2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Stochastic Consensus Seeking with Measurement Noise: Convergence and Asymptotic Normality

Minyi Huang and Jonathan H. Manton

Abstract— **We consider consensus seeking with measurement noise in directed graphs containing a spanning tree. By using stochastic approximation type algorithms, we show the state of each agent converges in mean square and almost surely to the same limit. Furthermore, we show that the approximation error, as the difference between the state vector and its limit, is asymptotically normal after normalization, which in turn characterizes the convergence rate of the algorithm. Finally, we generalize the algorithm to networks with random link failures.**

I. INTRODUCTION

A fundamental problem in the study of spatially distributed multi-agent systems is the so-called consensus problem, and a consensus algorithm or protocol provides a mechanism to propagate shared information across the population of agents. Due to its crucial role in such distributed systems, consensus problems and various closely related formulations have been intensively investigated in multi-agent control systems and distributed computing [16], [21], [19]. A comprehensive survey on recent research can be found in [22], [26].

In a basic setting with networked agents, a consensus algorithm is to form a weighted averaging rule [16], [3], where each agent uses data obtained from its neighbors and itself, such that the iterates of all individual states converge to a common value. Most existing algorithms assume exact state averaging, which in general necessitates perfect state exchange among the agents. Recently, there is an increasing attention on models with noise or quantization effect [25], [31], [5]. The early work [28] developed stochastic gradient based consensus algorithms for distributed function optimization. The communication or sensing noise issue also arises in the setting of distributed function computation in sensor networks [8] and formation control [1].

In consensus models with noisy measurements, the traditional algorithms involving constant or lowered bounded averaging weights in general cannot ensure convergence. On the other hand, however, in recursive computation of stochastic systems for the purpose of either function optimization, root-finding, or system identification, it has long been known that a properly decreasing gain for the correction term is essential for obtaining convergence, and there has existed a vast literature on these broad areas [20], [2], [6], [4]. In [10], [11], [12], a stochastic approximation type algorithm with a decreasing step size was proposed for consensus seeking where the state information of other agents is corrupted by white noise (see Fig. 1). In particular, almost sure (a.s., i.e., probability one) convergence results were obtained in [10] via a double array analysis in directed graph (also called digraph) models satisfying a circulant invariance property. Mean square convergence was proved for connected undirected graphs by using a stochastic Lyapunov function [11]. The analysis in [10], [11] was generalized to strongly connected digraphs in [13].

In this paper, we extend the analysis in [10], [11], and we adopt the spanning tree based model [24] to give a unified treatment of two scenarios: (i) leadless consensus seeking; (ii) leader following. We prove the consensus algorithm converges in mean square and almost surely. Subsequently, we show that the approximation error is asymptotically normally distributed after a suitable normalization; this, in turn, characterizes the convergence rate of the algorithm. Finally, the consensus formulation is generalized to models with random independent link failures. For related random graph based modeling, see e.g. [9], [30], [17], [23]. Within the random graph model, our algorithm may be viewed as an equivalent one with a fixed network topology subject to unbiased perturbations. The paper is organized as follows. Section II formulates the stochastic consensus algorithm and some algebraic preliminaries are summarized in Section III. After giving some preliminary lemmas in Section IV, the main consensus results are stated in Section V. Asymptotic normality results are established in Section VI. Section VII deals with random link failures, and Section VIII presents simulations. Section IX concludes the paper.

II. THE PROBLEM FORMULATION

Consider n agents distributed according to a digraph $G = (\mathscr{N}, \mathscr{E})$ consisting of a set of nodes $\mathscr{N} = \{1, 2, \cdots, n\}$ and a set of directed edges $\mathscr{E} \subset \mathscr{N} \times \mathscr{N}$. For brevity, a directed edge will be simply called an edge. 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 directed path (from i_1 to i_l, to be simply called a path) 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 $1 \leq k \leq l-1$. We say node i is connected to node $j (\neq i)$ if there exists a path from i to j. The digraph G is said to be strongly connected if each node is connected to any other node by a path. A directed tree is a digraph where each node, except the root node, has exactly one parent node. Hence, the root node is connected to any other node by a path. The

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

J. H. Manton is with Department of Information Engineering, Research School of Information Science and Engineering, The Australian National University, Canberra, ACT 0200, Australia. Email: j.manton@ieee.org.

978-1-4244-2079-7/08/$25.00 ©2008 AACC.

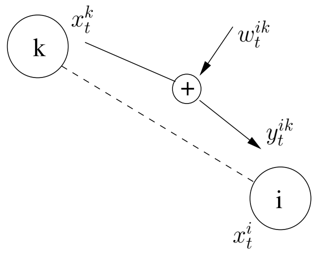

Fig. 1. Measurement with noise w_t^{ik}.

digraph G is said to contain a spanning tree $G_s = (\mathcal{N}_s, \mathcal{E}_s)$ if G_s is a directed tree such that $\mathcal{N}_s = \mathcal{N}$ and $\mathcal{E}_s \subset \mathcal{E}$. A strongly connected digraph always contains a spanning tree.

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 $(k,i) \in \mathcal{E}$ where $k \neq i$. Denote the neighbor set of node i by $\mathcal{N}_i = \{k | (k,i) \in \mathcal{E}\}$.

