Proceedings of the
46th IEEE Conference on Decision and Control
New Orleans, LA, USA, Dec. 12-14, 2007

Stochastic Approximation for Consensus Seeking: Mean Square and Almost Sure Convergence

Minyi Huang and Jonathan H. Manton

Abstract— **We consider stochastic consensus problems in strongly connected directed graph models where each agent has noisy measurements of its neighbors' states. For consensus seeking, we develop stochastic approximation type algorithms with a decreasing step size and establish mean square and almost sure convergence of the agents' states to the same limit.**

I. INTRODUCTION

Consensus problems are of importance, and in recent years have been an intensively researched area in the context of coordination and control of distributed multi-agent systems, though they have a much longer history. The steady accumulation of an enormous literature on this topic is, to a large extent, due to its connection with a diverse range of disciplines related to statistical decision theory, management science, computer science, biology [30], [10], [5], [9], [29], distributed computing, wireless ad hoc and sensor networks, and multi-agent control systems [16], [1], [7], [8], [14], [15], [4], [17], [19], [20], [25]. A comprehensive survey on the recent research on consensus problems can be found in [23].

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 [14], [2], [31], based upon the local information of each agent, such that the iterates of all individual states converge to a common value. The basic formulation may be generalized to deal with asynchronous state update, dynamic topologies or unreliable communication links (see the survey [23]). In the literature, most existing algorithms assume exact state exchange between the agents with only very few exceptions (see, e.g., [22], [32]). A least mean square optimization method was used in [32] to choose the constant coefficients in the averaging rule so that the long term consensus error is minimized. Also, in the early work [3], [27], [28] convergence of consensus problems was studied in a stochastic setting, but the exchange of random messages between the agents was assumed to be error-free. In particular, [28] obtained consensus results for a group of agents minimizing their common cost function via stochastic gradient based optimization.

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. In the work [12], [13], [11], a stochastic approximation type algorithm was proposed for consensus seeking where the data transmitted from other agents are corrupted by noises (see Fig. 1). 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. In particular, almost sure convergence results are obtained in directed graph models satisfying a circulant invariance property [12], and mean square convergence is established for connected undirected graphs by a stochastic Lyapunov analysis [13].

In this paper, we generalize the analysis in [12], [13] to strongly connected directed graphs. First, we analyze mean square convergence by a stochastic Lyapunov analysis. In this case, the useful properties of a graph Laplacian are no longer available, and we need to construct suitable Lyapunov functions. This, in turn, leads to the in-depth analysis of a class of degenerate algebraic Lyapunov equations. Next, we generalize the double array analysis in [12], and prove almost sure convergence of the algorithm.

II. THE PROBLEM FORMULATION

Consider n agents distributed according to a directed graph (or digraph) $G=(\mathscr{N},\mathscr{E})$ consisting of a set of nodes $\mathscr{N}=\{1,2,\cdots,n\}$ and a set of edges $\mathscr{E}\subset\mathscr{N}\times\mathscr{N}$. In the digraph, an edge from node i to node j is denoted as an ordered pair (i,j) where $i\neq j$ (so there is no edge between a node and itself). A path (from i_1 to i_l) 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 $k=1\cdots,l-1$. We say node i is connected to node $j(\neq i)$ if there exists a path from i to j. The graph G is said to be strongly connected if each node i is connected to any other node j by a path.

For convenience of exposition, 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 $\mathscr{N}_i=\{k|(k,i)\in\mathscr{E}\}$.

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

M. Huang is with the School of Mathematics and Statistics, Carleton University, Ottawa, ON K1S 5B6, Canada. Email: mhuang@math.carleton.ca.

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

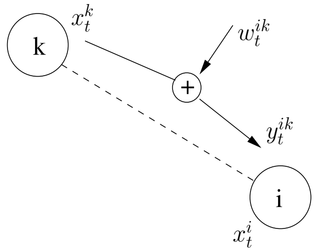

Fig. 1. Measurement with noise w_t^{ik}.

the resulting measurement by agent A_i of agent A_k's state by

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

where $w_t^{ik} \in \mathbb{R}$ is the additive noise; see Fig. 1 for illustration. The underlying probability space is denoted by $(\Omega, \mathscr{F}, P)$. We 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 [22]. We introduce the assumptions:

(**A1**) The graph $G = (\mathcal{N}, \mathscr{E})$ is strongly connected. □

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

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

B. The Stochastic Approximation Algorithm

The state of each agent is updated by the rule

$$x_{t+1}^i = (1 - a_t b_{ii}) x_t^i + a_t \sum_{k\in\mathcal{N}_i} b_{ik} y_t^{ik}, \quad i \in \mathcal{N}, \ t \geq 0, \tag{2}$$

where the step size $a_t \geq 0$, $b_{ik} > 0$ for $k \in \mathcal{N}_i$, and $b_{ii} = \sum_{k\in\mathcal{N}_i} b_{ik}$. We call b_{ik}, $k \in \mathcal{N}_i$, the relative weight that A_i assigns to its neighbor A_k. We restrict that $a_t b^* \in [0,1]$, where

$$b^* \triangleq \max_{i\in\mathcal{N}} b_{ii}.$$

Thus the right hand side of (2) is a convex combination of the agent's state and its $|\mathcal{N}_i|$ observations. Here we use $|S|$ to denote the cardinality of a set S. The objective of the consensus problem is to select $\{a_t, t \geq 0\}$ so that the individual states converge to a common limit in a certain sense.

For each i, we further define

$$b_{ik} = 0, \quad \text{for } k \notin \mathcal{N}_i \cup \{i\}. \tag{3}$$

Define the matrix

$$B = \begin{pmatrix} -b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & -b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & -b_{nn} \end{pmatrix}. \tag{4}$$

Let $\tilde{w}_t^i = \sum_{k\in\mathcal{N}_i} b_{ik} w_t^{ik}$ and define

$$x_t = (x_t^1, \cdots, x_t^n)^T, \quad \tilde{w}_t = (\tilde{w}_t^1, \cdots, \tilde{w}_t^n)^T. \tag{5}$$

Then we write algorithm (2) in the vector form

$$x_{t+1} = x_t + a_t B x_t + a_t \tilde{w}_t. \tag{6}$$

We may also rewrite (2) in the form

$$x_{t+1}^i = x_t^i + a_t (m_t^i - b_{ii} x_t^i) \tag{7}$$

where $m_t^i = \sum_{k\in\mathcal{N}_i} b_{ik} y_t^{ik}$ and $m_t^i - b_{ii} x_t^i$ provides a correction term controlled by the step size a_t. Since the additive noise is contained in $\{m_t^i, t \geq 0\}$, each state x_t^i will have long term fluctuations if the step size a_t is selected as a constant. 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) $a_t \in [0, (b^*)^{-1}]$ and ii) there exists $T_0 \geq 1$ such that

$$\frac{\alpha}{t^\gamma} \leq a_t \leq \frac{\beta}{t^\gamma} \tag{8}$$

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

Note that $b^* > 0$ under (**A1**). 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

$$\sum_{t=0}^{\infty} a_t = \infty, \qquad \sum_{t=0}^{\infty} a_t^2 < \infty, \tag{9}$$

which is a typical property for step size sequences used in classical stochastic approximation theory. 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 Notions in Stochastic Models

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 $\lim_{t\to\infty} x_t^i = x^*$ for all $i \in \mathcal{N}$. □

Convergence with probability one is also called almost sure (a.s.) convergence. In the above mean square and strong consensus, the states x_t^i, $i \in \mathcal{N}$, must converge to a common limit. However, the limit x^* as a random variable may depend upon the initial states, noises and the consensus algorithm.

In this paper, we only consider scalar individual states and the analysis may be easily generalized to the case of vector individual states; see related discussions in [12].

III. Mean Square Convergence

We prove the mean square convergence of algorithm (6) by a stochastic Lyapunov function approach.

Lemma 4: Under (**A1**), all eigenvalues of B define by (4) is inside the circle with radius $b^*>0$ on the complex plane:

$$\{s: |s+b^*| \le b^*\} \tag{10}$$

and $s=0$ is an eigenvalue with multiplicity one.

Proof: Denote the eigenvalues of the stochastic matrix $I+B/b^*$ by λ_i, $1\le i\le n$, where $\lambda_1=1$ and $|\lambda_i|\le 1$ for $i\ge 2$. Since G is strongly connected, $I+B/b^*$ is irreducible. This leads to analyzing the two scenarios below.

Case 1. If $I+B/b^*$ is aperiodic, then $|\lambda_i|<1$ for all $i\ge 2$.

Case 2. If $I+B/b^*$ is periodic with period $d\ge 2$, then there are a total of d eigenvalues, denoted by $\lambda_1,\cdots,\lambda_d$ with absolute value equal to 1, and $\lambda_k = e^{2\pi(k-1)\mathbf{i}/d}$ where $1\le k\le d$ and $\mathbf{i}$ is the imaginary unit [24]. And $|\lambda_k|<1$, for $d+1\le k\le n$.

By combining Cases 1 and 2 about the distribution of the eigenvalues of $I+B/b^*$, the lemma follows. □

Let $S^{n\times n}$ denote the set of $n\times n$ real symmetric matrices, and denote $1_n=[1,\cdots,1]^T$. Define the set of matrices:

$$\mathscr{D} \triangleq \{D\in S^{n\times n}:\ D\ge 0,\ \mathrm{Null}(D)=\mathrm{span}\{1_n\}\}.$$

Obviously, each $D\in\mathscr{D}$ has rank $n-1$.

