-dependence in both η_k and $\eta^\mu(\cdot)$.

The following theorem is analogous to Theorem 2.1 but deals with average-consensus of the empirical CDF.

THEOREM 3.2. *Under condition* (A), $(\eta^\mu(\cdot), \theta^\mu(\cdot))$ *converges weakly to* $(\eta(\cdot), \theta(\cdot))$ *such that* $\eta(\cdot)$ *satisfies the switching ODE*

$$\dot{\eta} = \pi^j(\theta(t))\Pi^j(\theta(t), x). \tag{3.5}$$

Remark 3.3. In analogy to Theorem 3.1 (which dealt with the probability mass function), we now comment on the scaled tracking error for average-consensus on the CDF when using (3.4). To proceed, define

$$\xi^{\mu,j}(t,x) = \sqrt{\mu} \sum_{k=0}^{\lfloor t/\mu \rfloor - 1} [I_{\{X_k^j \le x\}} - \Pi^j(\theta_k, x)].$$

Using a combination of the techniques in the proof, Markov averaging, and standard results in centered and scaled errors for empirical measures, we can show that $\xi^{\mu,j}(t,x)$ converges weakly to $\xi(t,x)$, a switching Brownian bridge process. However, unlike the proof of Theorem 2.1, we can no longer use the martingale problem formulation. Nevertheless, we can use the basic techniques of weak convergence; see [6] for the related study on Brownian bridge processes. We illustrate in section 5 a sample path of the Brownian bridge resulting from the sensor scaled tracking error when a consensus regarding the average $\bar{\Pi}(\theta(t)) = \frac{1}{n}\sum_{i=1}^n \Pi^i(\theta(t))$ is obtained via (1.1).

Compared to $\pi(\theta(t))$, estimation of $\Pi(\theta(t))$ implies that one less state-value of X^i for each $i \in \mathcal{V}$ need be transmitted between sensors, thus reducing by a factor of $(1 - 1/S)$ any number of physical constraints on the sensor operations per communication link. If the number of states S is near unity, then estimation of $\Pi(\theta(t))$ may result in a significant reduction of the resources required for sensor communication and thus consensus. The results of estimating $\Pi(\theta(t))$ as compared to $\pi(\theta(t))$ are illustrated in section 5, where it is clear that estimation of the distribution function also results in greater scaled tracking error among the midstates of S than the tracking error among extremum states.

4. Average-consensus exchange graph conditions. The continuous-time sensor estimates $s(t)$ resulting from (1.1) were stated in section 2 to converge weakly to the solutions of (2.1) under conditions (A)–(C). In particular we assumed $\{\mathcal{G}^v, \mathcal{G}^o\}$ are such that all eigenvalues of $H = (\mathcal{L} + \mathcal{D}^o)$ have either positive real parts or non-negative real parts. These basic constraints on the network exchange graphs are not sufficient to ensure each sensor $i \in \mathcal{V}$ can asymptotically obtain the average-consensus estimate $\bar{\pi}(\theta(t))$. To proceed we assume the following property of $\mathcal{G}^v$ holds in addition to conditions (A)–(B) stated in section 2.

(C) The graph $\mathcal{G}^v$ is *strongly connected* (SC); thus for any two nodes $i, j \in \mathcal{V}$, there exists a path (a sequence of directed edges in $\mathcal{E}^v$) that starts at node i and ends at j.

The reason for condition (C) is to greatly simplify the weight conditions required for average-consensus. In fact, with regard to consensus formation, only when assuming this connectivity condition on $\mathcal{G}^v$ may we assume a completely decentralized structure in $\mathcal{G}^o$. To see this, and as well to motivate condition (C), we note that if $\mathcal{E}^v$ were not SC, then there would exist a nonempty set of ordered pairs $\mathcal{V}_{\setminus SC} = \{(i_1, j_1)\ldots, (i_h, j_h)\}$ for each element of which $\mathcal{E}^v$ contains no path. To then reach consensus by DDA would require $(i_l, j_l) \in \mathcal{E}^o$ for each $(i_l, j_l) \in \mathcal{V}_{\setminus SC}$, thus implying a nonscalable and "centralized" structure in $\mathcal{G}^o$. Furthermore, to ensure a network agreement each node contained in an element of $\mathcal{V}_{\setminus SC}$ may require individual treatment. Although this is possible, we assume $\mathcal{G}^v$ is SC to avoid centralization assumptions on $\mathcal{G}^o$ and as well avoid a more detailed scrutiny that may be left as future research.

It is emphasized that for generality we seek to reach the average-consensus $\bar{\pi}(\theta(t))$ under the condition that in each regime the observed stationary distributions $\{\pi^1(\theta(t)), \ldots, \pi^n(\theta(t))\}$ are *distinct*, that is,

$$\pi^i(\theta(t)) \neq \pi^j(\theta(t)) \quad \text{for all } i, j \in \mathcal{V} \text{ and any } \theta(t) \in \mathcal{M}. \tag{4.1}$$

If it were specifically known that two sensors may observe the *same* stationary distribution, i.e., $\pi^i(\theta^*) = \pi^j(\theta^*)$ for some $i, j \in \mathcal{V}$ and $\theta^* \in \mathcal{M}$, then the average-consensus requirements stated in the lemma below could be relaxed conditional on the fact that $\theta(t) = \theta^*$. We consider the most general case and do not make any assumptions regarding equality among the $\pi^i(\theta)$; in other words we assume (4.1) holds.

LEMMA 4.1. *Assuming conditions* (A)–(C)*, each sensor estimate based on the discrete-time algorithm* (1.1) *will in the limit as* $\mu \to 0$ *possess the steady-state* $\bar{\pi}(\theta(t))$ *conditional on* $\theta(t)$ *if and only if one of the following two conditions holds:*

- *Condition* (1): $0 \preccurlyeq (\mathcal{L} + \mathcal{D}^o)$*,* $\mathcal{L}$ *is balanced, and* $\mathcal{W}^o = \alpha(\mathcal{L} + \mathcal{D}^o)$*, where* $\alpha \in \mathbf{R}$.
- *Condition* (2): $0 \prec (\mathcal{L} + \mathcal{D}^o)$*, and the following equation holds:*

  $$(\mathcal{L} + \mathcal{D}^o)^{-1}\mathcal{W}^o = \Lambda \tag{4.2}$$

  for any matrix Λ *with the form*

  $$\Lambda = I + \beta(\mathbb{1}\mathbb{1}' - nI), \quad \beta = \mathrm{diag}(\beta_1 I_{S\times S}, \ldots, \beta_n I_{S\times S}) \in \mathbf{R}^{Sn\times Sn}, \tag{4.3}$$

  where β *is nonsingular.*

Remark 4.2. We here clarify that under condition (C), if $0 \preccurlyeq (\mathcal{L} + \mathcal{D}^o)$, then we have $\mathcal{D}^o = 0$ and the matrix $H = \mathcal{L} + \mathcal{D}^o$ will have only *one* eigenvalue with a real part equal to zero, and the corresponding eigenvector will be $\mathbb{1}$. To see this requires only noting that $\mathcal{D}^o$ is diagonal and $\mathcal{L}v = 0$ if and only if $v = \mathbb{1}$ when $\mathcal{G}^v$ is SC [31]; the result that $Hv = 0$ if and only if $v = \mathbb{1}$ follows immediately, as does the requirement $\mathcal{D}^o = 0$. Thus under Condition (1) and condition (C) we have both $\mathcal{W}^o = \alpha\mathcal{L}$ and $\mathcal{E}^o = \mathcal{E}^v$. Despite this simplification, we may still write $\mathcal{W}^o = \alpha(\mathcal{L}+\mathcal{D}^o)$ since it technically holds and also maintains a clear notational consistency between Conditions (1) and (2).

Rationale for Lemma 4.1. Assume all sensor computations possess the linear form (1.1). Then the average-consensus $\bar{\pi}(\theta(t))$ can be obtained under Condition (1) if each sensor maintains, in addition to (1.1), two distinct estimates based on the following subalgorithms of (1.1):

1. the estimate $\hat{s}_k$ based on only the local tracking subalgorithm of (1.1),

   $$\hat{s}^i_{k+1} = \hat{s}^i_k + \mu(X^i_k - \hat{s}^i_k), \quad \hat{s}^i(0) = X^i(0),\ i \in \mathcal{V}; \tag{4.4}$$

2. the estimate $\underline{s}^i_k$ based only on the (static) Laplacian consensus subalgorithm of (1.1),

   $$\underline{s}_{k+1} = (I - \mu\mathcal{L})\underline{s}_k, \quad \underline{s}(0) = X(0).$$

As $\mu \to 0$, the local tracking estimates $\hat{s}_k^i$ converge weakly to $\pi^i(\theta(t))$ for each $i \in \mathcal{V}$. In comparison, it is well known [31] that for any graph Laplacian $0 \preccurlyeq \mathcal{L}$ with left-right eigenvectors ω_ℓ, ω_r corresponding to the zero eigenvalue and satisfying $\omega_\ell' \omega_r = 1$, $\omega_\ell > 0$, the estimate $\underline{s}_k^i$ will then approach a convex-consensus $\underline{s}_\infty = \omega_r \omega_\ell' \underline{s}(0) = \omega_r \omega_\ell' X(0)$ for all $i \in \mathcal{V}$. Regardless, the linear combination denoted

$$s^i(t)_{(1)} = \frac{1}{\alpha}(\underline{s}^i(t) - s^i(t)) + \hat{s}^i(t) \tag{4.5}$$

has a steady-state $\bar{\pi}(\theta(t))$ for each $i \in \mathcal{V}$; see section 6 for details.

Under Condition (2) the linear combination

$$s^i(t)_{(2)} = \frac{1}{n\beta_i}(s^i(t) + (n\beta_i - 1)\hat{s}^i(t)) \tag{4.6}$$

has a steady-state $\bar{\pi}(\theta(t))$ for each $i \in \mathcal{V}$, as follows from (6.12) in section 6. ☐

We make two remarks here.

1. The average-consensus estimate $s(t)_{(1)}$ in (4.5) assumes that $\mathcal{L}$ is balanced and that each sensor knows the scale factor α between $(\mathcal{L}+\mathcal{D}^o)$ and $\mathcal{W}^o$. In Condition (2), the state-value exchange graph Laplacian $\mathcal{L}$ need not be balanced, but each sensor $i \in \mathcal{V}$ must know the total network size n, as well as the ith diagonal element of the matrix β in (4.3). Computation of any diagonal element in β will by (4.2) require near complete knowledge of the exchange graphs $\{\mathcal{G}^o, \mathcal{G}^v\}$, and so under Condition (2) we must presume each sensor has this knowledge.

2. Under Condition (2), it follows by (4.6) that $s(t)_2 = s(t)$ if and only if $\beta = \frac{1}{n}I$. In this case the extra local tracking estimate $\hat{s}(t)$ is not required. However, when $\beta = \frac{1}{n}I$, (4.2) cannot hold unless $\mathcal{G}^o$ is complete; that is, $\mathcal{E}^o$ contains every ordered pair of $(i,j) \in \mathcal{V}$. Equivalently this implies every sensor observes all Markov chains X^i, $i \in \mathcal{V}$, in which case consensus may be formed trivially.