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 $i \in \mathcal{N}$, A_i receives noisy measurements of the states of its neighbors if $\mathcal{N}_i \neq \emptyset$, where $\emptyset$ denotes the empty set. We denote the resulting measurement by A_i of A_k's state by

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

where $w_t^{ik} \in \mathbb{R}$ is the additive noise; see Fig. 1 for illustration. The underlying probability space is denoted by $(\Omega, \mathcal{F}, P)$. We call y_t^{ik} the observation of the state of A_k obtained by A_i, and 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 [25]. We introduce the assumption:

(**A1**) The digraph $G = (\mathcal{N}, \mathcal{E})$ contains a spanning tree. □

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

$$x_t = (x_t^1, \cdots, x_t^n)^T, \qquad t \geq 0. \tag{2}$$

(**A2**) The sequence $\{\mathbf{w}_t, t \in \mathbb{Z}^+\}$ constitutes a sequence of independent random vectors which is independent of x_0 and satisfies $\sup_{t\geq 0} E|\mathbf{w}_t|^2 < \infty$. In addition $E|x_0|^2 < \infty$. □

B. The Stochastic Approximation Algorithm

The state of each agent is updated by the rule

$$x_{t+1}^i = (1 - a_t b_{ii}) x_t^i + a_t \sum_{k \in \mathcal{N}_i} b_{ik} y_t^{ik}, \qquad t \geq 0, \tag{3}$$

where $i \in \mathcal{N}$, $a_t > 0$ and the parameters b_{ij} will be specified subsequently. 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 [10]. Throughout our analysis, we adopt the convention: $\sum_{k\in\emptyset} = 0$ regardless of the summand. For specifying b_{ij} in (3), we consider two cases in terms of $\mathcal{N}_i$.

Case 1. If $\mathcal{N}_i \neq \emptyset$, we take:

$$\begin{cases} b_{ik} > 0, & \text{if } k \in \mathcal{N}_i, \\ b_{ik} = 0, & \text{if } k \notin \mathcal{N}_i \cup \{i\}, \\ b_{ii} = \sum_{k\in\mathcal{N}_i} b_{ik}, \end{cases}$$

and we call b_{ik}, $k \in \mathcal{N}_i$, the relative weight that A_i assigns to its neighbor A_k.

Case 2. If $\mathcal{N}_i = \emptyset$, we define $b_{ik} \equiv 0$ for all $k \in \mathcal{N}$ and the state of agent i is fixed:

$$x_t^i \equiv x_0^i. \tag{4}$$

For instance, (4) naturally arises in leader following. By our earlier convention $\sum_{k\in\emptyset} = 0$, (4) may be interpreted as a special case of (3).

Define the matrix

$$B = \begin{pmatrix} -b_{11} & \cdots & b_{1n} \\ \vdots & \vdots & \vdots \\ b_{n1} & \cdots & -b_{nn} \end{pmatrix}. \tag{5}$$

Let $w_t^i = \sum_{k\in\mathcal{N}_i} b_{ik} w_t^{ik}$ where $i \in \mathcal{N}$, and define $w_t = (w_t^1, \cdots, w_t^n)^T$. In the case $\mathcal{N}_i = \emptyset$, we accordingly have $w_t^i = \sum_{k\in\emptyset} = 0$. Write (3) in the vector form

$$x_{t+1} = x_t + a_t B x_t + a_t w_t, \qquad t \geq 0. \tag{6}$$

(**A3**) The sequence $\{a_t, t \geq 0\}$ satisfies i) $a_t \in (0, (\max_{i\in\mathcal{N}} b_{ii})^{-1}]$ and ii) $\sum_{t=0}^{\infty} a_t = \infty$, $\sum_{t=0}^{\infty} a_t^2 < \infty$. □

Under (**A1**), $\max_{i\in\mathcal{N}} b_{ii} > 0$ since at least one node has a nonempty neighbor set. In (**A3**)-i), we restrict $a_t \leq (\max_{i\in\mathcal{N}} b_{ii})^{-1}$ so that the coefficients for x_t^i and y_t^{ik} in (3) are all nonnegative. This gives a weighted averaging as in typical consensus algorithms. However, our convergence analysis for (6) holds without imposing the upper bound in (**A3**)-i).

Definition 1: (**mean square consensus**) The agents are said to reach mean square consensus if $E|x_t|^2 < \infty$, $t \geq 0$, 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 2: (**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 $i \in \mathcal{N}$. □

III. Algebraic Preliminaries

Let B be given by (5), and define the matrix

$$M = (m_{ij})_{1\leq i,j\leq n} = I + \gamma B, \tag{7}$$

where $b_{ij} > 0$ if and only if $(j,i) \in \mathcal{E}$ by the definition of B, and $0 < \gamma < (\max_{i\in\mathcal{N}} b_{ii})^{-1}$. Recall that $\max_{i\in\mathcal{N}} b_{ii} > 0$ when (**A1**) holds. It is obvious that M is a stochastic matrix and can be naturally associated with a discrete time Markov chain with state space $S = \{1, 2, \cdots, n\}$.

Recall that a nonnegative matrix is irreducible if all its associated states communicate with each other. Under (**A1**), the key observation is that the set of states $S = \{1, \cdots, n\}$ associated with the stochastic matrix M contains a communicating subclass S_c such that each state in S can reach S_c by finite transitions. By use of the method in [27], the following result may be proved.