Theorem 5: Assuming (**A1**), for B defined by (4) and any given $D\in\mathscr{D}$, there exists a unique $Q\in\mathscr{D}$ to satisfy

$$QB+B^TQ=-D. \tag{11}$$

Proof: See Appendix. □

Compared with the usual application of Lyapunov equations in stability analysis of linear systems, we have a more adverse situation since B is not strictly stable. Consequently, for the right hand side of (11) we only use $D\in\mathscr{D}$, instead of a positive definite matrix, and accordingly, the solution Q is not required to be positive definition. But it turns out such a "weaker" requirement for the pair (Q,D) is sufficient for our convergence analysis. Due to the degenerate nature of Q and D, we shall call (11) a degenerate algebraic Lyapunov equation.

We use the solution matrix $Q\in\mathscr{D}$ of (11) to construct the stochastic Lyapunov function

$$P_{\mathscr{N}}(t)=x_t^TQx_t, \qquad t\ge 0,$$

where x_t is generated by (6). Denote $V(t)=EP_{\mathscr{N}}(t)$. We have the following decay property of the Lyapunov function.

Theorem 6: Under (**A1**)-(**A3**), we have (i)

$$\begin{aligned} V(t+1) =& V(t)+a_tEx_t^T(QB+B^TQ)x_t \\ &+a_t^2Ex_t^TB^TQBx_t+O(a_t^2), \end{aligned} \tag{12}$$

(ii) there exist constants $c_1>0$ and $c_2>0$, determined by the matrices B,Q and D, such that

$$V(t+1)\le(1-c_1a_t+c_2a_t^2)V(t)+O(a_t^2), \tag{13}$$

for all $t\ge T_c$, where T_c is selected such that $1-c_1a_t+c_2a_t^2\ge 0$ for all $t\ge T_c$, and (iii) $\lim_{t\to\infty}V(t)=0$.

Proof: The theorem may be proved by following the argument in proving Theorem 5 in [13]. □

Theorem 7: Under (**A1**)-(**A3**), algorithm (6) achieves mean square consensus.

Proof: First, by Theorem 6 we have

$$\lim_{t\to\infty} Ex_t^TQx_t=0, \tag{14}$$

where $Q\in\mathscr{D}$. Next, we define the function

$$F(x_t)=\sum_{k=1}^{n-1}(x_t^{k+1}-x_t^k)^2+(x_t^1-x_t^n)^2, \qquad t\ge 0,$$

and may write $F(x_t)=x_t^TQ_Fx_t$, where $Q_F\in\mathscr{D}$. Since Q and Q_F both have the null space span$\{1_n\}$ and are positive definite when restricted to the orthogonal complementary subspace of span$\{1_n\}$, by following the method in proving Theorem 5 in [13], we can show that there exists a constant $c_3>0$ such that $Q_F\le c_3Q$, and therefore $F(x_t)\le c_3x_t^TQx_t$ which combined with (14) implies $\lim_{t\to\infty}EF(x_t)=0$; hence weak consensus follows.

We continue to prove mean square consensus. For $a\in(0,(b^*)^{-1})$, we write the equation

$$\pi^T(I+aB)=\pi^T \tag{15}$$

where $\pi=(\pi_1,\cdots,\pi_n)^T$. For the given value a, $I+aB$ is the transition matrix of an irreducible and aperiodic Markov chain with no transient states, hence there exists a unique invariant probability measure π satisfying (15) and having n positive entries. By (15), we have the recursion

$$\pi^Tx_{t+1}=\pi^Tx_t+a_t\pi^T\tilde{w}_t, \qquad t\ge 0.$$

By (**A2**)-(**A3**), π^Tx_t converges in mean square to a limit x^*. Recalling the weak consensus result, we have

$$\lim_{t\to\infty}E|x_t^i-x^*|^2=\lim_{t\to\infty}E\Big|\sum_{k=1}^n\pi_k(x_t^i-x_t^k)+\pi^Tx_t-x^*\Big|^2=0,$$

for each $i\in\mathscr{N}$. □

Remark: Theorems 6 and 7 hold when (**A3**)-ii) is replaced by (9). □

IV. Almost Sure Convergence

For each $t\in\mathbb{Z}^+$, the set of noises $\{w_t^{ik}, i\in\mathscr{N}$ and $k\in\mathscr{N}_i\neq\emptyset\}$ is listed into a vector $\mathbf{w}_t$ in which the position of w_t^{ik} depends only on (i,k) but not on t.