4.1. Necessary exchange graph edge sets. The necessary and sufficient condition for average-consensus to be obtained by sensors under (1.1) and its subalgorithms is that either one of the conditions in Lemma 4.1 holds. Each of these conditions specifically assume either $0 \prec H$ or $0 \preccurlyeq H$ (recall $H = \mathcal{L} + \mathcal{D}^o$). In both cases, the constraints posed on the exchange graphs $\{\mathcal{G}^o, \mathcal{G}^v\}$ are based on the requirement that for average-consensus Λ must have, in each row, identical nondiagonal terms that do not equal zero. By then supplementing the estimates $s(t)$ with the local tracking estimates $\hat{s}(t)$ and normalizing by $(nB)^{-1}$, the average $\bar{\pi}(\theta(t))$ is obtained under Condition (2). Under Condition (1) a similar procedure (4.5) results in the estimate $\bar{\pi}(\theta(t))$ at all nodes.

We now characterize the exchange graph edge sets $\{\mathcal{E}^o, \mathcal{E}^v\}$ for which either of the conditions stated in Lemma 4.1 is feasible. Without loss of generality we take $S = 1$, and we also define the neighborhood of sensor i as $\mathcal{E}_i^v = \{j : (i,j) \in \mathcal{E}^v\}$ and the *complementary* edge set of $\mathcal{E}$ as $\bar{\mathcal{E}}$; that is, $(i,j) \in \bar{\mathcal{E}}$ if and only if $(i,j) \notin \mathcal{E}$. We assume that no sensor receives information directly from all other sensors. Otherwise we would have a centralized data fusion problem that would render a distributed algorithm such as (1.1) unnecessary. As a consequence of this assumption each row of either matrix $\{\mathcal{W}^o, \mathcal{W}^v\}$ has at least one nondiagonal element equal to zero; thus there exists for each row i an element j_i^o such that $(i, j_i^o) \notin \mathcal{E}^o$ and an element j_i^v such that $(i, j_i^v) \notin \mathcal{E}^v$.

LEMMA 4.3. *If $\mathcal{E}^o = \mathcal{E}^v$, then average-consensus is possible only if Condition* (1) *of Lemma* 4.1 *holds.*

Proof. We prove the above statement by showing (4.2) is infeasible when $0 \prec H$ and $\mathcal{E}^o = \mathcal{E}^v$. Since $0 \prec H$ we can rearrange (4.2) such that $\mathcal{W}^o = H\Lambda$. Denote the ith row of H by the row vector w_i. Since each row of $\mathcal{W}^v$ has at least one zero nondiagonal element, there exists an element j_i such that $w_{j_i} = 0$ for each row $i \in \mathcal{V}$. Since $\mathcal{E}^o = \mathcal{E}^v$ we then have by (4.3) and (4.2) that $\mathbb{1}'\beta w_i' = 0$ for each row i. This implies the rows of H are linearly dependent and thus H has an eigenvalue at 0, which contradicts our assumption $0 \prec H$. Thus (4.2) is infeasible when $\mathcal{E}^o = \mathcal{E}^v$ and $0 \prec H$. On the other hand, if $\mathcal{E}^o = \mathcal{E}^v$, then we may set $\mathcal{W}^o = \alpha(\mathcal{L} + \mathcal{D}^o)$ and Condition (1) holds if $\mathcal{L}$ is balanced. ☐

LEMMA 4.4. *If $\mathcal{E}^o \subset \mathcal{E}^v$, then average-consensus is not possible.*

Proof. If $\mathcal{E}^o = \mathcal{E}^v$, then by Lemma 4.3 average-consensus is possible only if $\mathcal{W}^o = \alpha(\mathcal{L} + \mathcal{D}^o)$. If $\mathcal{E}^o \subset \mathcal{E}^v$, then $\mathcal{W}^o = \alpha(\mathcal{L} + \mathcal{D}^o)$ cannot hold and so average-consensus is not possible. ☐

LEMMA 4.5. *If $\mathcal{E}^o \neq \mathcal{E}^v$, then average-consensus is possible only if Condition* (2) *of Lemma* 4.1 *holds and $\mathcal{E}_i^o \supseteq \bar{\mathcal{E}}_i^v$ for at least one sensor $i \in \mathcal{V}$.*

Proof. By our assumption of a decentralized network, there exists for each row i an element j_i^o such that $(i, j_i^o) \notin \mathcal{E}^o$ and an element j_i^v such that $(i, j_i^v) \notin \mathcal{E}^v$. For any row i, if $j_i^o = j_i^v$, then we have $\mathbb{1}'\beta w_i' = 0$. By Lemma 4.3, if $\mathbb{1}'\beta w_i' = 0$ holds for all $i \in \mathcal{V}$, then average-consensus is infeasible when $\mathcal{E}^o \neq \mathcal{E}^v$; thus there must be at least one row i^* such that $\mathbb{1}'\beta w_{i^*}' \neq 0$. This implies $(i^*, j_{i^*}^v) \in \mathcal{E}_{i^*}^o$ for at least one $j_{i^*}^v$ such that $(i^*, j_{i^*}^v) \notin \mathcal{E}_{i^*}^v$, which implies $(i^*, j_{i^*}^v) \in \mathcal{E}_{i^*}^o$ for all $j_{i^*}^v$ such that $(i^*, j_{i^*}^v) \notin \mathcal{E}_{i^*}^v$. Thus $\mathcal{E}_{i^*}^o \supseteq \bar{\mathcal{E}}_{i^*}^v$. On the other hand, Condition (1) of Lemma 4.1 cannot be satisfied if $\mathcal{E}^o \neq \mathcal{E}^v$; thus in this case we require Condition (2). ☐

The statement $\mathcal{E}_i^o \supseteq \bar{\mathcal{E}}_i^v$ implies that every sensor that does not send sensor i state-value data must send sensor i observation data. For sparse networks we might presume the cardinality of $\mathcal{E}^v$ is $|\mathcal{E}^v| \ll n$, and thus when $\mathcal{E}^o \neq \mathcal{E}^v$ by Lemma 4.5 an average-consensus requires $|\mathcal{E}_{i^*}^o| \approx n$ for some $i^* \in \mathcal{V}$. We note that this implies either that a large subgroup of sensors $h \subset \mathcal{V}$ sends observation data to sensor i^*, or equivalently that sensor i^* can itself observe the Markov chains that are observed by sensors in h.

4.2. Minimum trajectory adaptation rates. We now derive the minimum asymptotic exponential rate of the sensor trajectories to the average-consensus $\bar{\pi}(\theta)$ for any fixed $\theta \in \mathcal{M}$. We define a sensor trajectory as any solution to (2.1) conditional on $\pi(\theta)$ as well as the initial conditions $s(0)$.

Case 1: $\mathbf{0 \preccurlyeq \boldsymbol{H}}$. When $0 \preccurlyeq H$, by Lemma 4.1 average-consensus requires that $\mathcal{L}$ be balanced and $\mathcal{W}^o = \alpha\mathcal{L}$ for some $\alpha \in \mathbf{R}$ (see Remark 4.2). For any fixed $\theta \in \mathcal{M}$ we have $\frac{d\pi(\theta)}{dt} = 0$, and thus (2.1) implies

$$\frac{dy(t)}{dt} = -\mathcal{L}y(t), \quad y(t) = s(t) - \alpha\pi(\theta).$$

Similar to (4.5) we define

$$\breve{s}(t)_{(1)} = \frac{1}{\alpha}(\underline{s}(t) - y(t)) = e^{-\mathcal{L}t}\pi(\theta), \tag{4.7}$$

where the second equality holds since $\underline{s}(0) = s(0)$. Since $y(t)$ is a function of $\pi(\theta)$, and $\pi(\theta)$ is unknown to the sensors, we indicate by the symbol $\breve{}$ that $\breve{s}(t)_{(1)}$ is not an actual sensor estimate. We use $\breve{s}(t)_1$ simply as a mathematical bounding device.

The initial conditions $\underline{s}(0) = s(0)$ imply $\breve{s}(0)_{(1)} = \pi(\theta)$, and thus (4.7) implies $\frac{d\breve{s}(t)_{(1)}}{dt} = -\mathcal{L}\breve{s}(t)_{(1)}$. As follows from Theorem 8 in [31], the Euclidean norm $||\cdot||$ of the sensor *disagreement measure* $\breve{\delta}(t)_{(1)} = \breve{s}(t)_{(1)} - \bar{\pi}(\theta)$ then vanishes exponentially at the minimum rate $\lambda_2(\tilde{\mathcal{L}})$, where $\lambda_2(\cdot)$ denotes the smallest positive real part of the eigenvalues and $\tilde{\mathcal{L}}$ is the Laplacian of the *mirror graph* $\tilde{\mathcal{G}}^v$ induced by $\mathcal{G}^v$. Specifically,

$$\tilde{\mathcal{L}} = \frac{1}{2}(\mathcal{L} + \mathcal{L}'), \quad \tilde{\mathcal{G}}^v = \left\{ \mathcal{V}, \mathcal{E} \cup \mathcal{E}', \frac{1}{2}(\mathcal{W}^v + (\mathcal{W}^v)') \right\}, \tag{4.8}$$

where $(i,j) \in \mathcal{E}'$ if and only if $(j,i) \in \mathcal{E}$. For brevity we denote $\lambda_2 = \lambda_2(\tilde{\mathcal{L}})$.

It is clear by (4.5) and (4.7), together with the initial conditions $\underline{s}(0) = s(0)$, that $\breve{s}(t)_{(1)} - s(t)_{(1)} = e^{-\gamma t}(s(0) - \pi(\theta))$. Thus $s(t)_{(1)}$ approaches $\breve{s}(t)_{(1)}$ at an exponential rate γ. We now define $\breve{s}(0)_{(1)} = s(0) + \pi(\theta)$; in this case $s(t)_{(1)}$ approaches $\breve{s}(t)_{(1)} + s(0)$ asymptotically at an exponential rate γ. Together with $s(0)_{(1)} = s(0)$ this implies the sensor disagreement measure $\delta(t)_{(1)} = s(t)_{(1)} - \bar{\pi}(\theta)$ will diminish exponentially with the minimum asymptotic rate $r = \min(\gamma, \lambda_2)$,

$$\begin{aligned} ||\delta(t)_{(1)}|| &= ||s(t)_{(1)} - \bar{\pi}(\theta)|| \\ &\le ||s(t)_{(1)} - \breve{s}(t)_{(1)}|| + ||\breve{s}(t)_{(1)} - \bar{\pi}(\theta)|| \\ &\le ||s_0 - \pi(\theta)||e^{-\gamma t} + ||s_0 + \pi(\theta) - \bar{\pi}(\theta) - \bar{s}_0||e^{-\lambda_2 t} + ||s(0)|| \\ &\le (||s_0 - \pi(\theta)|| + ||s_0 - \bar{\pi}(\theta)||)e^{-rt} + ||\bar{s}_0 - \pi(\theta)||e^{-\lambda_2 t} + ||s(0)|| \\ &\le (||s_0 - \pi(\theta)|| + ||\delta(0)_{(1)}|| + ||\bar{s}(0) - \pi(\theta)||)e^{-rt} + ||s(0)||. \end{aligned} \tag{4.9}$$

We note that since $s(0)$, θ, and $\pi(\theta)$ are independent of the exchange graphs $\{\mathcal{G}^o, \mathcal{G}^v\}$, we obtain by (4.9) that $r = \min(\gamma, \lambda_2)$ is the minimum asymptotic adaptation rate of $s(t)_{(1)}$ to $\bar{\pi}(\theta)$ for any fixed $\theta \in \mathcal{M}$.

