Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

WeC12.6

Convergence Rate for Stochastic Consensus Algorithms with Time-Varying Noise Statistics: Asymptotic Normality

Minyi Huang

Abstract— This paper studies consensus seeking over noisy networks with time-varying noise statistics. Stochastic approximation type algorithms can ensure consensus in mean square and with probability one. For performance evaluation, we examine the long term behavior of the approximation error which consists of two naturally defined components. We show that the two components and their sum are each asymptotically normal after being normalized by the square root of time. This, in turn, characterizes the convergence rate of the algorithm. We also give the asymptotic formula for the scaled error covariances.

I. INTRODUCTION

Recent technological advances have made it possible to build large distributed systems in which different constituent components or agents may cooperatively perform complex tasks. In these systems, consensus protocols provide a basic mechanism for the agents to agree on key information concerning system operation. Consensus problems and various closely related formulations have been intensively investigated for multi-agent systems [14], [17], [20]. A comprehensive survey on recent research can be found in [18], [22]. While most existing consensus algorithms have assumed exact state averaging, which in general necessitates perfect state exchange, recently, there is an increasing attention on models with noise or quantization effect [21], [27], [11], [3], [1], [9], [4], [24]. The work [25] made early effort introducing stochastic gradient based consensus algorithms. For consensus or synchronization problems based on random graphs, see [8], [26], [19].

In consensus models with noisy measurements, the traditional algorithms involving constant or lowered bounded averaging weights in general cannot ensure convergence. In [9], [10], [11], stochastic approximation type algorithms with a decreasing step size were applied for consensus seeking where the state information of other agents is corrupted by white noise (see Fig. 1). In particular, probability one convergence results were obtained in [9] via a double array analysis in digraph models satisfying a circulant invariance property. Mean square convergence was proved for connected undirected graphs by using a stochastic Lyapunov function [10]. The analysis in [9], [10] was generalized to digraphs containing a spanning tree in [12], [13].

In this paper, we aim to develop performance analysis for stochastic consensus algorithms on digraphs. We examine the asymptotic behavior of the approximation error, i.e., the difference between the state vector and its limit, the latter

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

being a vector in span$\{1_n\}$ where $1_n = [1, \cdots, 1]^T$. We give a linear decomposition of the error into two parts where the first characterizes the oscillation within span$\{1_n\}$. We will show that after normalization, the distributions of these two error components each converge weakly to a normal distribution; this, in turn, characterizes the convergence rate of the algorithm. Some preliminary asymptotic normality analysis has been developed in [12] under i.i.d. noise assumptions. In this paper, we consider much more general noise sequences without assuming independence. We allow time-varying noise statistics by only specifying certain long term average behavior of the covariances and conditional covariances; such assumptions are applicable to models occasionally experiencing burst receiver noises while the previous i.i.d. assumptions are not. Our proof adopts the classic central limit theorem approach [23], [16], [5], [7], but some new techniques need to be developed in order to deal with the time-varying (non-convergent) noise covariances. For general asymptotic error analysis in stochastic approximation, the reader is referred to [2], [5], [15], [16].

The organization of the paper is as follows. In Section II, we formulate the stochastic consensus problem and review our previous convergence results. The main theorem on asymptotic normality is stated in Section III. Section IV contains simulations and Section V concludes the paper.

II. THE STOCHASTIC CONSENSUS PROBLEM

Consider n agents distributed according to a digraph $G = (\mathcal{N}, \mathcal{E})$ consisting of a set of nodes $\mathcal{N} = \{1, \cdots, n\}$ and a set of directed edges $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$. 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) consists of a sequence of nodes $i_1, i_2, \cdots, i_l$, $l \geq 2$, such that $(i_k, i_{k+1}) \in \mathcal{E}$ for $k = 1 \cdots, l-1$. We say node i is connected to node $j (\neq i)$ if there exists a directed path from i to j. The digraph G is said to be strongly connected if each node i is connected to any other node j by a directed 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 directed path. The digraph G contains 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 $\mathcal{N}_i = \{k | (k, i) \in \mathcal{E}\} \subset \mathcal{N}$.