(**A2'**) The initial state vector satisfies $P\{|x_0|<\infty\}=1$. The sequence $\{\mathbf{w}_t, t\in\mathbb{Z}^+\}$ constitutes i.i.d. vector random variables with zero mean and $E|\mathbf{w}_t|^\tau<\infty$ for some $\tau\in(1,2]$. □

Theorem 8 below is based on Theorem 3 in [26] and is useful for studying sample path behavior of algorithm (6).

Theorem 8: [26] Let $\{w,w_t,t\ge 1\}$ be i.i.d. real-valued random variables with zero mean, and $\{a_{ki}, 1\le i\le l_k\uparrow\infty, k\ge 1\}$ a double array of constants. Assume (i) $\max_{1\le i\le l_k}|a_{ki}|h_i=O(1/\log k)$, where $0<h_i\uparrow$, $h_i=O(i^{1/\delta})$ for some $\delta\in[1,2]$, (ii) $\sum_{i=1}^\infty P\{|w|>h_i\}<\infty$, and (iii) $h_i/i\downarrow$ and $\sum_{i=1}^{l_k}|a_{ki}|^2h_i^{2-\delta}=o(1/\log k)$, $\sum_{i=1}^{l_k}|a_{ki}|^2h_i^{2-\delta}=O(1/\log l_k)$. Then we have $\lim_{k\to\infty}\sum_{i=1}^{l_k}a_{ki}w_i=0$ a.s.. □

Corollary 9: If $\{w, w_t, t \geq 1\}$ are i.i.d. $\mathbb{R}^n$-valued random variables with zero mean and $\{A_{ki}, 1 \leq i \leq l_k \uparrow \infty, k \geq 1\}$ an $\mathbb{R}^{n\times n}$-valued double array, then $\lim_{k\to\infty} \sum_{i=1}^{l_k} A_{ki} w_i = 0$ a.s., if conditions (i) and (iii) of Theorem 8 hold after replacing $|a_{ik}|$ by the matrix norm $\|A_{ik}\|$ and if condition (ii) of Theorem 8 is satisfied by the vector random variable w. □

We proceed to prove the sample path convergence of algorithm (6), which is rewritten below:

$$x_{t+1} = x_t + a_t B x_t + a_t \tilde{w}_t.$$

By (28) in Appendix, we have a nonsingular real matrix $\Phi = (1_n, \phi_{n\times(n-1)})$ such that $\Phi^{-1} B \Phi = \begin{pmatrix} 0 & \\ & \tilde{B}_{n-1} \end{pmatrix} \triangleq \tilde{B}$, where the $n-1$ eigenvalues of $\tilde{B}_{n-1}$ have strictly negative real parts. Letting $z_t = \Phi^{-1} x_t$ and $\tilde{v}_t = \Phi^{-1}\tilde{w}_t$, we have

$$z_{t+1} = z_t + a_t \tilde{B} z_t + a_t \tilde{v}_t, \qquad t \geq 0.$$

Let $z_t = [z_t^1, \cdots, z_t^n]^T$, $\tilde{v}_t = [\tilde{v}_t^1, \cdots, \tilde{v}_t^n]^T$, and $z^{(n-1)} = [z_t^2, \cdots, z_t^n]^T$, $\tilde{v}_t^{(n-1)} = [\tilde{v}_t^2, \cdots, \tilde{v}_t^n]^T$. We have the relation:

$$z_{t+1}^1 = z_t^1 + a_t \tilde{v}_t^1, \tag{16}$$

$$z_{t+1}^{(n-1)} = (I + a_t \tilde{B}_{n-1}) z_t^{(n-1)} + a_t \tilde{v}_t^{(n-1)}. \tag{17}$$

Lemma 10: Assuming (**A1**) and (**A3**), there exist constants $\hat{\delta} \in (0, (\sup_{t\geq 0}\{a_t\})^{-1}]$ and $C > 0$ such that

$$\|\prod_{i=k}^{l} (I + a_i \tilde{B}_{n-1})\| \leq C \prod_{i=k}^{l} (1 - \hat{\delta} a_i), \quad \forall\ l \geq k \geq 1. \tag{18}$$

