Stochastic Lyapunov Analysis for Consensus Algorithms with Noisy Measurements

Minyi Huang and Jonathan H. Manton

***Abstract*— This paper studies the coordination and consensus of networked agents in an uncertain environment. We consider a group of agents on an undirected graph with fixed topology, but differing from most existing work, each agent has only noisy measurements of its neighbors' states. Traditional consensus algorithms in general cannot deal with such a scenario. For consensus seeking, we introduce stochastic approximation type algorithms with a decreasing step size. We present a stochastic Lyaponuv analysis based upon the total mean potential associated with the agents. Subsequently, the so-called direction of invariance is introduced, which combined with the decay property of the stochastic Lyapunov function leads to mean square convergence of the consensus algorithm.**

I. Introduction

Consensus problems are of importance, and in recent years have been a heavily researched area in the context of coordination and control of spatially distributed multi-agent systems, though they have a much longer history. The accumulation of the enormous literature on this topic is, to a large extent, due to its broad connection with a diverse range of disciplines related to statistical decision, management science, medical applications, computer science, biology [25], [10], [4], [8], [24], distributed computing, ad hoc networks, and multi-agent control systems [14], [1], [5], [7], [12], [13], [15], [16], [17], [21]. A comprehensive survey on the recent research on consensus problems can be found in [20].

For a typical formulation within the context of multi-agent coordination, one has a group of agents with individual states, and the associated consensus algorithm is to form an averaging rule [12], [2], [26], based upon the local information of each agent, such that the iterates of all individual states converge to a common value. Various consensus algorithms have been developed to deal with practical scenarios such as asynchronous state update, dynamic topologies or unreliable communication links (see the survey [20]). In the literature, most existing algorithms assume exact state exchange between the agents with only very few exceptions; see, e.g., [19], [27]. A least mean square optimization method was used in [27] to choose the constant coefficients in the averaging rule so that the long term consensus error is minimized. In a continuous time model, deterministic disturbances were treated in [6] in the dynamics of the consensus algorithm. Also, in the early work [3], [22], [23] convergence of consensus problems was studied in a stochastic setting, but the exchange of random messages was assumed to be error-free. In particular, Tsitsiklis, et. al., [23] obtained consensus results in the context of a group of agents minimizing their common cost function.

This work was supported by the Australian Research Council.

M. Huang and J. H. Manton are with the Department of Information Engineering, Research School of Information Sciences and Engineering, The Australian National University, Canberra, ACT 0200, Australia. Email: {minyi.huang, jonathan.manton}@rsise.anu.edu.au.

In practical applications, the information exchange between different agents may involve the usage of sensors, quantization and wireless fading channels, which makes it unlikely to have noise free data delivery. In such models with noisy measurements, the traditional algorithms involving a constant (or non-vanishing) step size in general cannot ensure convergence. Owing to this fact, in the companying paper [11], a stochastic approximation type algorithm was proposed and a stochastic double array analysis was developed for proving convergence. The main assumption there is a certain symmetry property for the underlying directed graph, which facilitates matrix product estimates and leads to convergence both in mean square and along almost all sample paths (i.e., with probability one).

In this paper, we consider a general network topology with noises in inter-agent communication. Specifically, we consider a group of agents in an undirected graph. We develop a stochastic Lyapunov analysis, and convergence is established for connected graphs. Our modelling is different from [23] since in the asynchronous state updating rule of the latter, the exogenous term, which may be interpreted as the local noisy gradient information, is assigned with a small controlled weight while the weights for the exact massages received from other agents are separately selected to be above a fixed level; such a particular structure enables the authors in [23] to obtain consensus with a sufficiently small constant step size, or with only an upper bound condition for the deceasing rate of the step size sequence. In contrast, in our model the data transmitted from other agents are corrupted by noises (see Fig. 1), and consequently, in developing the averaging scheme it is critical to maintain a trade-off in attenuating the noise and ensuring a suitable stabilizing capability to drive the individual states toward each other. To achieve this objective, the step size can be decreased neither too slowly, nor too quickly.

Compared to [11], this paper develops a different approach by exploiting the algebraic properties of the graph Laplacian. The convergence analysis is accomplished by the decay rate estimate of the stochastic Lyapunov function and by the construction of the so-called direction of invariance.