Lemma 3: [14], [15] If G contains a spanning tree and M is defined by (7), then we have:

(i) there exists a permutation $(l_1, \cdots, l_n)$ of $(1, 2, \cdots, n)$ such that the blockwise lower triangular stochastic matrix

$$M^o = \begin{bmatrix} M^o_{11} & 0 & \cdots & 0 \\ M^o_{21} & M^o_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ M^o_{k1} & M^o_{k2} & \cdots & M^o_{kk} \end{bmatrix} \quad (8)$$

satisfies $M^o(i,j) = m_{l_i l_j}$ and is called the canonical form of M, where $M^o(i,j)$ denotes the element of M^o at the ith row and the jth column, each M^o_{ii} is a square matrix and we make the convention that the right hand side of (8) reduces to M^o_{11} in the case $k = 1$;

(ii) the $d_1 \times d_1$ matrix M^o_{11} is an irreducible stochastic matrix;

(iii) if $k \geq 2$, each M^o_{ii}, $i \geq 2$, is strictly substochastic (i.e., at least one row sum is less than 1) and irreducible;

(iv) the algebraic multiplicity of the eigenvalue 1 for M^o is one, and all other eigenvalues have a real part less than 1;

(v) the stochastic matrix M^o has a unique invariant probability measure of the form $\pi^o = (\pi_{l_1}, \cdots, \pi_{l_{d_1}}, 0, \cdots, 0)$ where each $\pi_{l_j} > 0$ for $1 \leq j \leq d_1$ and d_1 is the order of M^o_{11}. Moreover, π^o is independent of $\gamma \in (0, (\max_{i \in \mathcal{N}} b_{ii})^{-1})$. □

Corollary 4: [14] If G contains a spanning tree, we have:

(i) the algebraic multiplicity of the eigenvalue 1 (resp., 0) of M (resp., B) is one, and all other $n-1$ eigenvalues of M (resp., B) have a real part less than 1 (resp., 0);

(ii) with π^o given in Theorem 3, the row vector $\pi = \sum_{i=1}^{d_1} \pi_{l_i} e_{l_i}$ satisfies $\pi B = 0$, where $e_i = (0, \cdots, 0, 1, 0, \cdots, 0)$;

(iii) $\text{Null}(B) = \text{span}\{1_n\}$. □

By using the structure of M^o and translating back into the neighboring relation in G, we may classify the network structure into two scenarios.

(i) The order of M^o_{11} is greater than one; this implies leaderless consensus seeking (LLCS) where each agent receives measurements from some neighbors for updating its state.

(ii) The order of M^o_{11} is equal to one; then clearly, there exists exactly one node which does not receive data from others and its constant state is transmitted to other agents. This situation corresponds to leader following (LF).

IV. The Equivalent State Space Model

Although (6) is a linear model, most existing methods in stochastic approximation cannot be directly applied since B is unstable. We introduce the class of matrices

$$\mathscr{C}(B) = \left\{ \phi \in \mathbb{R}^{n \times (n-1)} | \text{span}\{\phi\} = \text{span}\{B\} \right\}, \quad (9)$$

where $\text{span}\{H\}$ denotes the linear space spanned by the columns of H. Under (**A1**), $\text{rank}(B) = n-1$, and accordingly, each matrix in $\mathscr{C}(B)$ has rank $n-1$. Denote $1_n = [1, \cdots, 1]^T$.

Lemma 5: Assuming (**A1**), for algorithm (6) we have:

(i) For any given $\phi_{n \times (n-1)} \in \mathscr{C}(B)$, the matrix $\Phi = (1_n, \phi_{n \times (n-1)})$ is nonsingular and

$$\Phi^{-1} B \Phi = \begin{pmatrix} 0 & \\ & \tilde{B}_{n-1} \end{pmatrix}, \quad (10)$$

where the $(n-1) \times (n-1)$ matrix $\tilde{B}_{n-1}$ is stable, i.e., all its eigenvalues have negative real parts.

(ii) Letting $z_t = (z^1_t, \cdots, z^n_t)^T = \Phi^{-1} x_t$ and $v_t = (v^1_t, \cdots, v^n_t)^T = \Phi^{-1} w_t$, we have the relation

$$z^1_{t+1} = z^1_t + a_t v^1_t, \quad (11)$$

$$z^{(n-1)}_{t+1} = (I + a_t \tilde{B}_{n-1}) z^{(n-1)}_t + a_t v^{(n-1)}_t, \qquad t \geq 0, \quad (12)$$

where $z^{(n-1)} = (z^2_t, \cdots, z^n_t)^T$ and $v^{(n-1)}_t = (v^2_t, \cdots, v^n_t)^T$.