Proof: We solve the algebraic Lyapunov equation $\tilde{B}_{n-1}^T \tilde{Q} + \tilde{Q}\tilde{B}_{n-1} = -I$ to get a unique $\tilde{Q} > 0$. Let the constant T_1 be selected such that $a_t \tilde{B}_{n-1}^T \tilde{Q} \tilde{B}_{n-1} \leq (1/2) I$ for all $t \geq T_1$. It suffices to prove (18) for all $l \geq k \geq T_1$. For $t \geq T_1$,

$$\begin{aligned}(I + a_t \tilde{B}_{n-1})^T \tilde{Q} (I + a_t \tilde{B}_{n-1}) &= \tilde{Q} - a_t I + a_t^2 \tilde{B}_{n-1}^T \tilde{Q} \tilde{B}_{n-1} \\ &\leq \tilde{Q} - (a_t/2) I \\ &\leq \tilde{Q} - (a_t/(2\lambda_{\max}))\tilde{Q} \\ &\triangleq (1 - \delta a_t)\tilde{Q} \end{aligned} \tag{19}$$

where $\lambda_{\max} > 0$ is the largest eigenvalue of $\tilde{Q}$. Hence

$$\begin{aligned}&(I + a_l \tilde{B}_{n-1})^T \cdots (I + a_k \tilde{B}_{n-1})^T \tilde{Q} (I + a_k \tilde{B}_{n-1}) \cdots (I + a_l \tilde{B}_{n-1}) \\ &\leq (1 - \delta a_l) \cdots (1 - \delta a_k) \tilde{Q}, \qquad l \geq k \geq T_1. \end{aligned} \tag{20}$$

We may take any $0 < \hat{\delta} < (\delta/2) \wedge (\sup_{t\geq 0}\{a_t\})^{-1}$ and the lemma follows. □

For any $\delta^* \in (0, (\sup_{t\geq 0}\{a_t\})^{-1}]$, we define

$$\Pi_{l,k} = a_k \prod_{i=k+1}^{l} (1 - \delta^* a_i), \tag{21}$$

where $l \geq k \geq 1$. We have the lemmas.

Lemma 11: For $\{a_t, t \geq 0\}$ satisfying (**A3**), we have the upper bound estimate: (i) If $\gamma = 1$ and $\varepsilon \in [0, 1)$

$$\sum_{k=1}^{t} \Pi_{t,k}^2 k^{\varepsilon} = \begin{cases} O(t^{-2\alpha\delta^*}) & \text{if} \quad 0 < \alpha < (1-\varepsilon)/(2\delta^*) \\ O(t^{\varepsilon-1}\ln t) & \text{if} \quad \alpha = (1-\varepsilon)/(2\delta^*) \\ O(t^{\varepsilon-1}) & \text{if} \quad \alpha > (1-\varepsilon)/(2\delta^*). \end{cases}$$

(ii) If $1/2 < \gamma < 1$ and $\varepsilon \in [0, \gamma)$, then $\sum_{k=1}^{t} \Pi_{t,k}^2 k^{\varepsilon} = O(t^{\varepsilon-\gamma})$.

Proof: We obtain the estimates by the same approach as in proving Lemma 5 of [12]. □

Lemma 12: Given $\varepsilon \in [0, \gamma - 1/2)$, $\max_{1\leq k\leq t} \Pi_{t,k} k^{\varepsilon+1/2} = O((\ln t)^{-1})$.

Proof: Case (i) $\gamma = 1$. Similar to Lemma 4 in [12], we can show $\Pi_{t,k} \leq [\beta(k+1)^{\alpha\delta^*}]/[k(t+1)^{\alpha\delta^*}]$. Hence

$$\max_{1\leq k\leq t} \Pi_{t,k} k^{\varepsilon+1/2} = \begin{cases} O(t^{-\alpha\delta^*}) & \text{if} \quad \alpha\delta^* + \varepsilon \leq 1/2 \\ O(t^{\varepsilon-1/2}) & \text{if} \quad \alpha\delta^* + \varepsilon > 1/2, \end{cases}$$

which implies that the lemma holds for $\gamma = 1$.

Case (ii) $1/2 < \gamma < 1$. Again, similar to Lemma 4 in [12], we have

$$\Pi_{t,k} \leq \exp\{-\frac{\alpha\delta^*}{1-\gamma}[(t+1)^{1-\gamma} - (k+1)^{1-\gamma}]\}\frac{\beta}{k^{\gamma}}. \tag{22}$$

Then in parallel to Lemma 8 of [12], we obtain $\max_{1\leq k\leq t} \Pi_{t,k} k^{\varepsilon+1/2} = O(t^{\varepsilon+1/2-\gamma})$ for $1/2 < \gamma < 1$. □

Theorem 13: Assume (**A1**), (**A2'**) and (**A3**) hold with $\gamma\tau > 1$. Then z_t converges a.s. to a random variable $z_\infty = (z_\infty^1, 0)^T$ as $t \to \infty$.

Proof: Since $z_{t+1}^1 = \sum_{i=0}^{t} a_i \tilde{v}_i^1$ and $\sum_{t=0}^{\infty} a_t^{\tau} E|\tilde{v}_t^1|^{\tau} < \infty$ there exists z_∞^1 such that $\lim_{t\to\infty} z_t^1 = z_\infty^1$ a.s. (see [6], pp. 114)).