II. The Problem Formulation

We describe the multi-agent system in terms of the standard graph model in the literature. Consider a set of n agents distributed with a spatial structure which is represented by an undirected graph (to be simply called a graph) $G = (\mathscr{N}, \mathscr{E})$ consisting of a set of nodes $\mathscr{N} = \{1, 2, \cdots, n\}$ and a set of edges $\mathscr{E} \subset \mathscr{N} \times \mathscr{N}$. We denote each edge as an unordered pair (i, j) where $i \neq j$, which implies there is no edge between a node and itself. A path in G consists of a sequence of nodes $i_1, i_2, \cdots, i_l$, $l \geq 2$, such that $(i_k, i_{k+1}) \in \mathscr{E}$ for all $1 \leq k \leq l-1$. Two distinct nodes i and j are said to be connected if there exists a path connecting them. The graph G is connected if any two distinct nodes in G are connected. For convenience of exposition, we often refer node i as agent A_i. The two names, agent and node, will be used alternatively. The agent A_k (resp., node k) is a neighbor of A_i (resp., node i) if $(k, i) \in \mathscr{E}$ where $k \neq i$. Denote the neighbors of node i by $\mathscr{N}_i \subset \mathscr{N}$. Throughout this paper, the analysis is for undirected graphs. We make the following assumption:

(**A1**) The graph G is connected. □

In below we follow similar steps as in [11] by introducing the measurement model, the stochastic algorithm and convergence notions. But we note that the exposition below is given in the context of undirected graphs.

A. The Measurement Model

For agent A_i, we denote its state at time t by $x_t^i \in \mathbb{R}$, where $t \in \mathbb{Z}^+ = \{0, 1, 2, \cdots\}$. For each $i \in \mathscr{N}$, agent A_i receives noisy measurements of the states of its neighbors. We denote the resulting measurement by agent A_i of agent A_k's state by

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

where w_t^{ik} is the additive noise; see Fig. 1 for illustration. The underlying probability space is denoted by $(\Omega, \mathscr{F}, P)$. We shall call y_t^{ik} the observation of the state of A_k obtained by A_i, and we assume each A_i knows its own state x_t^i exactly. There may be various interpretations for the additive noise; a natural one is that x_t^i is corrupted by noise during inter-agent communication [19]. We introduce the assumption:

(**A2**) The noises $\{w_t^{ik}, t \in \mathbb{Z}^+, i \in \mathscr{N}, k \in \mathscr{N}_i\}$ are independent with respect to the indices i, k, t and also independent of the initial states x_0^i, $i \in \mathscr{N}$, and each w_t^{ik} has zero mean and variance $Q_t^{i,k} \geq 0$. In addition, $\sup_{i \in \mathscr{N}} E|x_0^i|^2 < \infty$ and $\sup_{t \geq 0, i \in \mathscr{N}} \sup_{k \in \mathscr{N}_i} Q_t^{ik} < \infty$. □

Condition (**A2**) means that the noises are all independent random variables with respect to both space (as indexed by different nodes) and time.

B. The Stochastic Approximation Type Algorithm

The state of each agent is updated by:

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

where $i \in \mathscr{N}$, $t \in \mathbb{Z}^+$ and $a_t \in [0, 1]$. This gives an averaging rule in that the right hand side is a convex combination of the agent's state and its $|\mathscr{N}_i|$ observations, where we use $|S|$ to denote the cardinality of a set S. The objective for the multi-agent consensus problem is to select the sequence $\{a_t, t \geq 0\}$ so that the individual states of the agents will converge to a common limit in a certain sense.

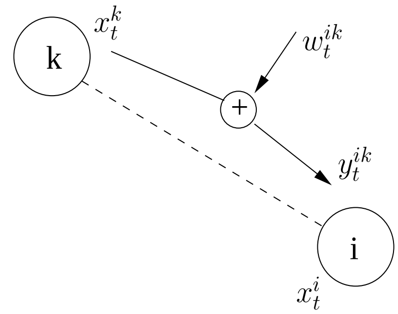

Fig. 1. Measurement with noise w_t^{ik}.

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

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

where

$$m_t^i = \frac{1}{|\mathscr{N}_i|} \sum_{k \in \mathscr{N}_i} y_t^{ik}. \tag{4}$$

Note that the structure of (3) is very similar to the recursion used in classical stochastic approximation algorithms in that $m_t^i - x_t^i$ provides a correction term with the step size a_t. Indeed, after introducing the so-called local potential $P_i(t)$ in Section III, $m_t^i - x_t^i$ may be represented as the noisy measurement of a scaled negative gradient of $P_i(t)$ along the direction x_t^i. Since the additive noise is contained in $\{m_t^i, t \geq 0\}$, each state x_t^i will fluctuate randomly. These fluctuations will not die off if the step size a_t is selected as a constant, and this situation will be illustrated by the simulations in Section V.

With the aim of getting a stable behavior for the agents, a vanishing sequence $\{a_t, t \geq 0\}$ will be used below.