Proof: (i) We show that $1_n \notin \text{span}\{\phi_{n \times (n-1)}\} = \text{span}\{B\}$; otherwise, there exists $\xi \in \mathbb{R}^n$ such that $1_n = B\xi$, which gives the contradiction $0 < \pi 1_n = \pi B \xi = 0$ where π is determined in Corollary 4. Hence $\Phi = (1_n, \phi_{n \times (n-1)})$ is nonsingular. Let $\Phi^{-1} = \begin{bmatrix} \Psi_1 \\ \Psi_2 \end{bmatrix}$, where Ψ_1 is the first row. Then $\Psi_1 1_n = 1$ and $\Psi_1 \phi_{n \times (n-1)} = 0$. Since $\text{span}\{\phi_{n \times (n-1)}\} = \text{span}\{B\}$, $\Psi_1 B = 0$. Recalling Corollary 4, we obtain $\Psi_1 = \pi$ and

$$\Phi^{-1} = \begin{bmatrix} \pi \\ \Psi_2 \end{bmatrix}. \quad (13)$$

Now (10) is easily verified and in fact $\tilde{B}_{n-1} = \Psi_2 B \phi_{n \times (n-1)}$. By the eigenvalue distribution of B in Corollary 4, all eigenvalues of $\tilde{B}_{n-1}$ have negative real parts.

(ii) This part follows from (6) and (10). □

Equation (12) is a linear stochastic approximation model, and since $\tilde{B}_{n-1}$ is stable, the convergence of $z^{(n-1)}_t$ can be handled by existing methods (see e.g. [29], [2]).

Lemma 6: [14], [15] Assume (**A1**) and let z_t be defined in Lemma 5. The n agents reach mean square (resp., strong) consensus if and only if z^1_t converges in mean square (resp., a.s.) to a random variable z^1_∞ and $z^{(n-1)}_t$ converges in mean square (resp., a.s.) to 0, as $t \to \infty$. □

The convergence of z_t may be completely determined by the noise sample path without assuming any statistical properties of the noise sequence.

Lemma 7: With the notation in Lemma 5, algorithm (6) ensures strong consensus if and only if the two conditions hold:

(i) $\sum_{t=0}^k a_t v^1_t$ converges a.s., as $k \to \infty$, and

(ii) $\lim_{k \to \infty} \sup_{k \leq t \leq \kappa(k,T)} |\sum_{j=k}^t a_j v^{(n-1)}_j| = 0$ a.s., where $\kappa(k,T) = \max\{j | a_k + \cdots + a_j \leq T\}$ for some constant $0 < T < \infty$.

Proof: We write (12) in the equivalent form

$$z^{(n-1)}_{t+1} = z^{(n-1)}_t + a_t \left[\tilde{B}_{n-1} z^{(n-1)}_t + v^{(n-1)}_t \right].$$

Since $\tilde{B}_n$ is stable, for any fixed sample $\omega_0 \in \Omega$, $z^{(n)}_t(\omega_0)$ converges to zero if and only if

$$\lim_{k \to \infty} \sup_{k \leq t \leq \kappa(k,T)} |\sum_{j=k}^t a_j v^{(n-1)}_j(\omega_0)| = 0. \quad (14)$$

For a proof of this fact, see e.g. [29] (Theorem 4). Equality (14) is usually called the Kushner-Clark condition [18], [29] along that noise sample path.

Hence, if (i) and (ii) hold, we have the a.s. convergence of z^1_t to z^1_∞, and $z^{(n-1)}_t$ to 0, as $t \to \infty$. By Lemma 6, strong consensus follows. Conversely, if algorithm (6) ensures strong consensus, Lemma 6 implies condition (i) and also $z_t^{(n-1)} \to 0$, a.s., which further implies condition (ii). ☐

V. Mean Square and Almost Sure Convergence

Theorem 8: Under (**A1**)-(**A3**), algorithm (6) achieves mean square consensus.

Proof: By $\sum_{i=0}^{\infty} a_i^2 < \infty$ and $\sup_t E|\mathbf{w}_t|^2 < \infty$, z_t^1 converges in mean square. We now consider $z_t^{(n-1)}$. For any $K > 0$, the Lyapunov equation $Q\tilde{B}_{n-1} + \tilde{B}_{n-1}^T Q = -K$ has a unique solution $Q > 0$. Denote $V_t = E(z_t^{(n-1)})^T Q z_t^{(n-1)}$. Similar to the stochastic Lyapunov analysis in [11], we can show that there exist constants $c_1 > 0, c_2 > 0$ such that

$$V_{t+1} \le (1 - c_1 a_t)V_t + c_2 a_t^2, \quad \forall\ t \ge T_0,$$

where T_0 is a large constant. By $\sum_{i=0}^{\infty} a_i = \infty$ and $\sum_{i=0}^{\infty} a_i^2 < \infty$, we can show $\lim_{t\to\infty} V_t = 0$, and hence $\lim_{t\to\infty} E|z_t^{(n-1)}|^2 = 0$. Mean square consensus follows from Lemma 6. ☐

Theorem 9: Under (**A1**)-(**A3**), algorithm (6) ensures strong consensus.

Proof: Since v_t defined in Lemma 5 is linear in $\mathbf{w}_t$ and $\sum_{t=0}^{\infty} a_t^2 E|\mathbf{w}_t|^2 < \infty$ by (**A2**), it follows that $\sum_{t=0}^{\infty} a_t^2 E|v_t^1|^2 < \infty$ and $\sum_{t=0}^{\infty} a_t^2 E|v_t^{(n-1)}|^2 < \infty$. By the Khintchine-Kolmogorov theorem, it follows that both $\sum_{t=0}^{k} a_t v_t^1$ and $\sum_{t=0}^{k} a_t v_t^{(n-1)}$ converge a.s., as $k \to \infty$. The convergence of $\sum_{t=0}^{k} a_t v_t^{(n-1)}$ clearly implies condition (ii) in Lemma 7. Hence, strong consensus follows from Lemma 7. ☐

Corollary 10: Given $p \in (1,2)$, we assume (**A1**) holds, (**A2**) holds after replacing $\sup_{t\ge0} E|\mathbf{w}_t|^2 < \infty$ by $\sup_{t\ge0} E|\mathbf{w}_t|^p < \infty$, and (**A3**) holds after replacing $\sum_{i=0}^{\infty} a_i^2 < \infty$ by $\sum_{i=0}^{\infty} a_i^p < \infty$. Then algorithm (6) ensures strong consensus.