For the sequence $\{z_t^{(n-1)}, t \geq 1\}$, we have the relation

$$\begin{aligned} z_{k+1}^{(n-1)} &= \Big[\prod_{t=1}^{k} (I + a_t \tilde{B}_{n-1})\Big] z_1^{(n-1)} \\ &\quad + \sum_{i=1}^{k} \Big[\prod_{t=i+1}^{k} (I + a_t \tilde{B}_{n-1})\Big] a_i \tilde{v}_i^{(n-1)} \\ &= \Big[\prod_{t=1}^{k} (I + a_t \tilde{B}_{n-1})\Big] z_1^{(n-1)} + \sum_{i=1}^{k} \Pi_{k,i}^M \tilde{v}_i^{(n-1)}, \end{aligned} \tag{23}$$

where the matrix $\Pi_{k,i}^M$ is defined in an obvious manner. By Lemma 10, we see that

$$\lim_{k\to\infty} \Big[\prod_{t=1}^{k} (I + a_t \tilde{B}_{n-1})\Big] z_1^{(n-1)} = 0, \qquad \text{a.s.} \tag{24}$$

By Lemma 10 again, we obtain $|\Pi_{k,i}^M| \leq C \Pi_{k,i}$, for all $k \geq i \geq 1$, where C is a fixed constant. Lemmas 11 and 12 give

$$\max_{1\leq i\leq k} |\Pi_{k,i}^M| i^{1/\tau} = O((\ln k)^{-1}), \quad \sum_{i=1}^{k} |\Pi_{k,i}^M|^2 i^{2/\tau - 1} = o((\ln k)^{-1}), \tag{25}$$

where we can verify that the exponents $1/\tau$ and $2/\tau - 1$ satisfy the conditions in Lemmas 11 and 12.

On the other hand, note that $\{\tilde{v}_i^{(n-1)}, i \geq 1\}$ is a sequence of i.i.d. vector valued random variables. We have

$$\begin{aligned} \sum_{k=1}^{\infty} P\{|\tilde{v}_1^{(n-1)}| > k^{1/\tau}\} &= \sum_{k=1}^{\infty} kP\{k^{1/\tau} < |\tilde{v}_1^{(n-1)}| \leq (k+1)^{1/\tau}\} \\ &= \sum_{k=1}^{\infty} kP\{k < |\tilde{v}_1^{(n-1)}|^{\tau} \leq k+1\} \leq E|\tilde{v}_1^{(n-1)}|^{\tau} < \infty. \end{aligned} \tag{26}$$

We apply Corollary 9 by taking $l_k = k$, $\delta = \tau$, $h_i = i^{1/\tau}$, $A_{ki} = \Pi_{k,i}^M$, $w_t = \tilde{v}_t^{(n-1)}$, $t \geq 1$, and all its conditions have been verified by (25) and (26). Hence $\lim_{t\to\infty} \sum_{k=1}^{t} \Pi_{t,k}^M \tilde{v}_k^{(n-1)} = 0$

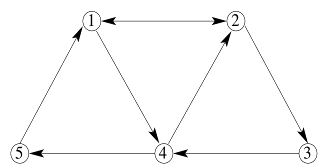

Fig. 2. The digraph with 5 nodes.

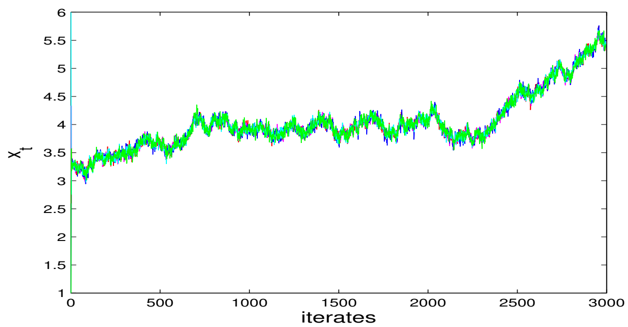

Fig. 3. The 5 trajectories fail to converge when fixed weights are used.

a.s., which combined with (24) implies $\lim_{t\to\infty} z_t^{(n-1)} = 0$, a.s. Hence $\lim_{t\to\infty} z_t = (z_\infty^1, 0)^T$ a.s. □

Theorem 14: Under (**A1**), (**A2'**) and (**A3**), algorithm (6) ensures strong consensus.