(**A3**) The sequence $\{a_t, t \geq 0\}$ satisfies i) $a_t \in [0, 1]$ and ii) there exists $T_0 \geq 1$ such that

$$\frac{\alpha}{t^\gamma} \leq a_t \leq \frac{\beta}{t^\gamma} \tag{5}$$

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

It is worth discussing the role of T_0 in (5). By starting from a suitable T_0 and requiring $\frac{\alpha}{t^\gamma} \leq a_t$ only for $t \geq T_0$, where $a_t \in [0, 1]$, we may allow large values for α. This gives more flexibility in choosing the step size sequence and otherwise α greater than one would be excluded. For clarity, we emphasize that in further analysis, the parameters $T_0, \alpha, \beta, \gamma$ are treated as fixed constants associated with $\{a_t, t \geq 0\}$.

Note that (**A3**) implies

$$\sum_{t=0}^{\infty} a_t = \infty, \qquad \sum_{t=0}^{\infty} a_t^2 < \infty, \tag{6}$$

which is a typical property for step size sequences used in classical stochastic approximation theory. The vanishing rate of the sequence is important for convergence analysis. We can see that when $a_t \to 0$ in (2), the signal x_t^k (contained in y_t^{ik}), as the state of A_k, is attenuated together with the noise. Hence, a_t cannot decrease too fast since otherwise, the agents may prematurely converge to different individual limits.

C. Consensus Notion in Stochastic Models

In a stochastic setting, the conventional definition of consensus is no longer adequate. We introduce the following definitions on the asymptotic behavior of the agents' states.

Definition 1: (**weak consensus**) The agents are said to reach weak consensus if $E|x_t^i|^2 < \infty$, $t \geq 0$, $i \in \mathcal{N}$, and $\lim_{t\to\infty} E|x_t^i - x_t^j|^2 = 0$ for all distinct $i, j \in \mathcal{N}$. □

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

Definition 3: (**strong consensus**) The agents are said to reach strong consensus if there exists a random variable x^* such that with probability one (w.p.1) $\lim_{t\to\infty} x_t^i = x^*$ for all $i \in \mathcal{N}$. □

It is obvious that mean square consensus implies weak consensus. Note that in the above definitions for mean square and strong consensus, the states x_t^i, $i \in \mathcal{N}$, must converge to a common value. The limit x^*, as a random variable, may depend on the initial states, noise terms and the consensus algorithm itself. Strong consensus has been treated in [11].

D. The Generalization to Vector States

We give some discussions for the vector case where each individual state $\mathbf{x}_t^k \in \mathbb{R}^d$ with dimension $d > 1$. It is easy to extend (1)-(2) to the vector case by taking a vector noise term. For the vector version of these equations, we see that each of the d components in $\mathbf{x}_t^k$ is decoupled from the other $d-1$ components during the iteration. Hence we may decompose the vector equation to d scalar equations. After adapting assumption (**A2**) to the vector case, the consensus result in the paper is easily generalized to the case of vector individual states.

III. Stochastic Lyapunov Functions

In this section, we develop the stochastic Lyapunov analysis. For agent A_i, we define its local potential as

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

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

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

It is easy to show that the term $m_t^i - x_t^i$ in (3) may be decomposed into the form

$$m_t^i - x_t^i = -\frac{1}{|\mathcal{N}_i|}\frac{\partial P_i(t)}{\partial x_t^i} + \frac{1}{|\mathcal{N}_i|}\sum_{j\in\mathcal{N}_i} w_t^{ij}. \tag{7}$$

This indicates the state of each agent is updated along the descending direction of the local potential subject to an additive noise, and justifies a stochastic approximation interpretation of the algorithm (2).

Under assumption (**A1**), it is easy to show that $P_{\mathcal{N}}(t) = 0$ if and only if $x_t^1 = \cdots = x_t^n$. For convergence analysis, we will use $P_{\mathcal{N}}(t)$ as a stochastic Lyapunov function. We introduce the graph Laplacian for G as a matrix $L = (a_{ij})_{1\leq i,j\leq n}$, where

$$a_{ij} = \begin{cases} d_i & \text{if } j = i, \\ -1 & \text{if } j \in \mathcal{N}_i, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

for which $d_i = |\mathcal{N}_i|$ is the degree (i.e., the number of neighbors) of node i. Recall that for a matrix $M \in \mathbb{R}^{n\times n}$, its null space is the solution space of the linear equation $Mx = 0$ for $x \in \mathbb{R}^n$. We denote $1_n = [1, 1, \cdots, 1]^T \in \mathbb{R}^n$. The rank of L is $n-1$ for the connected graph G and the null space of L is $\{c1_n, c \in \mathbb{R}\}$ [9], [18].