Proof: We can first show the a.s. convergence of $\sum_{t=0}^{k} a_t v_t^1$ and $\sum_{t=0}^{k} a_t v_t^{(n-1)}$ ([7], pp. 114). Then strong consensus follows as in Theorem 9. ☐

VI. Asymptotic Behavior of Normalized Approximation Errors

In this section we show that asymptotically the error $x_t - x_\infty$ is normally distributed after a suitable scaling. However, since the limit of the state vector cannot be specified in advance except the (LF) scenario, the error term depends on future behavior of the iteration. This differs from the typical analysis in the literature [2]. To simplify the analysis, we make additional restrictions for the noise sequence. Let $\mathbf{w}_t$ be specified as in Section II.

(**A2'**) The sequence $\{\mathbf{w}_t, t \in \mathbb{Z}^+\}$ constitutes i.i.d. vector random variables with zero mean and finite covariance $Q_\mathbf{w}$ and is independent of x_0, where $E|x_0|^2 < \infty$. ☐

(**A3'**) The positive sequence $\{a_t, t \ge 0\}$ satisfies $a_t = at^{-1}$, where $a > 0$, for $t \ge \hat{T}$ for some fixed $\hat{T} > 0$. ☐

Letting $x_\infty = x_\infty^1 1_n$ be the limit state vector and $\phi_{n\times(n-1)}$ be selected as in Lemma 5, we have the decomposition

$$\begin{aligned} x_t &= z_t^1 1_n + \phi_{n\times(n-1)} z_t^{(n-1)} \\ &= x_\infty^1 1_n + (z_t^1 - z_\infty^1)1_n + \phi_{n\times(n-1)} z_t^{(n-1)} \end{aligned} \tag{15}$$

Define

$$x_t^{e,a} = (z_t^1 - z_\infty^1)1_n, \quad x_t^{e,b} = \phi_{n\times(n-1)} z_t^{(n-1)}. \tag{16}$$

Thus, the approximation error for x_t is decomposed into two components $x_t^{e,a}$ and $x_t^{e,b}$ to give

$$x_t - x_\infty^1 1_n = x_t^{e,a} + x_t^{e,b}. \tag{17}$$

For the (LF) scenario with agent k_0 being the leader, $x_t^{e,a} = 0$ a.s. since $z_t^1 \equiv x_0^{k_0}$ for all $t \ge 0$.

Lemma 11: Under (**A1**)-(**A3**), the decomposition in (16)-(17) does not depend on the choice of $\phi_{n\times(n-1)} \in \mathscr{C}(B)$ where $\mathscr{C}(B)$ is defined by (9).

Proof: First, x_t does not depend on $\phi_{n\times(n-1)}$. Also, in

$$x_t^{e,a} = (\pi x_t - x_\infty^1)1_n, \tag{18}$$

π is determined by B and does not change with $\phi_{n\times(n-1)}$. So $x_t^{e,a}$ and consequently $x_t^{e,b}$, do not change with $\phi_{n\times(n-1)}$. ☐

Remark: In fact, by elementary linear algebra, we can directly verify that $\phi_{n\times(n-1)}\Psi_2$ and hence $x_t^{e,b}$ remain the same for all $\phi_{n\times(n-1)} \in \mathscr{C}(B)$. ☐

Lemma 11 shows an invariance property of the decomposition (17) with respect to all $\phi_{n\times(n-1)} \in \mathscr{C}(B)$. This property is useful in practical computation since it suffices to take any fixed $\phi_{n\times(n-1)} \in \mathscr{C}(B)$. We note that the error component $x_t^{e,b}$ may be evaluated on-line since $x_t^{e,b}$ is a function of x_t, but $x_t^{e,a}$ can not be observed on-line since it depends on the future noise sequence.

Let $\zeta_i, i \ge 0$ be a sequence of (scalar or vector) random variables. If ζ_i converges in distribution to a normal random variable with mean μ and covariance Σ, we denote $\zeta_i \xrightarrow{d} N(\mu, \Sigma)$. For additional materials on convergence in distribution or weak convergence, see e.g. [7] (pp. 286).

Let

$$Q' = a^2 \int_0^\infty e^{(a\tilde{B}_{n-1} + I/2)t} Q_{v_0^{(n-1)}} e^{(a\tilde{B}_{n-1}^T + I/2)t} dt \tag{19}$$

where $Q_{v_0^{(n-1)}}$ is the covariance of $v_0^{(n-1)}$, and set $Q_b = \phi_{n\times(n-1)} Q' \phi_{n\times(n-1)}^T$. Denote the variance of v_0^1 by $\sigma^2(v_0^1)$, and $Q_a = a^2\sigma^2(v_0^1)1_n 1_n^T$.