Case 2: $\mathbf{0} \prec \boldsymbol{H}$. We now show that $\delta(t)_{(2)} = s(t)_{(2)} - \bar{\pi}(\theta)$ vanishes exponentially with the minimum asymptotic rate $r = \min(\gamma, \lambda_2)$, where $\lambda_2 = \lambda_2(H)$. Let $\Lambda = H^{-1}\mathcal{W}^o$ and define $y(t) = s(t) - \Lambda\pi(\theta)$; we then have

$$\frac{dy(t)}{dt} = -Hy(t),$$

and thus $s(t) - \Lambda\pi(\theta)$ vanishes at a minimum rate λ_2. Independent of $s(t)$, it is clear by (4.4) that the trajectory of $\hat{s}(t)$ converges to $\pi(\theta)$ at an exponential rate γ; thus,

$$\begin{aligned} ||\delta(t)_{(2)}|| = ||s(t)_{(2)} - \bar{\pi}(\theta)|| &= \big|\big|(n\beta)^{-1}\big(s(t) - \Lambda\pi(\theta) + (n\beta - I)(\hat{s}(t) - \pi(\theta))\big)\big|\big| \\ &\le \|\tfrac{s(t)-\Lambda\pi(\theta)}{n\beta}\| + \|\tfrac{(n\beta-I)(\hat{s}(t)-\pi(\theta))}{n\beta}\| \\ &\le \|\tfrac{s(0)-\Lambda\pi(\theta)}{n\beta}\|e^{-\lambda_2 t} + \|\tfrac{(n\beta-I)(\hat{s}(0)-\pi(\theta))}{n\beta}\|e^{-\gamma t} \\ &\le \left(\|\tfrac{s(0)-\Lambda\pi(\theta)}{n\beta}\| + \|\tfrac{(n\beta-I)(\hat{s}(0)-\pi(\theta))}{n\beta}\|\right)e^{-rt}. \end{aligned} \tag{4.10}$$

Again the values of the initial conditions, θ, and $\pi(\theta)$ are independent of the weight matrices $\{\mathcal{W}^o, \mathcal{W}^v\}$. Thus (4.10) yields the minimum asymptotic adaptation rate r of $s(t)_{(2)}$ to $\bar{\pi}(\theta)$ for any fixed $\theta \in \mathcal{M}$.

Together the above cases yield $r = \min(\gamma, \lambda_2)$ as the minimum asymptotic adaptation rate of the adjusted sensor estimates to $\bar{\pi}(\theta)$ for any fixed $\theta \in \mathcal{M}$, where we define

$$\lambda_2 = \begin{cases} \lambda_2(\tilde{\mathcal{L}}) & \text{if } 0 \preccurlyeq H, \\ \lambda_2(H) & \text{if } 0 \prec H. \end{cases} \tag{4.11}$$

4.3. Adaptation rate and sensor uncertainty. Let $\gamma_2 = m\gamma_1$ for some scale factor $m > 1$. Assuming the weight matrices $\{\mathcal{W}^o, \mathcal{W}^v\}$ and observed parameters $\{X, \theta\}$ are unaffected by the scaling of γ, it is clear that the estimates $\{s(t)_{(1)}, s(t)_{(2)}\}$ will possess

(a) a minimum adaptation rate $r_2 = \min(\gamma_2, \lambda_2)$ that is greater than or equal to $r_1 = \min(\gamma_1, \lambda_2)$,

(b) increased uncertainty, measured here as the sampled scaled error averaged across all sensors and sampling period $[T_0, T_1]$,

$$U_{\text{samp}} = U(T_0, T_1) = \frac{1}{T_1 - T_0 + 1} \sum_{k=T_0}^{T_1} \frac{\mathbb{1}'|v_k - s(t)|}{n} \in \mathbf{R}^S. \tag{4.12}$$

Assume that we assign to each $\pi^i(\theta)$ some prior distribution $\zeta(\cdot)$. That is, for all $\theta \in \mathcal{M}$, $i \in \mathcal{V}$, and $j \in \{1, \ldots, S\}$ we have

$$P[\pi^i_j(\theta) = x] = \zeta(j, x), \;\; x \in \mathbf{R}. \tag{4.13}$$

In this case the covariance resulting from (4.13), denoted Σ_ζ, can be factored out of the expression $(\mathcal{W}^o \Sigma \mathcal{W}^{o\;\prime})$. Let $||\cdot||$ indicate the Frobenius norm and $\mathrm{tr}(\cdot)$ the trace of a matrix. We define $\bar{\sigma}^2 = ||(\mathcal{W}^o \Sigma \mathcal{W}^{o\;\prime})^{1/2}||^2 = \mathrm{tr}(\mathcal{W}^o \Sigma \mathcal{W}^{o\;\prime})$. Since $||(\mathcal{W}^o \mathcal{W}^{o\;\prime})^{1/2}||^2 = \mathrm{tr}(\mathcal{W}^o \mathcal{W}^{o\;\prime})$ we have, under (4.13),

$$\begin{aligned} \bar{\sigma}^2 = ||\Sigma_\zeta^{1/2}(\mathcal{W}^o \mathcal{W}^{o\;\prime})^{1/2}||^2 &\leq ||\Sigma_\zeta^{1/2}||^2 \,\mathrm{tr}(\mathcal{W}^o \mathcal{W}^{o\;\prime}) \\ &\propto \mathbb{1}'|\mathcal{W}^o|^2 \mathbb{1}. \end{aligned} \tag{4.14}$$

By (4.14) we may then expect the following ratio to hold for any two models that assume constant prior distributions ζ_1, ζ_2 for each $\pi^i(\theta)$:

$$\begin{aligned} \frac{U(T_0,T_1)_1}{U(T_0,T_1)_2} = \frac{\bar{\sigma}_1^2 + \bar{\sigma}_{\gamma_1}^2}{\bar{\sigma}_2^2 + \bar{\sigma}_{\gamma_2}^2} &= \left(\frac{||\Sigma_{\zeta_1}^{1/2}||}{||\Sigma_{\zeta_2}^{1/2}||}\right)^2 \frac{\mathrm{tr}(\mathcal{W}_1^o \mathcal{W}_1^{o\;\prime}) + nS\gamma_1^2}{\mathrm{tr}(\mathcal{W}_2^o \mathcal{W}_2^{o\;\prime}) + nS\gamma_2^2} \\ &\approx \left(\frac{||\Sigma_{\zeta_1}^{1/2}||}{||\Sigma_{\zeta_2}^{1/2}||}\right)^2 \frac{\mathbb{1}'(|\mathcal{W}_1^o|^2 + \gamma_1^2)\mathbb{1}}{\mathbb{1}'(|\mathcal{W}_2^o|^2 + \gamma_2^2)\mathbb{1}}, \end{aligned} \tag{4.15}$$

where we now let the underscore indicate model, with the exception of the sampling times T_0, T_1.

In our simulations the ratios of (4.15) were found to hold, as well as (4.14). Interestingly, when $\mathcal{W}^o$ was replaced with H the ratio (4.15) still held. Our simulations also consider the results of increasing λ_2. An increase in λ_2 has by logical necessity an effect similar to (a). We are unable to prove the necessity of (b), although our numerical examples support it.

Optimization. We now consider an optimization problem associated with maximizing λ_2 as given by (4.11) under either condition of Lemma 4.1. As shown in section 4.2, the minimum adaptation rate of the sensor trajectories is given by $r = \min(\gamma, \lambda_2)$. We consider γ a scale parameter and seek to maximize r with respect to λ_2; thus we here assume $\gamma > \lambda_2$.

Case 1: $\mathbf{0 \preccurlyeq} \boldsymbol{H}$. When $0 \preccurlyeq H$, average-consensus requires by Lemma 4.1 that $\mathcal{L}$ be balanced when assuming $\mathcal{G}^v$ is SC. The latter assumption implies the graph associated with the undirected edge set $(\mathcal{E}^v \cup \mathcal{E}^{v\;\prime})$ is also SC, where we recall $(i,j) \in \mathcal{E}^{v\;\prime}$ if and only if $(j,i) \in \mathcal{E}^v$. Denoting $\rho(\cdot)$ as spectral radius and $\mathcal{W} = I - \mu\tilde{\mathcal{L}}$, where $\tilde{\mathcal{L}}$ is defined by (4.8), we note that for all $(\tilde{\mathcal{L}}, \mu)$ such that $\mu\rho(\tilde{\mathcal{L}}) < 1$, the maximization of $\lambda_2(\tilde{\mathcal{L}})$ subject to $\rho(\mathcal{W}) < 1$ implies minimization of $||\mathcal{W} - \frac{1}{n}\mathbb{1}\mathbb{1}'||_2$. Thus solving (1.4) for $\mathcal{W}$ by the semidefinite program derived in [10],

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & \begin{bmatrix} sI & \mathcal{W} - \frac{1}{n}\mathbb{1}\mathbb{1}' \\ \mathcal{W} - \frac{1}{n}\mathbb{1}\mathbb{1}' & sI \end{bmatrix}, \\ & \mathcal{W} \in (\mathcal{E}^v \cup \mathcal{E}^{v\,\prime}),\ \mathcal{W}\mathbb{1} = \mathbb{1},\ \mathbb{1}'\mathcal{W} = \mathbb{1}' \end{array} \tag{4.16}$$

will yield both the fastest per-step convergence factor $r_{\text{step}}(\mathcal{W})$ as well as the maximum adaptation rate $r = \min(\gamma, \lambda_2)$. We find, however, that for small μ the solution to (4.16), denoted $\hat{\mathcal{W}}$, implies the optimal averaging weights $\hat{\mathcal{W}}^v$ approach infinity by the relation $\hat{\tilde{\mathcal{L}}} = \mu^{-1}(I - \hat{\mathcal{W}})$, in which case $\mu\rho(\tilde{\mathcal{L}}) < 1$ need not hold.