A. Recursion of Stochastic Lyapunov Functions

Denote by x_t the state vector for the n agents, i.e., $x_t = [x_t^1, \cdots, x_t^n]^T$. We have the relation [9]:

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

By (2), we have the state updating rule:

$$x_{t+1}^i = (1-a_t)x_t^i + (a_t/|\mathcal{N}_i|)\sum_{j\in\mathcal{N}_i} x_t^j + (a_t/|\mathcal{N}_i|)\sum_{j\in\mathcal{N}_i} w_t^{ij}. \tag{9}$$

Denote

$$\tilde{w}_t^i = (1/|\mathcal{N}_i|)\sum_{j\in\mathcal{N}_i} w_t^{ij}, \quad \tilde{w}_t = [\tilde{w}_t^1, \cdots, \tilde{w}_t^n]^T. \tag{10}$$

We further introduce the matrix $\hat{L} = (\hat{a}_{ij})_{1\leq i,j\leq n}$ where

$$\hat{a}_{ij} = \begin{cases} 1 & \text{if } j = i, \\ -d_i^{-1} & \text{if } j \in \mathcal{N}_i, \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

with $d_i = |\mathcal{N}_i|$, and we define the diagonal matrix $D_{\mathcal{N}} = \mathrm{Diag}(d_1^{-1}, \cdots, d_n^{-1})$. It is easy to verify that

$$\hat{L} = D_{\mathcal{N}} L.$$

Lemma 4: For $t \geq 0$, we have

$$\begin{aligned} P_{\mathcal{N}}(t+1) &= P_{\mathcal{N}}(t) - 2a_t x_t^T L D_{\mathcal{N}} L x_t + a_t^2 x_t^T L D_{\mathcal{N}} L D_{\mathcal{N}} L x_t \\ &\quad + 2a_t x_t^T L\tilde{w}_t - 2a_t^2 x_t^T L D_{\mathcal{N}} L\tilde{w}_t + a_t^2\tilde{w}_t^T L\tilde{w}_t, \end{aligned} \tag{12}$$

where the sequence $\{x_t, t \geq 0\}$ is generated by (1) and (2).

Proof: By (9), we get the vector equation

$$x_{t+1} = x_t - a_t\hat{L}x_t + a_t\tilde{w}_t, \qquad t \geq 0. \tag{13}$$

Equation (13) leads to the recursion of the total potential:

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

and the lemma follows. □

We denote the null spaces of the nonnegative definite matrices L, LDL, and $LD_{\mathcal{N}}LD_{\mathcal{N}}L$ by N_1, N_2 and N_3, respectively.

Theorem 5: Under (**A1**), we have the assertions:

(i) The null spaces of L, $LD_{\mathscr{N}}L$ and $LD_{\mathscr{N}}LD_{\mathscr{N}}L$ are given by the same one dimensional space, i.e., $N_i = \mathrm{span}\{1_n\}$, where $i = 1, 2, 3,$.

(ii) There exists positive constants $c_1 > 0$ and $c_2 > 0$ such that $LD_{\mathscr{N}}L \geq c_1 L$ and $LD_{\mathscr{N}}LD_{\mathscr{N}}L \leq c_2 L$.

(iii) In addition, we assume (**A2**)-(**A3**) and let T_c be such that $1 - 2a_t c_1 + a_t^2 c_2 \geq 0$ for all $t \geq T_c$. For the total mean potential, we have

$$V(t+1) \leq (1 - 2a_t c_1 + a_t^2 c_2) V(t) + O(a_t^2)$$

where $t \geq T_c$, and the algorithm (2) achieves weak consensus.

Proof: See Appendix. □

IV. The Direction of Invariance

Theorem 5 shows the difference between the states of any two agents converges to zero in mean square, as $t \to \infty$. However, this alone, does not guarantee that they will converge to a common limit. The asymptotic vanishing of the stochastic Lyapunov function only indicates that the state vector x_t will approach the subspace $\mathrm{span}\{1_n\}$. To obtain consensus results, we need some additional estimation. The strategy is to show that the oscillation of the sequence $\{x_t, t \geq 0\}$ along the direction 1_n will gradually die off. This is achieved by proving the existence of a vector η which is not orthogonal to 1_n and such that the linear combination $\eta^T x_t$ of the components in x_t converges. For convenience, η will be chosen to satisfy the additional requirement that $\eta^T x_{t+1}$ depends not on the whole of x_t but only on $\eta^T x_t$; this means $\eta^T x_t$ is a one-dimensional auto-regressive process, and its study is easier than that of the original process x_t.