Proof: By $x_t = \Phi z_t$ and Theorem 13, the limit $x_\infty = \lim_{t\to\infty} x_t$ exists a.s., and $x_\infty = \Phi z_\infty = (1_n, \phi_{n\times(n-1)})(z_\infty^1, 0)^T = z_\infty^1 1_n$ a.s., which implies strong consensus. □

V. Numerical Simulations

We consider a digraph with 5 nodes as shown in Fig. 2. The variance of the i.i.d. Gaussian measurement noises is $\sigma^2 = 0.01$. The initial state vector is $x_t|_{t=0} = [4,3,1,6,1]^T$. Fig. 3 shows the simulation of the standard averaging rule with equal weights for an agent's neighbors and itself (for instance, $x_{t+1}^1 = (x_t^1 + y_t^{12} + y_t^{15})/3$, $t \geq 0$), and no convergence is achieved. Fig. 4 shows mean square and strong consensus as achieved by algorithm (6) with $b_{ij} = |\mathcal{N}_i|^{-1}$, $j \in \mathcal{N}_i$, and the step size sequence $\{a_t = (t+5)^{-0.85}, t \geq 0\}$, where the 5 trajectories all merge toward a constant.

VI. Conclusions

We have analyzed stochastic consensus algorithms with measurement noise in strongly connected digraph models. Two different approaches, i.e., Lyapunov analysis and double array analysis, are developed, leading to mean square and almost sure convergence results, respectively. For future work, it is of interest to generalize the convergence analysis to dynamic network topologies.

VII. Appendix

Proof: We split the proof into 2 steps.

Step 1. We introduce the integral representation formula

$$Q = \int_0^\infty e^{B^T t} D e^{Bt} dt. \tag{27}$$

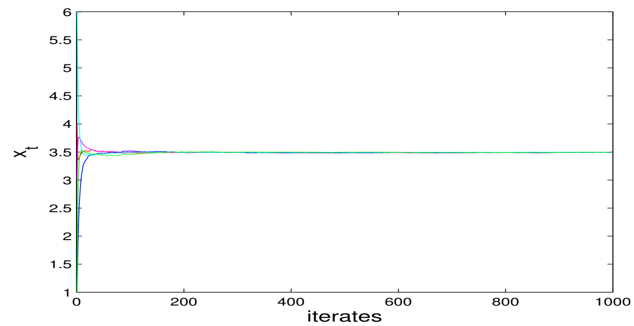

Fig. 4. The 5 trajectories converge to the same constant level with a decreasing step size.

We need to show that the right hand side is well defined and that it gives a solution to (11).

Since B has the eigenvalue 0 and another $n-1$ eigenvalues with strictly negative real parts, in below we show there exists a real matrix $\Phi \triangleq (1_n, \phi_{n\times n-1})$, where $\phi_{n\times(n-1)}$ is an $n\times(n-1)$ matrix, such that we have the block-wise diagonalization

$$\Phi^{-1} B \Phi = \begin{pmatrix} 0 & 0 \\ 0 & \tilde{B}_{n-1} \end{pmatrix}, \tag{28}$$

where $\tilde{B}_{n-1} \in \mathbb{R}^{(n-1)\times(n-1)}$ is a strictly stable matrix. Note that 1_n is the eigenvector of B associated with the eigenvalue 0. Since $\mathrm{rank}(B) = n-1$, there exist $n-1$ linearly independent vectors ζ_i, $1 \leq i \leq n-1$, such that $\mathbb{S} \triangleq \mathrm{span}\{\zeta_1, \cdots, \zeta_{n-1}\} = \mathrm{span}\{B\}$ where $\mathrm{span}\{B\}$ denotes the linear space spanned by the columns of B. Obviously, $1_n \notin \mathbb{S}$. We take $\phi_{n\times(n-1)} = (\zeta_1, \cdots, \zeta_{n-1})$ and compose a nonsingular matrix $(1_n, \phi_{n\times(n-1)})$. Since $\mathbb{S}$ is an invariant subspace of the linear transform associated with B, there exists an $(n-1)\times(n-1)$ matrix $\tilde{B}_{n-1}$ such that

$$B\phi_{n\times(n-1)} = \phi_{n\times(n-1)}\tilde{B}_{n-1},$$

and (28) follows.

Let $c > 0$ be a constant such that the real part of each eigenvalue $\bar{\lambda}_k$ of $\tilde{B}_{n-1}$ is strictly less than $-c$, i.e.,

$$\mathrm{Re}(\bar{\lambda}_k) < -c, \quad k = 1, \cdots, n-1. \tag{29}$$

For $D \in \mathscr{D}$, we use $D^{1/2}$ to denote the nonnegative definite matrix such that $D = (D^{1/2})^2$. It is easy to check that $\mathrm{Null}(D^{1/2}) = \mathrm{span}\{1_n\}$. Now it follows that

$$\begin{aligned} D^{1/2} e^{Bt} &= D^{1/2}(1_n, \phi_{n\times(n-1)}) \begin{pmatrix} 1 & 0 \\ 0 & e^{\tilde{B}_{n-1}t} \end{pmatrix} \Phi^{-1} \\ &= \left(0, D^{1/2}\phi_{n\times(n-1)} e^{\tilde{B}_{n-1}t}\right)\Phi^{-1}. \end{aligned}$$

Subsequently, for c determined in (29), we have

$$\|e^{B^T t} D e^{Bt}\| = O(e^{-2ct}), \tag{30}$$

which implies the integral in (27) converges.