Consider then the problem equivalent to (1.4) when $\mu = 1$ and $\tilde{\mathcal{L}}$ is the optimization variable,

$$\begin{array}{ll} \text{minimize} & \| I - \tilde{\mathcal{L}} - \frac{1}{n}\mathbb{1}\mathbb{1}' \|_2 \\ \text{subject to} & \tilde{\mathcal{L}} \in (\mathcal{E}^v \cup \mathcal{E}^{v\,\prime}),\ \tilde{\mathcal{L}}\mathbb{1} = 0,\ \ \mathbb{1}'\tilde{\mathcal{L}} = 0. \end{array} \tag{4.17}$$

The corresponding solution $\hat{\mathcal{W}}$ to (4.16) then implies $\hat{\tilde{\mathcal{L}}} = I - \hat{\mathcal{W}}$ will solve (4.17). By convexity of the spectral norm, this implies $\tilde{\mathcal{L}}^* = c(I - \hat{\mathcal{W}})$ will, for any $c \in [0, \mu^{-1}]$, then solve

$$\begin{array}{ll} \text{minimize} & \| I - \mu\tilde{\mathcal{L}} - \frac{1}{n}\mathbb{1}\mathbb{1}' \|_2 \\ \text{subject to} & \tilde{\mathcal{L}} \in (\mathcal{E}^v \cup \mathcal{E}^{v\,\prime}),\ \tilde{\mathcal{L}}\mathbb{1} = 0,\ \ \mathbb{1}'\tilde{\mathcal{L}} = 0,\ \mathbb{1}'|\tilde{\mathcal{L}}|\mathbb{1} = c\mathbb{1}'|\hat{\tilde{\mathcal{L}}}|\mathbb{1}, \end{array} \tag{4.18}$$

as well as ensure $\mu\rho(\tilde{\mathcal{L}}) < 1$. The last constraint in (4.18) implies $\mathbb{1}'|\mathcal{L}|\mathbb{1} = c\mathbb{1}'|\hat{\mathcal{L}}|\mathbb{1}$; thus we find (4.16) yields the averaging weights $\{\mathcal{W}^{v*} : \tilde{\mathcal{L}}^* = c(I - \hat{\mathcal{W}})\}$ that will maximize λ_2 (and also, for fixed μ, optimize r_{step}) under an absolute total weight constraint. Besides its necessary role to bound $\mu\rho(\tilde{\mathcal{L}})$, the absolute weight constraint is of practical significance since, given $0 \preccurlyeq H$, an average-consensus formation requires $\mathcal{W}^o = \alpha\mathcal{L}$, and thus the total variance of the sensor diffusion $\bar{\sigma}^2 = ||(\mathcal{W}^o\Sigma\mathcal{W}^{o\,\prime})^{1/2}||^2$ will increase in squared proportion to the constant c. We note that although (4.16) maximizes λ_2 subject to an absolute weight constraint (thus bounding $\bar{\sigma}^2$ for any fixed Σ), the solution $\hat{\mathcal{W}}$ to (4.16) does not necessarily *minimize* $\bar{\sigma}^2$.

We pose below the optimization problem related to minimizing $\bar{\sigma}^2$ when subject to the average-consensus requirement $\mathcal{W}^o = \alpha\mathcal{L}$ as well as an absolute weight constraint,

$$\begin{array}{ll} \text{minimize} & \| (\mathcal{W}^o\Sigma\mathcal{W}^{o\,\prime})^{1/2} \|^2 \\ \text{subject to} & \mathcal{W}^o \in \mathcal{E}^v,\ \mathcal{W}^o\mathbb{1} = 0,\ \ \mathbb{1}'\mathcal{W}^o = 0,\ \mathbb{1}'|\mathcal{W}^o|\mathbb{1} = c\mathbb{1}'|\hat{\tilde{\mathcal{L}}}|\mathbb{1}. \end{array} \tag{4.19}$$

Given the constant prior distribution (4.13) we may factor Σ_ζ out of the above optimization function, thus yielding

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{i=1}^{n} \upsilon_i^2 \\ \text{subject to} & \mathcal{W}^o \in \mathcal{E}^v,\ \mathcal{W}^o\mathbb{1} = 0,\ \ \mathbb{1}'\mathcal{W}^o = 0,\ \mathbb{1}'|\mathcal{W}^o|\mathbb{1} = c\mathbb{1}'|\hat{\tilde{\mathcal{L}}}|\mathbb{1}, \end{array} \tag{4.20}$$

since $\| (\mathcal{W}^o\mathcal{W}^{o\,\prime})^{1/2} \|^2 = \sum_{i=1}^n \upsilon_i^2$, for any real matrix $\mathcal{W}^o$, where υ_i denote the singular values of $\mathcal{W}^o$. By minimizing $||I - \tilde{\mathcal{L}} - \frac{1}{n}\mathbb{1}\mathbb{1}'||_2$ as in (4.16), we then will again bound only $\sum_{i=1}^n \upsilon_i^2$ and thus also $\bar{\sigma}^2$. That is, given an absolute weight constraint when $0 \preccurlyeq H$, the optimal averaging weights for fast adaptation are not necessarily the optimal weights for minimum variance of the sensor scaled error.

Case 2: $\mathbf{0} \prec \boldsymbol{H}$. In this case an average-consensus requires that $\{\mathcal{W}^v, \mathcal{W}^o\}$ satisfy (4.2) for some nonsingular diagonal matrix β. Under this constraint, similar to above we wish to maximize $\lambda_2(H)$ while also minimizing $||\mathcal{W}^o||$ for fixed edge sets $\{\mathcal{E}^v, \mathcal{E}^o\}$. We do not explore this optimization problem further.

5. Numerical examples. We provide numerical examples to illustrate that Lemma 4.1 together with the theorems in sections 2 and 3 results in the weak convergence of each sensor's estimate to the average-consensus distribution $\bar{\pi}(\theta(t))$. As our communication networks we consider an example of either condition stated in Lemma 4.1, specifically the following:

- (Model 1) $\mathcal{W}^o = -\mathcal{L}$, and $\mathcal{W}^v$ is balanced. $\{\mathcal{G}^v, \mathcal{G}^o\}$ are such that $\mathcal{E}^v = \mathcal{E}^o$.
- (Model 2) $\{\mathcal{W}^o, \mathcal{W}^v\}$ satisfy (4.2). $\{\mathcal{G}^v, \mathcal{G}^o\}$ are such that $\mathcal{E}^v \neq \mathcal{E}^o$.

In both cases we set the network size at $n = 24$ sensors and consider $S = 40$ observed states. Thus each sensor $i \in \mathcal{V}$ will observe X^i in one of 40 states upon each iteration. In our simulation θ takes on 6 different states. For each of these regimes, the stationary distribution $\pi^i(\theta)$ of each Markov chain X^i was randomly generated from the uniform distribution and normalized, thus validating the constant a priori assumption required for (4.14)–(4.15).

The total number of communication links, that is, the total number of elements in $\mathcal{E}^o$ and $\mathcal{E}^v$, is fixed at $2n(n-1) = 1104$. We also fix the sum of absolute averaging weights at approximately the constant 2755,

$$|\mathcal{E}^v| + |\mathcal{E}^o| = 2n(n-1), \quad \mathbb{1}'(|\mathcal{L}| + |\mathcal{W}^o|)\mathbb{1} \approx 2755.$$

Neither of these constraints is necessary; they are set identical for either model to better compare the two different average-consensus graph conditions stated in Lemma 4.1. We note that under Condition (1) the minimum number of edges required for an average-consensus can be proved to be $2n$, whereas under Condition (2) this number is strictly greater than $2n$ for all $n \geq 4$.

Setting $\gamma = 10$ and $\mu = 10^{-9}$ for both models, the sample path of the sensor iterates $s(t)$ under (1.1) is plotted as the *unadjusted* sensor estimates; see Figure 5.1.

The estimates $s(t)_{(1)}$ and $s(t)_{(2)}$ are plotted as the *adjusted* sensor estimates under Models 1 and 2, respectively. We find that, in accordance with the adaptation rate $r = \min(\gamma, \lambda_2)$ proposed in section 4, Model 1 converges quicker to $\bar{\pi}(\theta)$ as compared to Model 2; see also Table 5.1. Also indicated by Table 5.1 is that our simulations imply the ratio (4.15) holds in approximation (we note that the two values under U_{samp} refer to sensor estimation of $\pi(\theta)$ and $\Pi(\theta)$, respectively).

As a trade-off to the improved adaptation rate, there is an increase in the scaled error under Model 1 when averaged across sensors. This is shown in Figure 5.2, where we have plotted for both models the sample scaled tracking error when estimating $\pi(\theta)$, as well as estimation of $\Pi(\theta)$ by the empirical measure described in section 3. We see that Model 1 results in an increased average scaled error as compared to Model 2. The exact relation between the scaled error and adaptation rate r is not explored here in detail.

The analytic and sample variances of the scaled tracking error, as measured relative to the sensor trajectories (6.11), are plotted in Figure 5.3. Our purpose here is to illustrate how the coefficients of the Brownian motion dw vary between models. For state ℓ in X^i and sensor $i \in \mathcal{V}$ under (1.1), the variance of the Brownian motion in (3.2) was computed analytically as a function of the averaging weights,

$$\sigma_{i\ell}^2 = \sum_{j=1}^{Sn} (\mathcal{W}^o \cdot \Sigma)^2_{((i-1)S+1+\ell)j}, \quad (\mathcal{W}^o \cdot \Sigma) \doteq (\mathcal{W}^o \Sigma \mathcal{W}^{o\,\prime})^{1/2}. \tag{5.1}$$

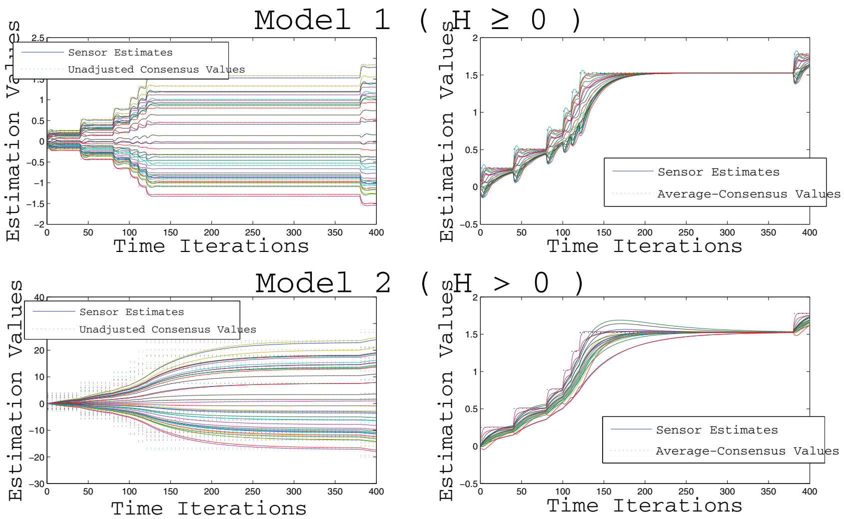

FIG. 5.1. *Sensor estimation of the average-consensus* $\bar{\pi}(\theta(t))$ *under both Models* 1 *and* 2*. For illustrative purposes we have, without loss of generality and for this figure only, set* $X(0) = 0$ *and magnified the sensor observations of* X^i *by a factor proportional to the number of previous states* θ *has occupied in the simulation.*

TABLE 5.1
Numerical results of the ratio comparisons. The average ratio between Models 1 *and* 2 *is* 0.2094*.*

Model	$r = \min(\gamma, \lambda_2)$	$\mathbb{1}'U_{\text{samp}}$	$\bar{\sigma}^2_{\text{model}}$	$\mathbb{1}'\lvert\mathcal{W}^o\rvert^2\mathbb{1}$	$\lvert\lvert\mathcal{W}^o\rvert\rvert^2$	$\mathbb{1}'\lvert H\rvert^2\mathbb{1}$
1	7.12	$3.563, 0.944$	$6.875\ 10^{-7}$	$442\ 10^6$	6343	$2.999\ 10^5$
2	1.44	$0.805, 0.207$	$1.556\ 10^{-7}$	$841.5\ 10^5$	1463	$1.3775\ 10^5$
Ratio $\frac{\text{model}_1}{\text{model}_2}$	0.1994	$0.2259, 0.2193$	0.2278	0.1843	0.2049	0.2177