Definition 6: Let $x_t = [x_t^1, \cdots, x_t^n]^T$ be generated by the algorithm (2). If $\eta = (\eta_1, \cdots, \eta_n)^T$ is a real-valued vector of unit length, i.e., $|\eta|^2 = \sum_{i=1}^n \eta_i^2 = 1$ and satisfies

$$\eta^T x_{t+1} = \eta^T x_t + a_t \eta^T \tilde{w}_t, \qquad t \geq 0, \tag{14}$$

for any initial condition x_0^i, $i \in \mathscr{N}$ and any step size sequence $a_t \in [0, 1]$, where $\tilde{w}_t$ is given in (10), then η is called a direction of invariance associated with (2). □

The directions of invariance associated with the consensus algorithm (2) are easily characterized in terms of the degrees of the nodes of the underlying graph.

Theorem 7: We have the assertions:

(i) There exists a real-valued vector $\eta = (\eta_1, \cdots, \eta_n)^T$ of unit length satisfying $\eta^T \hat{L} = 0$ where $\hat{L}$ is defined by (11).

(ii) If η is a unit length vector, then η is a direction of invariance associated with (2) if and only if $\eta^T \hat{L} = 0$.

(iii) Under (**A1**), the direction of invariance has the representation $\eta = c[d_1, \cdots, d_n]^T$ where $c = \pm(\sum_{i=1}^n d_i^2)^{1/2}$ and the integer $d_i = |\mathscr{N}_i|$ is the degree of node $i \in \mathscr{N}$.

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

We now show (ii). The condition $\eta^T \hat{L} = 0$ combined with (13) implies

$$\begin{aligned} \eta^T x_{t+1} &= \eta^T x_t - a_t \eta^T \hat{L} x_t + a_t \eta^T \tilde{w}_t \\ &= \eta^T x_t + a_t \eta^T \tilde{w}_t. \end{aligned}$$

The sufficiency part of (ii) follows easily. Conversely, if the unit length vector η satisfies (14) for all initial states x_0^i and the step size a_t as specified in Definition 6, then we necessarily have $\eta^T \hat{L} = 0$. So the necessity part of (ii) holds.

We continue to prove (iii) under (**A1**). By (ii) and the definition of $\hat{L}$, η with $|\eta| = 1$ is a direction of invariance if and only if $\eta^T D_{\mathscr{N}} L = 0$, which in turn, is equivalent to

$$LD_{\mathscr{N}} \eta = 0.$$

By (**A1**) and Theorem 5, we have $D_{\mathscr{N}} \eta = c 1_n$ where $c \neq 0$ is a constant to be determined. This gives the row vector

$$\eta = c[d_1, \cdots, d_n]^T, \qquad c \neq 0, \tag{15}$$

where c is determined by the unit length condition of η. The direction of invariance is unique up to sign. □

If η is a direction of invariance, then Theorem 7 shows under (**A1**) that all elements of η have the same sign. Therefore, η is not orthogonal to 1_n, and the requirement stated at the beginning of this section is met. Geometrically, the notion of the direction of invariance means under (2) and zero noise conditions, the projection (i.e., $(\eta^T x_t)\eta$) of $x_t \in \mathbb{R}^n$ along the direction η would remain a constant vector regardless of the value of $a_t \in [0, 1]$ used in the iterates.

A. Mean Square Consensus

Now we are in a position to establish mean square consensus. We state the following lemma.

Lemma 8: Assume (**A1**)-(**A3**) hold, and $\{x_t, t \geq 0\}$ is given by (13). Let $\eta_0 = [d_1, \cdots, d_n]^T$ where $d_i = |\mathscr{N}_i|$. Then there exists a random variable y^* such that $\lim_{t\to\infty} E|\eta_0 x_t - y^*|^2 = 0$.

Proof: By Theorem 7, $\eta_0/|\eta_0|$ is a direction of invariance. Hence, we have

$$\eta_0^T x_{t+1} = \eta_0^T x_0 + a_0 \eta_0^T \tilde{w}_0 + \cdots + a_t \eta_0^T \tilde{w}_t. \tag{16}$$

By (**A2**) and (**A3**), it follows that $\eta_0^T x_t$ converges in mean square, and the lemma follows. □

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

Theorem 9: Assume (**A1**)-(**A3**) hold. The algorithm (2) achieves mean square consensus.

Proof: By Theorem 5, we have weak consensus, i.e.,

$$\lim_{t\to\infty} E|x_t^i - x_t^k|^2 = 0, \qquad \forall i, k \in \mathscr{N}. \tag{17}$$

On the other hand, by Lemma 8, as $t \to \infty$,

$$\begin{aligned} \eta_0^T x_t &= [d_1, \cdots, d_n]^T x_t \\ &= \eta_0^T [x_t^1 - x_t^1, \cdots, x_t^n - x_t^1]^T + \eta_0^T [x_t^1, \cdots, x_t^1]^T \end{aligned} \tag{18}$$

converges in mean square, which further combined with (17) implies x_t^1 converges in mean square. By (17) again, we see that the mean square consensus result follows. □

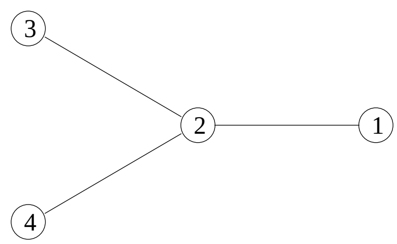

Fig. 2. The undirected graph with 4 nodes.