A. The Measurement Model

For agent A_i, denote its state at time t by $x_t^i \in \mathbb{R}$, where $t \in \mathbb{Z}^+ = \{0, 1, 2, \cdots\}$. Each A_i receives noisy measurements of the states of its neighbors if $\mathcal{N}_i \neq \emptyset$, where $\emptyset$ denotes the empty set. Denote the measurement by A_i of A_k's state by

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

where $w_t^{ik} \in \mathbb{R}$ is the additive noise; see Fig. 1. 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. For similar measurement modeling, see [9], [4], [24]. 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 $\{w_t^{ik}, i \in \mathcal{N} \text{ and } k \in \mathcal{N}_i \neq \emptyset\}$ 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

$$x_t = [x_t^1, \cdots, x_t^n]^T, \qquad t \geq 0. \qquad (2)$$

Denote the σ-algebras as follows: $\mathcal{F}_t = \sigma(x_0, \mathbf{w}_k, k \leq t)$ (i.e., the set of all events induced by these random variables) for $t \geq 0$, $\mathcal{F}_{-1} = \{\emptyset, \Omega\}$. Then $\mathbf{w}_t$ is adapted to (i.e., measurable on) $\mathcal{F}_t$ and $\mathcal{F}_t \subset \mathcal{F}_{t+1}$. We introduce the assumption:

(**A2**) The sequence $\{\mathbf{w}_t, t \in \mathbb{Z}^+\}$ constitutes a sequence of martingale differences with a uniformly bounded second order moment, i.e., $\mathbf{w}_t$ is adapted to $\mathcal{F}_t$, $E[\mathbf{w}_t | \mathcal{F}_{t-1}] = 0$ for $t \geq 0$ and $\sup_{t \geq 0} E|\mathbf{w}_t|^2 < \infty$. In addition, $E|x_0|^2 < \infty$. ☐

The following assumption with independent noises holds as a special case of (**A2**).

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

B. The Stochastic Approximation Algorithm

The state of each agent is updated by the rule

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

where $i \in \mathcal{N}$, $a_t > 0$ and the parameters b_{ij} will be specified subsequently. Throughout our analysis, we adopt the convention: $\sum_{k \in \emptyset} = 0$ regardless of the summand.

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

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

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 as its initial value: $x_t^i \equiv x_0^i$. Such a situation arises in leader following where the leader's state is fixed as a constant at all times.

Define the matrix

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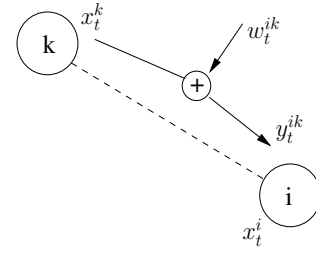

Fig. 1.　Measurement with noise w_t^{ik}.

Define

$$w_t^i = \sum_{k \in \mathcal{N}_i} b_{ik} w_t^{ik}, \qquad w_t = [w_t^1, \cdots, w_t^n]^T \qquad (5)$$

where $w_t^i = \sum_{k \in \emptyset} = 0$ if $\mathcal{N}_i = \emptyset$. Write (3) in the vector form

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

(**A3**) The sequence $\{a_t, t \geq 0\}$ satisfies i) $a_t > 0$ and ii) $\sum_{t=0}^{\infty} a_t = \infty$, $\sum_{t=0}^{\infty} a_t^2 < \infty$. ☐

The right hand side of (3) is a convex combination of the agent's state and its $|\mathcal{N}_i|$ observations if $a_t b_{ii} \leq 1$.

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

Convergence with probability one (w.p. 1) is also called almost sure (a.s.) convergence. We cite a convergence result.

Theorem 3: [13] Under (**A1**)-(**A3**), algorithm (6) achieves both mean square and strong consensus. ☐

C. Preliminary Decomposition Results

We introduce the following class of matrices in $\mathbb{R}^{n \times (n-1)}$

$$\mathscr{C}(B) = \left\{ \phi = (\zeta_1, \cdots, \zeta_{n-1}) | \text{span}\{\phi\} = \text{span}\{B\} \right\}, \qquad (7)$$

where $\text{span}\{\phi\}$ denotes the subspace spanned by the columns of ϕ. Under (**A1**), $\text{rank}(B) = n - 1$, and accordingly, each $\phi \in \mathscr{C}(B)$ has rank $n - 1$ (see [12]).