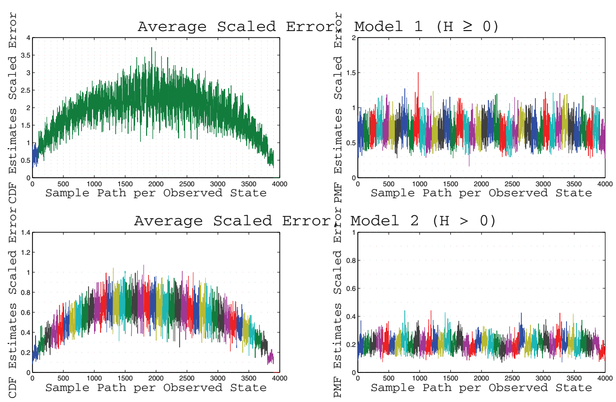

FIG. 5.2. *The absolute scaled error* U_{samp}*, plotted consecutively over the entire simulation time* $[T_0, T_1]$ *for each state in* S*. The results indicate a Brownian bridge across the state estimates* $(S = 40)$ *of the average cumulative distribution* $\bar{\Pi}(\theta(t))$*.*

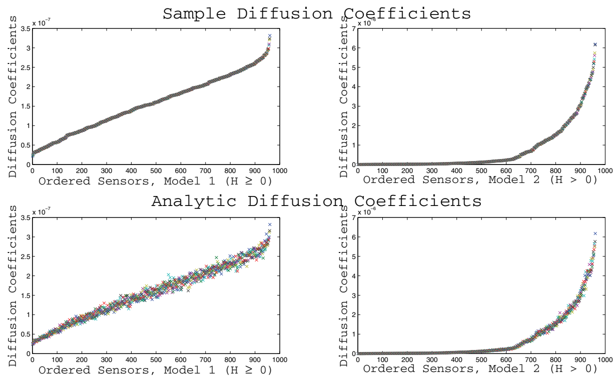

FIG. 5.3. *The analytic and sample variance $\sigma_{i\ell}$ of the scaled tracking error under both Models* 1 *and* 2, *as given by* (5.1). *A point exists for each sensor $i \in \mathcal{V}$ and observed state $\ell \in \{1, \ldots, S\}$. The results have been ordered according to the sample variances to better show the similarity between the analytic and simulated variances.*

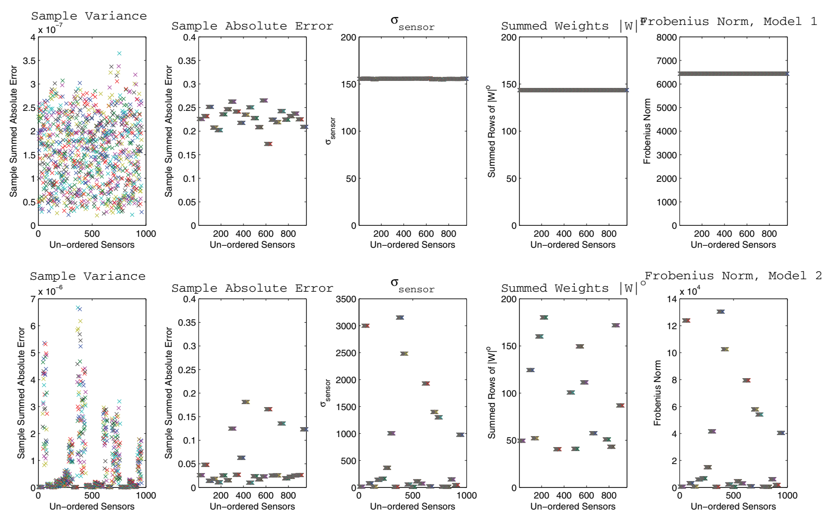

FIG. 5.4. *Unordered sensor diffusion coefficients, absolute error, and functions of weights. The similarities between each quantity are apparent by the localization of larger values under Model* 2, *as compared to the uniform values obtained from Model* 1. *The approximate ratio* (4.15) *is also seen here to generally hold* per sensor, *although only with $\mathcal{W}^o$ replaced by H as seen in Table* 5.1.

We have plotted in Figure 5.3 the sample values of $\sigma_{i\ell}$ in ascending order; the corresponding analytic values have been plotted in the same order to illustrate the similarity between the analytic and simulated results, thus verifying (3.2) as the correct expression for the scaled sensor tracking error. We note that $\bar{\sigma}^2 = \sum_{\ell=1}^{S} \sum_{i=1}^{n} \sigma_{i\ell}^2$.

Finally, in Figure 5.4, we have plotted various features of our simulations *per sensor*. It is apparent that the ratios between these models regarding analytic variance $\sigma_{sensor\ i}^2 = \sum_{\ell=1}^{S} \sigma_{i\ell}^2$, Frobenius norm $||\mathcal{W}^o||^2$, either summed squared weight values $\mathbb{1}'|\mathcal{W}^o|^2\mathbb{1}$ or $\mathbb{1}'|H|^2\mathbb{1}$, and the absolute sampled error U_{samp} hold in approximation.

In these numerical examples we employed the adaptation rate r and uncertainty measures (4.12) and $\bar{\sigma}^2$ to explain our results. We leave open the question of which model is superior in terms of maximizing r while minimizing $\bar{\sigma}^2$ and as well the weights and edges, or complexity thereof, required for $\{\mathcal{G}^o, \mathcal{G}^v\}$ to ensure average-consensus. This question is of importance since the features of $\{\mathcal{G}^o, \mathcal{G}^v\}$ can be naturally associated with the communication costs assumed of the sensor network [31].

6. Proofs of results.

Proof of Theorem 2.1. We divide the proof into several steps. Some of the steps are formulated as lemmas to improve clarity of the presentation.

Step 1. In this step, we show that a moment bound holds for the iterates under consideration. The assertion is stated as a lemma.

LEMMA 6.1. *For each $0 < T < \infty$, under the conditions of Theorem* 2.1,

$$\sup_{0 \le k \le T/\mu} E|s_k|^2 < \infty. \tag{6.1}$$

Proof of Lemma 6.1. Define $V(s) = s's/2$. The gradient and Hessian of $V(s)$ are given by $\nabla V(s) = s$ and $\nabla^2 V(s) = I$, respectively. Under condition (B) we assume H has no eigenvalues with negative real parts; hence there is a $\lambda(0) > 0$ such that $s'Hs \ge \lambda(0)V(s)$. In addition, since X_k is a conditional Markov chain with a finite state-space, it is bounded uniformly w.p.1, which implies that

$$\begin{aligned} s_k' \mathcal{W}^o X_k &\le \frac{1}{2}(|s_k|^2 + |\mathcal{W}^o|^2 |X_k|^2) \\ &\le K(V(s_k) + 1) \end{aligned}$$

for some $K > 0$. Here and hereafter, K is used to represent a generic constant, whose values may change for different appearances. Using (1.1) and the above estimates, we obtain

$$\begin{aligned} V(s_{k+1}) &= V(s_k) + \mu s_k'(-Hs_k + \mathcal{W}^0 X_k) + O(\mu^2)(V(s_k) + 1) \\ &\le V(s_k) - \lambda(0)\mu V(s_k) + \mu s_k' \mathcal{W}^o X_k + O(\mu^2)(V(s_k) + 1) \\ &\le (1 - \lambda(0)\mu)V(s_k) + O(\mu)(V(s_k) + 1). \end{aligned} \tag{6.2}$$

Taking expectation and iterating on the resulting sequence, we obtain

$$\begin{aligned} EV(s_{k+1}) \le (1 - \lambda(0)\mu)^{k+1} EV(s(0)) &+ O(\mu) \sum_{j=0}^{k} (1 - \lambda(0)\mu)^{k-j} EV(s_j) \\ &+ O(\mu) \sum_{j=0}^{k} (1 - \lambda(0)\mu)^{k-j}. \end{aligned}$$

An application of Gronwall's inequality yields that

$$EV(s_{k+1}) \le O(1) \exp(\mu k) \le O(1) \exp(\mu (T/\mu)) \le O(1).$$

Taking $\sup_{0 \le k \le T/\mu}$ on both sides above yields the desired result. □

First, note that by means of Chebyshev's inequality and Lemma 6.1, for any $\delta > 0$ sufficiently small, there is a $K_\delta = (1/\sqrt{\delta})$ sufficiently large such that

$$P(|s_k| \ge K_\delta) \le \frac{E|s_k|^2}{K_\delta^2} \le \frac{\sup_{0 \le k \le T/\mu} E|s_k|^2}{K_\delta^2} \le \frac{K}{K_\delta^2} \le K\delta.$$

Thus $\{s_k\}$ is tight. Since θ_k takes values in a finite set, it is also tight. Thus, we obtain the following corollary.

COROLLARY 6.2. *The pair of sequences* $\{s_k, \theta_k\}$ *is tight.*

Step 2. We show that the process $(s^\mu(\cdot), \theta^\mu(\cdot))$ is tight in $D([0,T] : \mathbf{R}^S \times \mathcal{M})$. Since the tightness of $\theta^\mu(\cdot)$ has been demonstrated in [53], we concentrated on the tightness of $s^\mu(\cdot)$. For any $\delta > 0$, $t > 0$, and $0 < u \le \delta$, we have

$$s^\mu(t+u) - s^\mu(t) = \mu \sum_{k=t/\mu}^{(t+u)/\mu} (-Hs_k + \mathcal{W}^o X_k).$$

Thus by repeating the proof of Lemma 6.1, we can show that

$$\begin{aligned} E_t^\mu |s^\mu(t+u) - s^\mu(t)|^2 &\le K\mu^2 \sum_{k=t/\mu}^{(t+u)/\mu - 1} (E_t^\mu |s_k|^2 + 1) \\ &\le K\mu^2 \left(\frac{t+u}{\mu} - \frac{t}{\mu} \right)^2 \le Ku^2 \le K\delta \end{aligned}$$

for u sufficiently small, where E_t^μ denotes the conditional expectation with respect to the σ-algebra $\mathbb{F}_t = \{s^\mu(u), \theta^\mu(u) : u \le t\}$. Thus

$$\lim_{\delta \to 0} \limsup_{\mu \to 0} E E_t^\mu |s^\mu(t+u) - s^\mu(t)|^2 = 0.$$

The claim then follows from the well-known tightness criteria [26, p. 47] (see also [27, sect. 7.3]).