B. A Three Node Example

For illustration, we give a three node model with $\mathscr{N} = \{1,2,3\}$, where $\mathscr{N}_1 = \{2\}$, $\mathscr{N}_2 = \{1,3\}$ and $\mathscr{N}_3 = \{2\}$. For this model, we have

$$P_1(t) = \frac{1}{2}|x_t^1 - x_t^2|^2, \quad P_2(t) = \frac{1}{2}(|x_t^2 - x_t^1|^2 + |x_t^2 - x_t^3|^2),$$
$$P_3(t) = \frac{1}{2}|x_t^3 - x_t^2|^2.$$

For illustrating the direction of invariance, we take $\zeta_t = (1/\sqrt{6})(x_t^1 + 2x_t^2 + x_t^3)$ for $t \geq 0$, and we can verify that

$$\zeta_{t+1} = \zeta_t + a_t \eta_0^T \tilde{w}_t$$

where $\tilde{w}_t$ is a sequence of independent vector noises and $\eta_0 = (1/\sqrt{6})[1,2,1]^T$ is a direction of invariance. We see that η_0 is consistent with the expression (15) since the degrees for the three nodes are, respectively, 1, 2 and 1.

V. Numerical Simulations

In the numerical studies, we consider an undirected graph with 4 nodes $\mathscr{N} = \{1,2,3,4\}$ and edges $\mathscr{E} = \{(1,2),(2,3),(2,4)\}$; see Fig. 2. The initial condition for the state vector $x_t = [x_t^1, \cdots, x_t^4]^T$ at $t = 0$ is $[5,1,3,2]^T$, and the variance of the i.i.d. Gaussian measurement noises is $\sigma^2 = 0.01$. The simulation of the standard averaging rule with equal weights to an agent's neighbors and itself is given in Fig. 3; hence we have $x_{t+1}^1 = (x_t^1 + y_t^{12})/2$ and $x_{t+1}^2 = (x_t^2 + y_t^{21} + y_t^{23} + y_t^{24})/4$, etc., where $t \geq 0$. It is seen that the 4 state trajectories in Fig. 3 move toward each other rather quickly at the beginning, but they maintain long term fluctuations as the state iteration continues. The stochastic algorithm (2) is used in Fig. 4 with the step size sequence $\{a_t = (t+5)^{-0.85}, t \geq 0\}$. Fig. 4 shows the 4 trajectories all merge into a constant level, and this is consistent with the mean square consensus result obtained in this paper.

VI. Conclusions

We have developed a stochastic Lyapunov analysis for consensus problems with noisy measurements. The convergence result is obtained by use of the decay property of the stochastic Lyapunov function and the direction of invariance. For future work, it is of interest to consider stochastic algorithms with network conditions such as dynamic topologies and asynchronous state updates.

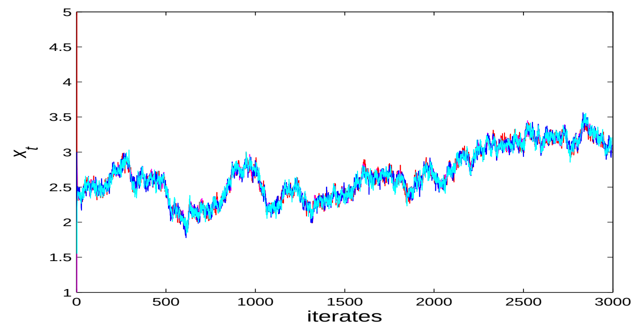

Fig. 3. The 4 node example using the fixed step size.

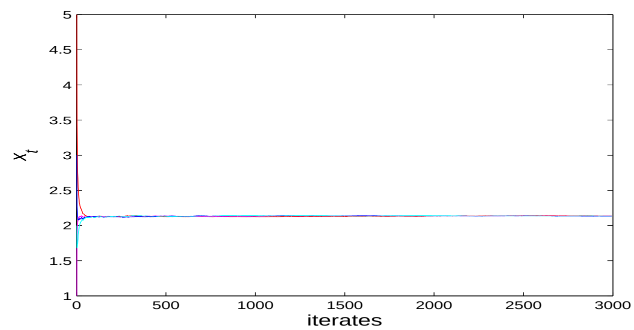

Fig. 4. The 4 node example using the decreasing step size.