We continue to show that $Q \in \mathscr{D}$. Since $D \geq 0$, we have $Q \geq 0$. By the power series expansion of e^{Bt}, we can show $e^{Bt}1_n = 1_n$ since $1_n \in \text{Null}(B)$. Then (27) leads to

$$Q1_n = \int_0^\infty e^{B^T t} D 1_n dt = 0.$$

On the other hand, if there exists a nonzero real vector ξ such that $Q\xi = 0$, then we have

$$\xi^T Q \xi = \int_0^\infty \xi^T e^{B^T t} D e^{Bt} \xi dt = 0.$$

By $\xi^T e^{B^T t} D e^{Bt} \xi \geq 0$ for all $t \geq 0$ and its continuity in t, we necessarily have $\xi^T e^{B^T t} D e^{Bt} \xi|_{t=0} = \xi^T D \xi = 0$, which implies $\xi \in \text{span}\{1_n\}$ since $D \in \mathscr{D}$. So we conclude $Q \in \mathscr{D}$.

Next, we verify Q defined in (27) is the desired solution. For each $C \in (0, \infty)$, we have

$$\begin{aligned} &(\int_0^C e^{B^T t} D e^{Bt} dt) B + B^T (\int_0^C e^{B^T t} D e^{Bt} dt) \\ = &\int_0^C \frac{d}{dt}(e^{B^T t} D e^{Bt}) dt = e^{B^T C} D e^{BC} - D. \end{aligned} \tag{31}$$

By letting $C \to \infty$ in (31), it follows from (30) that

$$QB + B^T Q = -D \tag{32}$$

where Q is defined by (27). This proves the existence of a solution to the algebraic Lyapunov equation (11).

Step 2. Now we prove uniqueness. Suppose there exists $\bar{Q} \in \mathscr{D}$ such that

$$\bar{Q}B + B^T \bar{Q} = -D. \tag{33}$$

Let $\Delta = \bar{Q} - Q$. By (32) and (33), we get $\Delta B = -B^T \Delta$, which leads to $\Delta (Bt)^k = (-1)^k (B^T t)^k \Delta$, for $k = 0, 1, 2, \cdots$, and therefore

$$\Delta e^{Bt} = e^{-B^T t} \Delta. \tag{34}$$

By (34), we get

$$\Delta = e^{B^T t} \Delta e^{Bt} = e^{B^T t} \bar{Q} e^{Bt} - e^{B^T t} Q e^{Bt}. \tag{35}$$

Similar to (30), we get the estimate

$$\|e^{B^T t} \bar{Q} e^{Bt}\| + \|e^{B^T t} Q e^{Bt}\| = O(e^{-2ct}), \tag{36}$$

as $t \to \infty$, by the fact that both Q and $\bar{Q}$ are in $\mathscr{D}$. Hence (35) and (36) imply that $\Delta = \bar{Q} - Q = 0$, and uniqueness follows. □

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. 44th 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. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Trans. Inf. Theory*, vol. 52, pp. 2508-2530, June 2006.

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

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

[7] 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.

[8] 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.

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

[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. Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior. Submitted for journal publication, Oct. 2006.

[12] M. Huang and J. H. Manton. Stochastic double array analysis and convergence of consensus algorithms with noisy measurements. *Proc. American Control Conference*, New York, pp. 705-710, July 2007.

[13] M. Huang and J. H. Manton. Stochastic Lyapunov analysis for consensus algorithms with noisy measurements. *Proc. American Control Conference*, New York, pp. 1419-1424, July 2007.

[14] 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.

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

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

[17] 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.

[18] B. Mohar. The Laplacian spectrum of graphs. In *Graph Theory, Combinatorics, and Applications*, vol. 2, pp. 871-898, Y. Alavi, G. Chartrand, O. R. Oellermann, and A. J. Schwenk Eds., Wiley, 1991.

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

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

[21] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *IEEE Proc.*, vol. 95, pp. 215-233, Jan. 2007.

[22] 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.

[23] 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.

[24] E. Seneta. *Non-negative Matrices and Markov Chains*. 2nd ed., Springer-Verlag, New York, 1981.

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

[26] H. Teicher. Almost certain convergence in double arrays. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 69, pp. 331-345, 1985.

[27] 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.

[28] 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.

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

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

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

[32] 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.