Step 3. Characterization of the limit. Owing to Step 2, $(s^\mu(\cdot), \theta^\mu(\cdot))$ is tight. By Prohorov's theorem, we can extract a weakly convergent subsequence. Without loss of generality, we still index the selected sequence by μ and denote the limit by $s(\cdot), \theta(\cdot)$. By Skorohod representation, we may assume (with slight abuse of notation) $(s^\mu(\cdot), \theta^\mu(\cdot))$ converges to $(s(\cdot), \theta(\cdot))$ w.p.1 and the convergence is uniform in any bounded t interval. For each $t, u > 0$, we partition $[t, t+u]$ into small segments with the use of $k_\mu \to \infty$ as $\mu \to 0$ but $\delta_\mu := \mu k_\mu \to 0$ as $\mu \to 0$. Then we have

$$s^\mu(t+u) - s^\mu(t) = \sum_{l:l\delta_\mu = t}^{t+u} \delta_\mu \frac{1}{k_\mu} \sum_{k=lk_\mu}^{lk_\mu + k_\mu - 1} (-Hs_k + \mathcal{W}^o X_k),$$

where $\sum_{l:l\delta_\mu=t}^{t+u}$ denotes the sum over l in the range $t \le l\delta_\mu < t+u$. For the following analysis, it is crucial to recognize the two-time-scale structure of the algorithm. In the segment $lk_\mu \le k \le lk_\mu + k_\mu - 1$, compared with s_k and θ_k, X_k varies much faster. Thus, in this segment, s_k and θ_k can be viewed as fixed. As a result, although X_k depends on θ, the slowly varying θ_k enables us to treat X_k as a "noise" with θ_k fixed at a specific value θ. In the end, X_k will be averaged out and replaced by its stationary measure. More precisely, let $l\delta_\mu \to \widetilde{u}$ as $\mu \to 0$. Then for all $lk_\mu \le k \le lk_\mu + k_\mu - 1$, $\mu k \to \widetilde{u}$. To emphasize the θ dependence in X_k, we write it as $X_k(\theta_k)$. It then follows that

$$\begin{aligned} \frac{1}{k_\mu} \sum_{k=lk_\mu}^{lk_\mu + k_\mu - 1} \mathcal{W}^o X_k &= \frac{1}{k_\mu} \sum_{k=lk_\mu}^{lk_\mu + k_\mu - 1} \mathcal{W}^o X_k(\theta_k) \\ &= \frac{1}{k_\mu} \sum_{k=lk_\mu}^{lk_\mu + k_\mu - 1} \mathcal{W}^o X_k(\theta_{lk_\mu}) + o(1) \\ &\to \mathcal{W}^o \pi(\theta(\widetilde{u})) \ \text{ in probability as } \ \mu \to 0, \end{aligned}$$

where $o(1) \to 0$ in probability as $\mu \to 0$. The last line follows from the ergodicity of the θ dependent Markov chain $X_k(\theta)$. As a result

$$\sum_{l:l\delta_\mu=t}^{(t+u)} \delta_\mu \frac{1}{k_\mu} \sum_{k=lk_\mu}^{lk_\mu+k_\mu-1} \mathcal{W}^o X_k \to \int_t^{t+u} \mathcal{W}^o \pi(\theta(\widetilde{u}))d\widetilde{u}. \tag{6.3}$$

Likewise, we can work with the term involving s_k by using the continuity in s of the following expression, which leads to

$$\sum_{l:l\delta_\mu=t}^{(t+u)} \delta_\mu \frac{1}{k_\mu} \sum_{k=lk_\mu}^{lk_\mu+k_\mu-1} H s_k \to \int_t^{t+u} Hs(\widetilde{u})d\widetilde{u}. \tag{6.4}$$

Now combining (6.3) and (6.4), following the argument as in the proof of Theorem 4.5 in [51], it can be shown that $(s(\cdot), \theta(\cdot))$ is a solution of the martingale problem associated with the operator given by

$$\mathbb{L}f(s,\theta^i) = \nabla f'(s,\theta^i)[-Hs + \mathcal{W}^o\pi(\theta^i)] + Qf(s,\cdot)(\theta^i), \tag{6.5}$$

where

$$Qf(s,\cdot)(\theta^i) = \sum_{j=1}^m q_{ij} f(s,\theta^j).$$

The desired results thus follow. □

Proof of Theorem 2.2. (i) Define

$$\widetilde{s}^\mu(\cdot) = s^\mu(\cdot + t_\mu),\ \widetilde{\theta}^\mu(\cdot) = \theta^\mu(\cdot + t_\mu),$$

where t_μ is given in the statement of the theorem. Then $(\widetilde{s}^\mu(\cdot), \widetilde{\theta}^\mu(\cdot))$ is tight, which can be proved as in Theorem 2.1. For $0 < T < \infty$, extract a convergent subsequence $\{\widetilde{s}^\mu(\cdot), \widetilde{s}^\mu(\cdot - T)\}$ with limit denoted by $(s(\cdot), s_T(\cdot))$. Note that $s(0) = s_T(T)$.

Note that $\{s_k\}$ is tight, which can be proved as in Corollary 6.2. The tightness of $\{s_k\}$ then implies that $\{s_T(0)\}$ is tight. By using the following representation of the solution of the switched ODE and noting that T is arbitrary, it then follows that

$$\begin{aligned} s_T(T) &= \exp(-HT)s_T(0) + \sum_{i=1}^m \int_0^T \exp(-H(T-u))\mathcal{W}^0\pi(\theta^i)I_{\{\theta(u)=i\}}du \\ &= \exp(-HT)s_T(0) + \sum_{i=1}^m \int_0^T \exp(-H(T-u))du\mathcal{W}^0\pi(\theta^i)\nu_i \\ &\quad + \sum_{i=1}^m \int_0^T \exp(-H(T-u))\mathcal{W}^0\pi(\theta^i)[I_{\{\theta(u)=i\}} - \nu_i]du. \end{aligned} \tag{6.6}$$

(ii) We claim that as $T \to \infty$, the last term above goes to 0 in probability. To show this, it suffices to work with a fixed i. Define

$$\xi(T) = E\left|\int_0^T \exp(-H(T-u))\mathcal{W}^0\pi(\theta^i)[I_{\{\theta(u)=i\}} - \nu_i]du\right|^2.$$

Then it is readily seen that

$$
\begin{aligned}
(6.7)\qquad |\xi(T)| \le K \bigg| E\int_0^T & (\mathcal{W}^0\pi(\theta^i))' \exp(-H'(T-t))dt \\
& \times \int_0^t \exp(-H(T-t))\mathcal{W}^0\pi(\theta^i)[I_{\{\theta(u)=i\}} - \nu_i][I_{\{\theta(t)=i\}} - \nu_i]du \bigg| .
\end{aligned}
$$

By using the Markov property, it can be shown that

$$
\begin{aligned}
&\big|E[I_{\{\theta(u)=i\}} - \nu_i][I_{\{\theta(t)=i\}} - \nu_i]\big| \\
&\quad = |[P(\theta(u)=i) - \nu_i][P(\theta(t)=i|\theta(u)=i) - \nu_i]| \\
&\quad \le K\exp(-\kappa_0 u)\exp(-\kappa_0(t-u)) \le K\exp(-\kappa_0 t),
\end{aligned}
$$

where $\kappa_0 > 0$ is a constant representing the spectrum gap in the Markov chain owing to the irreducibility of Q (see [52, p. 300]). In the above and hereafter in the proof, we use K as a generic positive constant whose value may change for different apparentness. Since $-H$ is Hurwitz,

$$
\int_0^t |\exp(-H(T-u))|du \le \int_0^t \exp(-\lambda_H(T-u))du \le K\exp(-\lambda_H(T-t)),
$$

where $\lambda_H > 0$. Then we have

$$
\begin{aligned}
&\left| E\int_0^t \exp(-H(T-u))\mathcal{W}^0\pi(\theta^i)[I_{\{\theta(u)=i\}} - \nu_i][I_{\{\theta(t)=i\}} - \nu_i]du \right| \\
&\quad \le K\exp(-\lambda_H(T-t))\exp(-\kappa_0 t).
\end{aligned}
$$

Then

$$
|\xi(T)| \le \left| K\int_0^T |\mathcal{W}^0\pi(\theta^i)||\exp(-H'(T-t))|\exp(-\lambda_H(T-t))\exp(-\kappa_0 t)dt \right| .
$$

If $\lambda_H \le \kappa_0$, $\exp((\lambda_H - \kappa_0)t) \le 1$, so

$$
|\xi(T)| \le \left| K\int_0^T \exp(-\lambda_H(T-t))\exp(-\lambda_H T)dt \right| \le K\exp(-\lambda_H T) \to 0 \ \text{ as } \ T\to\infty.
$$

If $\lambda_H > \kappa_0$,

$$
\begin{aligned}
|\xi(T)| &\le K\int_0^T \exp(-\lambda_H(T-t))\exp(-(\lambda_H-\kappa_0)T)\exp(-\kappa_0 T)dt \\
&\le K\exp(-\kappa_0 T) \to 0 \ \text{ as } \ T\to\infty.
\end{aligned}
$$

Thus the claim in (ii) is verified.

(iii) The tightness of $\{s_T(0)\}$ together with $-H$ being Hurwitz implies the first term on the right-hand side of (6.6) tends 0 as $T\to\infty$. The second term in (6.6) converges to s_* as $T\to\infty$ by the hypothesis of this theorem. The desired result thus follows from the above and (ii). ∎

Proof of Theorem 2.4. The proof here is similar to Theorem 2.1 with the modification of the inclusion. We omit the proof and refer the reader to [27] for related results on averaging leading to differential inclusions. Note that the main difference of the above results compared with Theorem 2.1 is the following: In Theorem 2.1, condition (A) implies the existence of the unique invariant measure for each θ. Here this condition is relaxed. For Markov chains with multiple ergodic classes, we refer the reader to [25, Chap. 1] for further details.

Proof of Theorem 3.1. First, using (1.1),

$$\begin{aligned} s_{k+1} - \Lambda E\pi(\theta_{k+1}) = [s_k - \Lambda E\pi(\theta_k)] - \mu H[s_k - \Lambda E\pi(\theta_k)] \\ + \Lambda E[\pi(\theta_k) - \pi(\theta_{k+1})] + \mu \mathcal{W}^o(X_k - E\pi(\theta_k)). \end{aligned}$$

In the last line above, we have used the fact that $H\Lambda = \mathcal{W}^0$; this holds by necessity under the conditions of Lemma 4.1 (see (6.12) below). The definition of (3.1) thus leads to

$$v_{k+1} = v_k - \mu H v_k + \frac{\Lambda E(\pi(\theta_k) - \pi(\theta_{k+1}))}{\sqrt{\mu}} + \sqrt{\mu}\mathcal{W}^o(X_k - E\pi(\theta_k)). \tag{6.8}$$

Comparing (6.8) with equation (5.3) in [51], we see that (6.8) is simply a modified version of (5.3) in [51]. We can thus proceed as in the aforementioned paper. Since $\{v_k\}$ is not a priori bounded, it is pertinent to first take an N truncation [27, p. 284] for an arbitrary large N, work with the truncated process, establish the weak convergence of the truncated process, and finally let $N \to \infty$ to conclude the proof of the result. However, we assume that the iterates are bounded here to simplify the presentation.

Considering the interpolated sequence $\{v^\mu(\cdot)\}$, essentially the same argument as in [51] shows that for $t, u > 0$,

$$\sum_{k=t/\mu}^{(t+u)/\mu-1} \frac{\Lambda E(\pi(\theta_k) - \pi(\theta_{k+1}))}{\sqrt{\mu}} \to 0 \text{ in probability,}$$

so

$$v^\mu(t+u) = v^\mu(0) - \mu \sum_{k=t/\mu}^{(t+u)/\mu-1} H v_k + \sqrt{\mu} \sum_{k=t/\mu}^{(t+u)/\mu-1} \mathcal{W}^o(X_k - E\pi(\theta_k)) + o(1), \tag{6.9}$$

where $o(1) \to 0$ in probability uniformly in t as $\mu \to 0$. As in [51], it can be shown that

$$\sqrt{\mu} \sum_{k=0}^{t/\mu-1} (X_k - E\pi(\theta_k)) \text{ converges weakly to } \Sigma^{1/2}(\theta(t))w(t),$$

where $w(\cdot)$ is a standard Brownian motion, and $\Sigma(\theta)$ is the covariance given by (3.3). As a result,

$$\sqrt{\mu} \sum_{k=0}^{t/\mu-1} \mathcal{W}^o(X_k - E\pi(\theta_k)) \text{ converges weakly to } \mathcal{W}^o\Sigma^{1/2}(\theta(t))w(t)$$

by the well-known Slutsky theorem. The corresponding covariance thus becomes $\mathcal{W}^o\Sigma(\theta(t))\mathcal{W}^{o\,\prime}$.