Appendix: Proof of Theorem 5

Proof: (i) In the following we use $A \Rightarrow B$ as the abbreviation for "A implies B", and $A \Leftrightarrow B$ for "A is equivalent to B". First, it is a well known fact [9], [18] that when the graph is connected, $N_1 = \text{span}\{1_n\}$.

Since L is nonnegative definite, there exists a nonnegative definite matrix, denoted as $L^{1/2}$ such that $L = (L^{1/2})^2$. We also write $D_{\mathscr{N}}^{1/2} = \text{Diag}(d_1^{-1/2}, \cdots, d_n^{-1/2})$ which gives $D_{\mathscr{N}} = (D_{\mathscr{N}}^{1/2})^2$. For $x \in \mathbb{R}^n$, we have $Lx = 0 \Rightarrow LD_{\mathscr{N}}Lx = 0 \Rightarrow LD_{\mathscr{N}}LD_{\mathscr{N}}Lx = 0$. On the other hand, we have

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

Hence, it immediately follows that

$$Lx = 0 \Leftrightarrow LD_{\mathscr{N}}Lx = 0 \Leftrightarrow LD_{\mathscr{N}}LD_{\mathscr{N}}Lx = 0,$$

and assertion (i) follows. It is evident that the rank for each of the matrices L, $LD_{\mathscr{N}}L$ and $LD_{\mathscr{N}}LD_{\mathscr{N}}L$ is equal to $n-1$.

(ii) We begin by proving the first part. Let

$$0 = \lambda_1, \quad 0 < \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n, \tag{19}$$

and

$$0 = \hat{\lambda}_1, \quad 0 < \hat{\lambda}_2 \leq \hat{\lambda}_3 \leq \cdots \leq \hat{\lambda}_n,$$

respectively, denote the eigenvalues of L and $LD_{\mathscr{N}}L$. Let $\Phi = (\alpha_1, \cdots, \alpha_n)$ and $\hat{\Phi} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_n)$ be two orthogonal matrices (i.e., $\Phi^T\Phi = I$, and $\hat{\Phi}^T\hat{\Phi} = I$) such that

$$L\Phi = \Phi\mathrm{Diag}(\lambda_1, \cdots, \lambda_n), \quad LD_{\mathscr{N}}L\hat{\Phi} = \hat{\Phi}\mathrm{Diag}(\hat{\lambda}_1, \cdots, \hat{\lambda}_n).$$

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

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

$$\sum_{i=2}^n y_i^2 = \sum_{i=2}^n \hat{y}_i^2. \tag{20}$$

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

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

On the other hand, we have

$$x^T Lx \leq \lambda_n\sum_{i=2}^n y_i^2 = \lambda_n\sum_{i=2}^n \hat{y}_i^2,$$

where the equality follows from (20). Hence

$$x^T LD_{\mathscr{N}}Lx \geq \hat{\lambda}_2\lambda_n^{-1}x^T Lx,$$

and the first part of (ii) is proved by taking $c_1 = \hat{\lambda}_2\lambda_n^{-1} > 0$.

We denote the eigenvalues of $LD_{\mathscr{N}}LD_{\mathscr{N}}L$ by

$$0 = \tilde{\lambda}_1, \quad 0 < \tilde{\lambda}_2 \leq \tilde{\lambda}_3 \leq \cdots \leq \tilde{\lambda}_n.$$

By a similar argument, we can show that for any $x \in \mathbb{R}^n$,

$$x^T LD_{\mathscr{N}}LD_{\mathscr{N}}Lx \leq \tilde{\lambda}_n\lambda_2^{-1}x^T Lx$$

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

(iii) The inequality follows by taking expectation on both sides of (12) and using (ii). Consequently, we select $\hat{T}_c \geq T_c$ to ensure $1 - 2c_1a_t + c_2a_t^2 \leq 1 - c_1a_t$ for all $t \geq \hat{T}_c$, and find a fixed constant $C > 0$ such that

$$V(t+1) \leq (1 - c_1a_t)V(t) + Ca_t^2, \tag{21}$$

for all $t \geq \hat{T}_c$. By lengthy but elementary product estimates under (**A3**), we get $\lim_{t\to\infty} V(t) = 0$. Then it follows that

$$\lim_{t\to\infty} E|x_t^i - x_t^k|^2 = 0, \qquad i \in \mathscr{N},\ k \in \mathscr{N}_i. \tag{22}$$

By connectivity of the graph, for any pair of nodes i and k, we can find a path from i to k. Then by repeatedly applying (22) to all pairs of neighboring nodes along that path, we can show that

$$\lim_{t\to\infty} E|x_t^i - x_t^k|^2 = 0, \qquad \forall i, k \in \mathscr{N}, \tag{23}$$

which implies weak consensus. $\square$

REFERENCES

[1] R. W. Beard and V. Stepanyan. Synchronization of information in distributed multiple vehicle coordination control. *Proc. IEEE CDC Conf.*, Maui, Hawaii, pp. 2029-2034, Dec. 2003.