Theorem 12: Assume (**A1**), (**A2'**) and (**A3'**) hold, and $a\tilde{B}_{n-1} + I/2$ is a stable matrix. Then $x_t^{e,a} \xrightarrow{d} N(0, Q_a)$, $\sqrt{t}x_t^{e,b} \xrightarrow{d} N(0, Q_b)$, $\sqrt{t}(x_t^{e,a} + x_t^{e,b}) \xrightarrow{d} N(0, Q_a + Q_b)$.

Proof (Sketch): We use the characteristic function approach. Let $\mathbf{i}$ denote the imaginary unit and ξ a real number. We have $z_t^{1,e} \triangleq \sqrt{t}(z_t^1 - z_\infty^1) = \sqrt{t}\sum_{k=t}^{\infty} a_k(-v_k^1)$ and

$$e^{\mathbf{i}\xi z_t^{1,e}} = \lim_{T\to\infty} E e^{\mathbf{i}\xi\sqrt{t}\sum_{k=t}^{T} a_k(-v_k^1)} = \prod_{k=t}^{\infty} g_{v_0^1}(\xi a\sqrt{t}/k),$$

where $g_{v_0^1}(\xi) = Ee^{\mathbf{i}\xi(-v_0^1)}$. By the second order Taylor expansion of $g_{v_0^1}$, we may show that for any given $C > 0$ and $\xi \in [-C, C]$, $Ee^{\mathbf{i}\xi z_t^{1,e}} = e^{-\xi^2 a^2 \sigma^2(v_0^1)/2} + o(1)$, as $t \to \infty$. Hence $z_t^{1,e} \xrightarrow{d} N(0, a^2\sigma^2(v_0^1))$, which implies $x_t^{e,a} \xrightarrow{d} N(0, Q_a)$.

By [20] (pp. 147), $\sqrt{t}z_t^{(n-1)} \xrightarrow{d} N(0,Q')$, which implies $\sqrt{t}x_t^{e,b} \xrightarrow{d} N(0,Q_b)$. By the independence of $x_t^{e,a}$ and $x_t^{e,b}$ for each t, we further obtain $\sqrt{t}(x_t^{e,a}+x_t^{e,b}) \xrightarrow{d} N(0,Q_a+Q_b)$. □

Remark: If, in addition to the conditions in Theorem 12, the noise vector w_t in (6) has a non-degenerate covariance, i.e., $Q_w \triangleq Ew_tw_t^T > 0$, we can show that $Q_a+Q_b > 0$ by first checking $\sigma^2(v_0^1) > 0$ and $Q_{v_0^{(n-1)}} > 0$. Thus, $N(0,Q_a+Q_b)$ has a density in $\mathbb{R}^n$. □

VII. Randomly Time-Varying Communication Links

Let us first introduce a fixed digraph $G=(\mathcal{N},\mathcal{E})$ which describes the maximal set of communication links when there is no link failure. At time t the inter-agent communication is described by a subgraph of G denoted by $G_t=(\mathcal{N},\mathcal{E}_t)$ where $\mathcal{E}_t \subset \mathcal{E}$; the edge $(i,j)\in\mathcal{E}_t$ if and only if there exists a communication link from i to j at time t where $(i,j)\in\mathcal{E}$. The digraph G_t is generated as the outcome of random link failures. Note that an edge (i,j) never appears in G_t if it is not an edge of G. The neighbor set of node i is $\mathcal{N}_{it}=\{k|(k,i)\in\mathcal{E}_t\}$ at time t.

At time $t\geq 0$, the adjacency matrix of G_t is defined as $A_t^{\mathcal{N}}=(a_t^{ij})_{1\leq i,j\leq n}$, where $a_t^{ij}=1$ if $(i,j)\in\mathcal{E}_t$, and $a_t^{ij}=0$ otherwise. The digraph G_t is completely characterized by the random matrix $A_t^{\mathcal{N}}$.

Now, the measurement relation is given as

$$y_t^{ik}=x_t^k+w_t^{ik} \qquad \text{if} \quad a_t^{ki}=1 \ \text{(i.e., } k\in\mathcal{N}_{it}),$$

where w_t^{ik} is the additive noise. The state of agent $i\in\mathcal{N}$ is updated by the rule

$$x_{t+1}^i=(1-a_t|\mathcal{N}_{it}|)x_t^i+a_t\sum_{k\in\mathcal{N}_{it}}y_t^{ik}. \tag{20}$$

If $\mathcal{N}_{it}=\emptyset$, (20) reduces to $x_{t+1}^i=x_t^i$. Here for simplicity we assign the same weight to the $|\mathcal{N}_{it}|$ observations y_t^{ik}.

For specifying the statistical properties of the noises, we introduce the array of measurement noises as a square matrix:

$$W_t=(w_t^{ik})_{1\leq i,k\leq n},$$

where it is restricted from the beginning that $w_t^{ik}\equiv 0$ if $(k,i)\notin\mathcal{E}$. It is sufficient to further specify w_t^{ik} with $(k,i)\in\mathcal{E}$. The combined link and noise assumption is stated below.