To proceed, we show that the sequence $(v^\mu(\cdot), \theta^\mu(\cdot))$ converges weakly to $(v(\cdot), \theta(\cdot))$ such that the limit is a solution of the martingale problem with the operator given by

(6.10)
$$\mathbb{L}_v f(v, \theta^i) = -\nabla f'(v, \theta^i) H v + \frac{1}{2} \operatorname{tr}[\nabla^2 f(v, i) \mathcal{W}^o \Sigma(\theta^i) \mathcal{W}^{o\,\prime}] + Q f(v, \cdot)(\theta^i), \;\; \theta^i \in \mathcal{M}.$$

The rest of the proof is similar to that of [51]. The details are omitted. □

Proof of Theorem 3.2. The well-known Glivenko–Cantelli theorem (see [6, p. 103]) for the mixing process implies that

$$\frac{1}{k_\mu} \sum_{k=lk_\mu}^{lk_\mu + k_\mu - 1} I_{\{X_k(\theta^{j_1}) \le x\}} \to F(\theta^{j_1}, x) \;\; \text{w.p.1 as} \;\; \mu \to 0 \;\; \text{and hence} \;\; k_\mu \to \infty.$$

The Markovian structure implies that

$$(A(\theta^{j_1}))^{k - lk_\mu} \to \mathbb{1}\pi(\theta^{j_1}) \;\; \text{as} \;\; \mu \to 0.$$

The limit is a matrix with identical rows containing the stationary distribution $\pi(\theta^{j_1})$. Since $I_{\{\theta_{lk_\mu} = \theta^{j_1}\}}$ can be written as $I_{\{\theta^\mu(l\delta_\mu) = \theta^{j_1}\}}$, as $\mu \to 0$ and $k\delta_\mu \to \widetilde{u}$, we can show that

$$\begin{aligned} \frac{1}{k_\mu} \sum_{k=lk_\mu}^{lk_\mu + k_\mu - 1} E_{lk_\mu} I_{\{X_k^j(\theta) \le x\}} &\to \sum_{j_1=1}^{m} \pi^j(\theta^{j_1}) F(\theta^{j_1}, x) I_{\{\theta(\widetilde{u}) = \theta^{j_1}\}} \\ &= \pi^j(\theta(\widetilde{u})) F(\theta(\widetilde{u}), x). \end{aligned}$$

Thus we obtain that the limit $\eta(\cdot)$ satisfies

$$\eta(t) - \eta(u) = \int_u^t \pi^j(\theta(\widetilde{u})) F(\theta(\widetilde{u}), x) d\widetilde{u}.$$

Thus the desired result follows. □

Proof of Lemma 4.1 *continued (exact solutions to* (2.1)*).* We first derive

$$s(t) = e^{-Ht}(s(0) - \Lambda\pi(\theta(t))) + \Lambda\pi(\theta(t)) \tag{6.11}$$

as an exact solution to (2.1) conditional on the state $\theta(t)$ of $\theta \in \mathcal{M}$. The equilibrium matrix Λ is defined by the following function of $\{\mathcal{G}^v, \mathcal{G}^o\}$:

$$\Lambda = \begin{cases} (\mathcal{L} + \mathcal{D}^o)^{-1} \mathcal{W}^o & \text{if} \;\; 0 \prec (\mathcal{L} + \mathcal{D}^o), \\ (I - \omega_r \omega_\ell') & \text{if} \;\; 0 \preccurlyeq (\mathcal{L} + \mathcal{D}^o) \;\; \text{and} \;\; \mathcal{W}^o = \mathcal{L} + \mathcal{D}^o. \end{cases} \tag{6.12}$$

If $0 \prec H$, then the above results follow immediately by standard techniques of solving systems of first order linear ODEs. On the other hand, if $0 \preccurlyeq H$, then (6.12) is derived based on the conservation property,

$$\frac{d}{dt} \omega_\ell' s(t) = \omega_\ell' \frac{ds(t)}{dt} = -\omega_\ell' H s(t) + \omega_\ell' \mathcal{W}^o \pi(\theta(t)) = 0, \tag{6.13}$$

which follows under the assumption $\mathcal{W}^o = H$, since $\omega_\ell' H = 0$ by definition. From (6.13) it is clear that Λ must satisfy $\omega_\ell' \Lambda = 0$, whereas by solving $\frac{ds(t)}{dt} = 0$ given (6.11) we have $H\Lambda = \mathcal{W}^o$. These two constraints then uniquely determine the second steady-state in (6.12).

For any SC graph $\mathcal{G}^v$, if and only if $\mathcal{L}$ is balanced will $\omega_r\omega_\ell' = \frac{1}{n}\mathbb{1}\mathbb{1}'$ [31]. On the other hand, if $0 \preccurlyeq H$, then $\omega_r = c\mathbb{1}$ for any nonzero $c \in \mathbf{R}$; thus since $\omega_\ell'\omega_r = 1$ it is clear that if $\omega_r\omega_\ell' \neq \frac{1}{n}\mathbb{1}\mathbb{1}'$, then the average-consensus cannot be obtained by any linear combination of $s(t)$, $\hat{s}(t)$, or $\underline{s}(t)$. Furthermore, when $0 \preccurlyeq H$ and $\mathcal{W}^o \neq H$, then (1.1), although stable in the limit as μ vanishes, is asymptotically unbounded as the number of $\lfloor 1/\mu \rfloor$ iterations increases. To see this assume $t \leq 0$ and $0 < \mu \ll 1$ are such that $(t/\mu) \in \mathbb{N}$. The (t/μ) iteration of (1.1) yields the sensor estimates

$$s_{t/\mu} = (I - \mu\mathcal{L} - \mu\mathcal{D}^o)^{t/\mu}s(0) + \sum_{l=0}^{t/\mu-1}(I - \mu\mathcal{L} - \mu\mathcal{D}^o)^l\mu\mathcal{W}^oX_{t/\mu-1-l}. \tag{6.14}$$

For arbitrary weight matrices $\{\mathcal{W}^v, \mathcal{W}^o\}$ the limit of (6.14) when μ approaches zero yields the same expression for the coefficient of $s(0)$ as the exact solution (6.11),

$$\lim_{\mu\to 0}(I - \mu\mathcal{L} - \mu\mathcal{D}^o)^{t/\mu} = e^{-(\mathcal{L}+\mathcal{D}^o)t}.$$

Similarly in the limit as μ vanishes, by (6.3) above we replace $X_{t/\mu-1-l}$ with the stationary measure $\pi(\theta(t))$, the summed coefficients of which yield the following limiting values:

$$\lim_{\mu\to 0}\sum_{l=0}^{t/\mu-1}(I - \mu\mathcal{L} - \mu\mathcal{D}^o)^l\mu\mathcal{W}^o = \begin{cases} H^{-1}(I - e^{-Ht})\mathcal{W}^o & ,\ H \prec 0, \\ \alpha(I - e^{-Ht}) & ,\ 0 \preccurlyeq H,\ \mathcal{W}^o = \alpha H,\ \alpha \in \mathbf{R}. \end{cases} \tag{6.15}$$

Without loss of generality set $\alpha = 1$. As t increases, the above limits become those expressed in (6.12). If, however, $0 \preccurlyeq H$ and $\mathcal{W}^o \neq H$, the above iteration results in an unbounded steady-state in the limit as t increases, as can be seen by re-expressing (6.15) in terms of the eigendecompositions $H = UJU^{-1}$ and $\mathcal{W}^o = ARA^{-1}$,

$$\sum_{l=0}^{t/\mu-1}(I - \mu\mathcal{L} - \mu\mathcal{D}^o)^l\mu\mathcal{W}^o = U\sum_{l=0}^{t/\mu-1}(I - \mu J)^l\mu U^{-1}ARA^{-1}. \tag{6.16}$$

In the limit as μ vanishes this expression becomes $UJ^*U^{-1}ARA^{-1}$, where the zero eigenvalue of J is replaced by t in the matrix J^* and all others replaced by $\frac{1-e^{-\lambda_i t}}{\lambda_i}$; for convenience we let the first element of J denote this zero eigenvalue, $J_{11} = 0$. As t increases it is clear that J^* approaches J^{-1} except in its first element, which grows linearly with t. For (6.16) to remain bounded then as t increases, the multiplications $UJ^*U^{-1}ARA^{-1}$ must eliminate the presence of t in J^*_{11}. It is straightforward to see that this occurs if and only if $\mathcal{W}^o = H$, in which case $A = U$, $R = J$, and the zero eigenvalue in R eliminates the presence of t in J^* regardless of the value of μ. In this case, the denominator of J^* is also eliminated by right multiplication of ARA^{-1} and we have the second steady-state of (6.12).

Since we have considered all possible solutions to (2.1) with bounded steady-states, there exist no other methods by which to achieve average-consensus through (1.1) and its implied subalgorithms. Our rationale has been to place conditions on $\{\mathcal{G}^v, \mathcal{G}^o\}$ such that the sensor steady-states under (1.1) can be adjusted by their local tracking estimates $\hat{s}(t)$ to obtain the average-consensus estimate $\bar{\pi}(\theta(t))$. This requires that the equilibrium matrix Λ have each of its rows $i = 1, \ldots, n$ comprised of a nonzero scalar β_i, with diagonal elements defined as $(1 - (n-1)\beta_i)$; see (4.3). The diagonals are defined as such due to the property $\Lambda\mathbb{1} = \mathbb{1}$, which holds only when $0 \prec H$, as can be seen by (6.12) and the fact that

$$(\mathcal{L} + \mathcal{D}^o)^{-1}\mathcal{W}^o\mathbb{1} = \mathbb{1} \;\Rightarrow\; \mathcal{D}^o\mathbb{1} = (\mathcal{L} + \mathcal{D}^o)\mathbb{1}, \tag{6.17}$$

which holds since $\mathcal{L}\mathbb{1} = 0$ by definition. Assuming $0 \prec H$, the necessity of $\Lambda\mathbb{1} = \mathbb{1}$ is evident simply from the contradiction entailed by (6.17) when $\Lambda\mathbb{1} \neq \mathbb{1}$.