[2] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. *Proc. IEEE CDC-ECC'05*, Seville, Spain, pp. 2996-3001, Dec. 2005.

[3] V. Borkar and P. Varaiya. Asymptotic agreement in distributed estimation. *IEEE Trans. Automat. Control*, vol. 27, pp. 650-655, June 1982.

[4] S. Chatterjee and E. Seneta. Towards consensus: some convergence theorems on repeated averaging. *J. Appl. Probab.*, vol. 14, pp. 89-97, 1977.

[5] J. Cortes, S. Martinez, and F. Bullo. Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions. *IEEE Trans. Automat. Control*, vol. 51, no. 8, pp. 1289-1298, Aug. 2006.

[6] G. A. de Castro and F. Paganini. Convex synthesis of controllers for consensus. *Proc. ACC*, pp. 4933-4938, Boston, MA, 2004.

[7] J. P. Desai, V. Kumar, and J. P. Ostrowski. Control of changes in formation for a team of mobile robots. *Proc. IEEE Int. Conf. on Robotics and Automat.*, Detroit, Michigan, pp. 1556-1561, May 1999.

[8] M. J. Fischer. The consensus problem in unreliable distributed systems (a brief survey). In *Foundations of Computation Theory*, M. Karpinsky Ed., Lec. Notes Comput. Sci., vol. 158, pp. 127-140, Springer-Verlag, 1983.

[9] C. Godsil and G. Royle. *Algebraic Graph Theory*, Springer-Verlag, New York, 2001.

[10] M. H. de Groot. Reaching a consensus. *J. Amer. Statist. Assoc.*, vol. 69, pp. 118-121, 1974.

[11] M. Huang and J. H. Manton. Stochastic double array analysis and convergence of consensus algorithms with noisy measurements. *Proc. American Control Conference*, New York, July 2007.

[12] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Trans. Automat. Contr.*, vol. 48, no. 6, pp. 988-1000, June 2003.

[13] Y. Liu and Y. R. Yang. Reputation propagation and aggreement in wireless ad hoc networks. *Proc. IEEE Wireless Commun. Network. Conf.*, 2003.

[14] N. Lynch. *Distributed Algorithms*. Morgan Kaufmann, San Matero, CA, 1996.

[15] J. A. Marshall and M. E. Broucke. On invariance of cyclic group symmetries in multiagent formations. *Proc. IEEE CDC-ECC'05*, Seville, Spain, pp. 746-751, Dec. 2005.

[16] R. Olfati-Saber. Flocking for multi-agent dynamic systems: algorithms and theory. *IEEE Trans. Automat. Contr.*, vol. 51, pp. 401-420, Mar. 2006.

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

[18] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *IEEE Proc.*, vol. 95, no. 1, pp. 215-233, 2007.

[19] W. Ren, R. W. Beard, and D. B. Kingston. Multi-agent Kalman consensus with relative uncertainty. *Proc. American Control Conf.*, Portland, OR, pp. 1865-1870, June 2005.

[20] W. Ren, R. W. Beard, and E. M. Atkins. A survey of consensus problems in multi-agent coordination. *Proc. American Control Conference*, Portland, OR, pp. 1859-1864, June 2005.

[21] H. G. Tanner, G. J. Pappas, and V. Kumar. Input-to-state stability on formation graphs. *Proc. IEEE CDC Conf.*, Las Vegas, NV, pp. 2439-2444, Dec. 2002.

[22] J. N. Tsitsiklis and M. Athans. Convergence and asymptotic agreement in distributed decision problems. *IEEE Trans. Automat. Control*, vol. 29, no. 1, pp. 42-50, Jan. 1984.

[23] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans. Automat. Contr.*, vol. 31, no. 9, pp. 803-812, 1986.

[24] P. K. Visscher. How self-organization evolves. *Nature*, vol. 421, pp. 799-800, Feb. 2003.

[25] R. L. Winkler. The consensus of subjective probability distributions. *Management Science*, vol. 15, B61-B75, 1968.

[26] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging, *Systems and Control Letters*, vol. 53, pp. 65-78, 2004.

[27] L. Xiao, S. Boyd, and S.-J. Kim. Distributed average consensus with least-mean-square deviation. *Journal of Parallel and Distributed Computing*, vol. 67, pp. 33-46, 2007.