(**A4**) (i) For $(i,j)\in\mathcal{E}$, $P\{a_t^{ij}=1\}=P\{(i,j)\in\mathcal{E}_t\}=p_{ij}>0$, (ii) the pair $(A_t^{\mathcal{N}},W_t)$ is independent of $(x_0,A_k^{\mathcal{N}},W_k,k\leq t-1)$ where $t\geq 0$, and (iii) conditioned on $A_t^{\mathcal{N}}=(a_t^{ij})_{1\leq i,j\leq n}$, the noises ($w_t^{ik}$, with all pairs $(k,i)\in\mathcal{E}$) are independent and satisfy

$$P(w_t^{ik}=0|a_t^{ki}=0)=1,$$
$$E(w_t^{ik}|a_t^{ki}=1)=0, \qquad \sup_{i,k,t}E(|w_t^{ik}|^2|a_t^{ki}=1)\leq C_w$$

where $C_w<\infty$ is a constant. The term $(x_0,A_k^{\mathcal{N}},W_k,k\leq t-1)$ is interpreted as x_0 when $t=0$. □

Remark: If we further define the distribution of w_t^{ik} conditioned on $\{a_t^{ki}=1\}$, then any finite dimensional distribution of $(x_0,A_k^{\mathcal{N}},W_k,k\leq t)$ is well defined. □

We still denote $x_t=(x_t^1,\cdots,x_t^n)^T$. Define $w_t^i=\sum_{k\in\mathcal{N}_{it}}a_t^{ki}w_t^{ik}$ and the noise vector

$$w(A_t^{\mathcal{N}},W_t)=(w_t^1,\cdots,w_t^n)^T.$$

Let $D(A_t^{\mathcal{N}})=\mathrm{Diag}(\sum_{k\in\mathcal{N}}a_t^{k1},\cdots,\sum_{k\in\mathcal{N}}a_t^{kn})$. We write (20) in the vector form

$$x_{t+1}=x_t+a_t[(A_t^{\mathcal{N}})^T-D(A_t^{\mathcal{N}})]x_t+a_tw(A_t^{\mathcal{N}},W_t), \qquad t\geq 0. \tag{21}$$

Let $B_t=[(A_t^{\mathcal{N}})^T-D(A_t^{\mathcal{N}})]$, $\overline{B}=EB_t$ and $\Delta B_t=B_t-\overline{B}$, and it can be shown that all row sums of both B_t and $\overline{B}$ are zero. Now (21) may be written in the form

$$x_{t+1}=x_t+a_t\overline{B}x_t+a_t\Delta B_tx_t+a_tw(A_t^{\mathcal{N}},W_t), \qquad t\geq 0. \tag{22}$$

Lemma 13: [14] Assuming (**A4**) holds, we have

(i) for $t\geq 0$, the pair $(\Delta B_t,w(A_t^{\mathcal{N}},W_t))$ is independent of x_t, and $Ew(A_t^{\mathcal{N}},W_t)=0$;

(ii) if, in addition, G contains a spanning tree, then $\overline{B}$ satisfies: a) it has a zero eigenvalue of algebraic multiplicity equal to one, and $n-1$ eigenvalues with negative real parts; b) there exists a unique probability measure $\bar{\pi}$ such that $\bar{\pi}\overline{B}=0$; c) $\mathrm{Null}(\overline{B})=\mathrm{span}\{1_n\}$. □

Equation (22) can be viewed as a perturbed version of (6) where the additional term $a_t\Delta B_tx_t$ is unbiased in the sense ΔB_tx_t has zero mean under (**A4**). The proof of convergence of (22) (or (21)) follows similar ideas as in analyzing (6).

Theorem 14: [14], [15] Under (**A1**), (**A3**) and (**A4**), algorithm (21) ensures mean square consensus, i.e., there exists x^* such that $\lim_{t\to\infty}E|x_t^i-x^*|^2=0$ for all $i\in\mathcal{N}$. □

Proof: The theorem may be proved by using (22) and a linear transform to generate a perturbed version of (11)-(12). See [14], [15] for details. □

VIII. Simulations–Asymptotic Behavior

Consider a digraph shown in Fig. 2. The noises $\{w_t^{12},w_t^{21},w_t^{23},w_t^{31},t\geq 0\}$ are i.i.d. and uniformly distributed on the interval $[-0.15,0.15]$. The initial state vector is $x_t|_{t=0}=[4,3,1]^T$. The algorithm (6) is applied by taking

$$B=\begin{bmatrix} -1 & 1 & 0 \\ 0.5 & -1 & 0.5 \\ 1 & 0 & -1 \end{bmatrix}, \tag{23}$$

and $a_0=0.5$, $a_t=0.5t^{-1}$ for $t\geq 1$. The 3 eigenvalues of B are 0, $-1.5\pm 0.5\mathbf{i}$. We use the first two columns in B to construct

$$\Phi=\begin{bmatrix} 1 & -1 & 1 \\ 1 & 0.5 & -1 \\ 1 & 1 & 0 \end{bmatrix}$$

and calculate $x_t^{e,b}$ by (16). The asymptotic normality conclusion of Theorem 12 holds for this example since $a\tilde{B}_{n-1}+I/2$ is stable with eigenvalues $-0.25\pm 0.25\mathbf{i}$ when B is given by (23) and $a=0.5$. The convergence of x_t is shown in Fig. 3 which displays the first 10^3 iterates, and the corresponding scaled error process $\{\sqrt{t}x_t^{e,b},t\geq 0\}$ is displayed in Fig. 4 with a maintained magnitude in the long term while $x_t^{e,b}$ itself vanishes when $t\to\infty$.