The equality among nondiagonal elements within each row i implies that the sensor estimate $s^i(t)$ approaches a linear average with uniform weights (β_i) assigned to the currently observed stationary distributions of all *other* sensors and a disproportionate weight $(1-(n-1)\beta_i)$ given to its own locally observed stationary distribution. If the weights β_i are known, then the correction with $\hat{s}^i_k$ and rescaling, as stated in (4.6), yield $\bar{\pi}(\theta(t))$. ∎

Remark 6.3. The consensus-tracking algorithm (1.1) requires either $0 \preccurlyeq (\mathcal{L}+\mathcal{D}^o)$ or $0 \prec (\mathcal{L}+\mathcal{D}^o)$ in order for the sensor estimates s_k to remain bounded in the limit as μ vanishes and t increases. This is in parallel with the constraints $0 \preccurlyeq \mathcal{L}$ and $0 \prec \mathcal{L}$ that are required for the static Laplacian consensus algorithm (1.3) to remain bounded under the same asymptotic limits. Thus, unlike the works of [10, 21, 31, 5], which consider the static algorithm (1.3), we may permit $\mathcal{L}$ to have negative eigenvalues. Denoting the minimum of these $\underline{\lambda}$, by then taking $\mathcal{D}^o = |\underline{\lambda}|I + V$ for any diagonal matrix $V \geq 0$, we have $(\mathcal{L} + \mathcal{D}^o) = U(J + |\underline{\lambda}|I)U^{-1} + V = UJ^*U^{-1} + V = \tilde{\mathcal{L}} + \tilde{\mathcal{D}}^o$, where now $0 \preccurlyeq \tilde{\mathcal{L}} = UJ^*U^{-1}$ and $\tilde{\mathcal{D}}^o = V$. Under these conditions, we may let $\tilde{\mathcal{W}}^o = \mathcal{W}^o - |\underline{\lambda}|I$ and the results of (6.11), (6.12) hold with respect to $\{\mathcal{L}, \mathcal{W}^o\}$ or equivalently $\{\tilde{\mathcal{L}}, \tilde{\mathcal{W}}^o\}$. In other words, consensus may be achieved even when $\mathcal{L}$ has negative eigenvalues, provided $\mathcal{W}^o$ has sufficiently large positive elements.

7. Conclusions. We have proven that under the Laplacian consensus dynamics and a local linear SA algorithm the sensor state-values within a network communication graph may under suitable averaging weights converge weakly to an average-consensus regarding the state of a two-time-scale Markovian system. In our framework this consensus is based on the slowly time-varying stationary distribution of a Markov chain observed by each sensor. The result has been extended to the case of multiple ergodic classes in the observed Markov chains, or equivalently to network exchange graphs $\{\mathcal{G}^v, \mathcal{G}^o\}$ that belong to a set rather than a fixed parameterization $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{W}\}$. The limiting switched ODE is in this case replaced by a switched differential inclusion, thus motivating the idea of set-valued consensus. In addition to this, the sensor scaled tracking error was proved to satisfy a switching diffusion equation and, in the case of CDF estimation by an empirical measure, a switching Brownian bridge.

Necessary and sufficient conditions were derived in regard to the network communication graph edge set and weights required to ensure average-consensus formation. As a result, computation of the correct consensus estimate requires (1.1) in conjunction with its two component subalgorithms. Thus, obtaining the average-consensus does not require any greater complexity in sensor computing ability or network communication than (1.1) itself assumes. Last, we considered the adaptation rate and magnitude of diffusion present in the sampled sensor trajectories, and proposed an optimization problem as well as an approximate ratio that both relate to various features of the sensor averaging weights. Future work may consider the edge set and averaging weights required for fast convergence to the average-consensus estimate without requiring strong connectivity assumptions or increased sum of the absolute averaging weights.

REFERENCES

[1] P. Alricksson and A. Rantzer, *Distributed Kalman filtering using weighted averaging*, in Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto, Japan, 2006.

[2] P. Antsaklis and L. Fang, *Information consensus of asynchronous discrete-time multi-agent systems*, in Proceedings of the 2005 American Control Conference, Vol. 3, 2005, pp. 1883–1888.

[3] J. Aspnes, *Time- and space-efficient randomized consensus*, J. Algorithms, 14 (1993), pp. 414–431.

[4] R. Beard and D. Kingston, *Discrete-time average-consensus under switching network topologies*, in Proceedings of the 2006 American Control Conference, Minneapolis, MN, 2006.

[5] R. Beard and W. Ren, *Consensus seeking in multi-agent systems under dynamically changing interaction topologies*, IEEE Trans. Automat. Control, 50 (2005), pp. 655–661.

[6] P. Billingsley, *Convergence of Probability Measures*, John Wiley, New York, 1968.

[7] V. Blondel, J. Hendrickx, A. Olshevsky, and J. Tsitsiklis, *Convergence in multiagent coordination, consensus, and flocking*, in Proceedings of the IEEE Conference on Decision and Control, Seville, Spain, 2005, pp. 2996–3000.

[8] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, *Randomized gossip algorithms*, IEEE Trans. Inform. Theory, 52 (2006), pp. 2508–2530.

[9] S. Boyd, S. Lall, and L. Xiao, *A scheme for robust distributed sensor fusion based on average consensus*, in Proceedings of the 4th International IEEE Symposium on Information Processing in Sensor Networks, Los Angeles, CA, 2005.

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

[11] M. Broucke, B. Francis, and Z. Lin, *Local control strategies for groups of mobile autonomous agents*, IEEE Trans. Automat. Control, 49 (2004), pp. 622–629.

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

[13] J. Cortes, *Distributed algorithms for reaching consensus on general functions*, Automatica J. IFAC, 44 (2008), pp. 726–737.

[14] G. Cybenko, *Load balancing for distributed memory multiprocessors*, J. Parallel Distrib. Comput., 7 (1989), pp. 279–301.

[15] M. H. DeGroot, *Reaching a consensus*, J. Amer. Statist. Assoc., 69 (1974), pp. 118–121.

[16] G. B. Ermentrout, *Stable periodic solutions to discrete and continuum arrays of weakly coupled nonlinear oscillators*, SIAM J. Appl. Math., 52 (1992), pp. 1665–1687.

[17] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, *Next century challenges: Scalable coordination in sensor networks*, in Proceedings of the ACM/IEEE International Conference on Mobile Computing and Networking, 1999, pp. 263–270.

[18] M. Franceschelli and A. Gasparri, *Decentralized centroid estimation for multi-agent systems in absence of any common reference frame*, in Proceedings of the 2009 American Control Conference, St. Louis, MO, 2009.

[19] S. Grime and H. Durrant-Whyte, *Data fusion in decentralized sensor networks*, Control Engineering Practice, 2 (1994), pp. 849–863.

[20] V. Gupta, B. Hassibi, and R. M. Murray, *Stability analysis of stochastically varying formations of dynamic agents*, in Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, HI, 2003, pp. 504–509.

[21] Y. Hatano and M. Mesbahi, *Agreement over random networks*, IEEE Trans. Automat. Control, 50 (2005), pp. 1867–1872.

[22] R. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1990.

[23] M. Huang and J. H. Manton, *Stochastic consensus seeking with measurement noise: Convergence and asymptotic normality*, in Proceedings of the 2008 American Control Conference, Seattle, WA, 2008, pp. 1337–1342.

[24] A. Jadbabaie, J. Lin, and A. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Trans. Automat. Control, 48 (2003), pp. 988–1001.

[25] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes*, Academic Press, New York, 1981.

[26] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*, MIT Press, Cambridge, MA, 1984.

[27] H. J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed., Springer-Verlag, New York, 2003.

[28] P. Lin and Y. Jia, *Average consensus in networks of multi-agents with both switching topology and coupling time-delay*, Phys. A, 387 (2008), pp. 303–313.
[29] M. Mehyar, D. Spanos, J. Pongsajapan, S. Low, and R. Murray, *Asynchronous distributed averaging on communication networks*, IEEE Trans. Netw., 18 (2007), pp. 535–547.
[30] L. Moreau, *Stability of multiagent systems with time-dependent communication links*, IEEE Trans. Automat. Control, 50 (2005), pp. 169–182.
[31] R. Murray and R. Olfati-Saber, *Consensus problems in networks of agents with switching topology and time-delays*, IEEE Trans. Automat. Control, 49 (2004), pp. 1520–1532.
[32] R. Murray, R. Olfati-Saber, and D. Spanos, *Distributed sensor fusion using dynamic consensus*, in Proceedings of the 16th IFAC World Congress, Prague, Czech Republic, 2005.
[33] E. Nakamura, A. Loureiro, and A. Frery, *Information fusion for wireless sensor networks: Methods, models, and classifications*, ACM Comput. Surveys, 39 (2007), article 9.
[34] R. Olfati-Saber, *Distributed Kalman filter with embedded consensus filters*, in Proceedings of the Joint 44th IEEE Conference on Decision and Control and European Control Conference, 2005, pp. 8179–8184.
[35] R. Olfati-Saber and J. F. Shamma, *Consensus filters for sensor networks and distributed sensor fusion*, in Proceedings of the 44th IEEE Conference on Decision and Control, Seville, Spain, 2005.
[36] R. Pagliari and A. Scaglione, *Implementation of average consensus protocols for commercial sensor networks platforms*, in Grid Enabled Remote Instrumentation, Springer-Verlag, New York, 2009, pp. 81–95.
[37] K. Plarre and P. Kumar, *Extended message passing algorithm for inference in loopy Gaussian graphical models*, Ad Hoc Networks, 2 (2004), pp. 153–169.
[38] M. Porfiri and D. Stilwell, *Consensus seeking over weighted directed graphs*, IEEE Trans. Automat. Control, 52 (2007), pp. 1767–1773.
[39] R. Rahman, M. Alanyali, and V. Saligrama, *Distributed tracking in multihop sensor networks with communication delays*, IEEE Trans. Signal Process., 55 (2007), pp. 4656–4668.
[40] W. Ren, R. W. Beard, and E. M. Atkins, *A survey of consensus problems in multi-agent coordination*, in Proceedings of the 2005 American Control Conference, Portland, OR, 2005, pp. 1859–1864.
[41] D. Scherber and H. Papadoupoulos, *Locally constructed algorithms for distributed computations in ad-hoc networks*, in Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks, Berkeley, CA, 2004, ACM, New York, pp. 11–19.
[42] D. Scherber and H. C. Papadopoulos, *Distributed computation of averages over ad hoc networks*, IEEE J. Sel. Areas Commun., 23 (2005), pp. 776–787.
[43] 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.
[44] J. Slotine and W. Wang, *A study of synchronization and group cooperation using partial contraction theory*, in Cooperative Control, Springer-Verlag, Berlin, 2003, pp. 207–228.
[45] S. Stankovic, M. Stankovic, and D. Stipanovic, *Decentralized parameter estimation by consensus based stochastic approximation*, in Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, LA, 2007.
[46] A. Tahbaz-Salehi and A. Jadbabaie, *A necessary and sufficient condition for consensus over random networks*, IEEE Trans. Automat. Control, 53 (2008), pp. 791–795.
[47] H. Tanner, A. Jadbabaie, and G. Pappas, *Stable flocking of mobile agents, Part* I: *Fixed topology*, in Proceedings of the 42nd IEEE Conference on Decision and Control, 2003, pp. 2010–2015.
[48] T. Vicsek, A. Czirók, E. Ben-Jacob, O. Cohen, and I. Shochet, *Novel type of phase transition in a system of self-driven particles*, Phys. Rev. Lett., 75 (1995), pp. 1226–1229.
[49] C. W. Wu, *Synchronization and convergence of linear dynamics in random directed networks*, IEEE Trans. Automat. Control, 51 (2006), pp. 1207–1210.
[50] 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.
[51] G. Yin, V. Krishnamurthy, and C. Ion, *Regime switching stochastic approximation algorithms with application to adaptive discrete stochastic optimization*, SIAM J. Optim., 14 (2004), pp. 1187–1215.
[52] G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Singular Perturbations Approach*, Springer-Verlag, New York, 1998.
[53] G. Yin and Q. Zhang, *Discrete-Time Markov Chains: Two-Time-Scale Methods and Applications*, Springer-Verlag, New York, 2005.