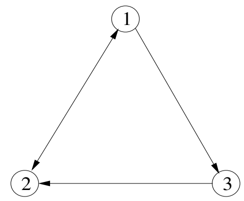

Fig. 2. The digraph with 3 nodes.

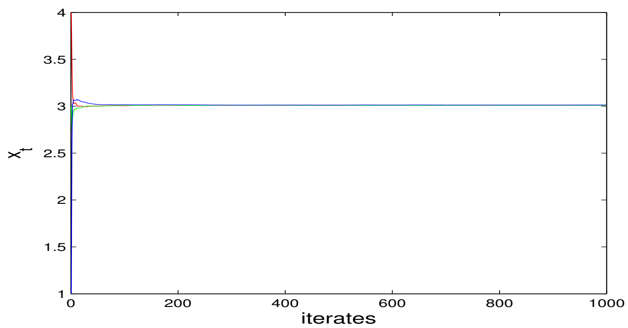

Fig. 3. Convergence of the 3 trajectories using a decreasing step size.

IX. Conclusion

We consider stochastic consensus problems with measurement noise. Stochastic approximate algorithms are applied to achieve consensus. We also present consensus error analysis via asymptotic normality results.

References

[1] P. Barooah and J. P. Hespanha. Graph effective resistence and distributed control: special properties and applications. *Proc. 45th IEEE CDC Conf.*, San Diego, CA, pp. 3479-3485, Dec. 2006.

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

[3] 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.

[4] P. E. Caines. *Linear Stochastic Systems*. John Wiley, New York, 1988.

[5] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri. Communication constraints in coordinated consensus problems. *Proc. American Control Conference*, Minneapolis, MN, pp. 4189-4194, June 2006.

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

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

[8] A. Giridhar and P. R. Kumar. Toward a theory of in-network computation in wireless sensor networks. *IEEE Commun. Mag.*, pp. 98-107, April 2006.

[9] Y. Hatano and M. Mesbahi. Agreement in random networks. *IEEE Trans. Automat. Control*, vol. 51, no. 11, pp. 1867-1872, Nov. 2005.

[10] 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.

[11] 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.

[12] M. Huang and J. H. Manton. Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior. Submitted to *SIAM J. Control Optimization*, Oct. 2006.

[13] M. Huang and J. H. Manton. Stochastic approximation for consensus-seeking: mean square and almost sure convergence. *Proc. 46th IEEE CDC Conference*, New Orleans, LA, pp. 306-311 Dec. 2007.

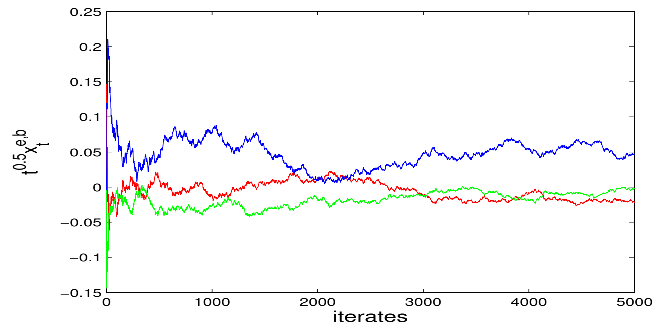

Fig. 4. The scaled error component $\sqrt{t}x_t^{e,b}$.

[14] M. Huang and J. H. Manton. Stochastic consensus seeking with directed inter-agent communication: convergence and asymptotic normality. Preprint, June 2007.

[15] M. Huang and J. H. Manton. Stochstic consensus seeking with noisy and directed inter-agent communication. Submitted to *IEEE Trans. Autom. Control* in June 2007, revised Dec. 2007.

[16] 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, pp. 988-1000, June 2003.

[17] S. Kar and J. M. Moura. Distributed average consensus in sensor networks with random link failures. *Proc. IEEE Int. Conf. Acoustics, Speech, and Sig. Process.*, Honolulu, HI, pp. II-1013-1016, April 2007.

[18] H. J. Kushner and D. S. Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer-Verlag, New York, 1978.

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

[20] M. B. Nevel'son and R. Z. Has'minskii. *Stochastic Approximation and Recursive Estimation*. AMS, Providence, Rhode Island, 1973.

[21] 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.

[22] 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, Jan. 2007.

[23] M. Porfiri and D. J. Stilwell. Consensus seeking over random weighted directed graphs. *IEEE Trans. Autom. Control*, vol. 52, pp. 1767-1773, Sept. 2007.

[24] W. Ren and R. W. Beard. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Trans. Automat. Control*, vol. 50, no. 5, pp. 655-661, 2005.

[25] 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.

[26] 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.

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

[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] I.-J. Wang, E. K. P. Chong, and S. R. Kulkarni. Weighted averaging and stochastic approximation. *Math. Control Signals and Systems*, vol. 10, pp. 41-60, 1997.

[30] C. W. Wu. Synchronization and convergence of linear dynamics in random directed networks. *IEEE Trans. Autom. Control*, vol. 51, pp. 1207-1210, July, 2006.

[31] 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.