Lemma 4: [12] Assuming (**A1**), for (6) we have:

(i) For $1_n = [1, \cdots, 1]^T$ and any given $\bar{\phi} \in \mathscr{C}(B)$, the matrix $\Phi = (1_n, \bar{\phi})$ is nonsingular and

$$\Phi^{-1} B \Phi = \begin{pmatrix} 0 & \\ & \tilde{B} \end{pmatrix}, \qquad (8)$$

where $\tilde{B} \in \mathbb{R}^{(n-1) \times (n-1)}$ is Hurwitz.

(ii) Letting $z_t = [z_t^1, \cdots, z_t^n]^T = \Phi^{-1} x_t$ and

$$v_t = [v_t^1, \cdots, v_t^n]^T = \Phi^{-1} w_t, \qquad (9)$$

we have the relation

$$z_{t+1}^1 = z_t^1 + a_t v_t^1, \qquad (10)$$

$$\tilde{z}_{t+1} = (I + a_t \tilde{B}) \tilde{z}_t + a_t \tilde{v}_t, \qquad t \geq 0, \qquad (11)$$

where $\tilde{z}_t = [z_t^2, \cdots, z_t^n]^T$ and $\tilde{v}_t = [v_t^2, \cdots, v_t^n]^T$.   □

In fact, the first row of Φ^{-1} is given as a unique nonnegative vector $\pi = [\pi_1, \cdots, \pi_n]$ satisfying $\pi B = 0$ and $\sum_{k=1}^n \pi_k = 1$. And under (**A1**)-(**A3**), z_t converges to $z_\infty = [z_\infty^1, 0, \cdots, 0]^T$ in mean square and w.p.1 (see [12]).

III. MAIN RESULTS ON ASYMPTOTIC NORMALITY

For Theorem 3, denote the limit state vector by $x_\infty = [x_\infty^1, \cdots, x_\infty^n]^T = x_\infty^1 1_n$. Since $z_t^1 = \pi x_t$, we obtain $z_\infty^1 = \pi(x_\infty^1 1_n) = x_\infty^1$. By the relation $x_t = x_\infty^1 1_n + \Phi z_t - x_\infty^1 1_n$, we obtain the following decomposition:

$$x_t = x_\infty^1 1_n + (z_t^1 - z_\infty^1) 1_n + \bar{\phi} \tilde{z}_t. \tag{12}$$

Thus, under (**A1**)-(**A3**) the approximation error for x_t is decomposed into two components $x_t^{e,a} = (z_t^1 - z_\infty^1) 1_n$ and $x_t^{e,b} = \bar{\phi} \tilde{z}_t$ to give

$$x_t - x_\infty^1 1_n = x_t^{e,a} + x_t^{e,b}. \tag{13}$$

Clearly, $x_t^{e,a} \in \text{span}\{1_n\}$ and $x_t^{e,b} \in \text{span}\{B\}$. Under (**A1**), using the property $\pi B = 0$, we may show that

$$\text{span}\{1_n\} \cap \text{span}\{B\} = \{0\}, \quad \mathbb{R}^n = \text{span}\{1_n\} \oplus \text{span}\{B\}$$

where $\oplus$ denotes direct sum. Hence, $x_t - x_\infty^1 1_n$ has a unique representation as the sum of two vectors (independent of $\bar{\phi}$) in $\text{span}\{1_n\}$ and $\text{span}\{B\}$, respectively. Alternatively, we may use the fact that x_t, $x_\infty^1 1_n$ and $z_t^1 = \pi x_t$, z_∞^1 are all independent of $\bar{\phi}$ to check that the decomposition in (13) does not depend on the choice of $\bar{\phi} \in \mathscr{C}(B)$. For the case of leader following, we can show that $v_t^1 \equiv 0$ and $x_t^{e,a} \equiv 0$, and the asymptotic error analysis reduces to checking $x_t^{e,b}$.

We introduce some assumption related to $\mathbf{w}_t$, and it will be convenient to give the condition based on v_t, which is defined via (5) and (9).

(**A4**) The sequence $\{\tilde{v}_t, t \in \mathbb{Z}^+\}$ constitutes vector random variables with zero mean and covariance $Q_t^{\tilde{v}}$ such that

$$\lim_{T \to \infty} (1/T) \sum_{i=k}^{k+T-1} Q_i^{\tilde{v}} = \overline{Q}^{\tilde{v}} \tag{14}$$

uniformly w.r.t. $k \geq 0$, and in addition

$$\lim_{T \to \infty} (1/T) \sum_{i=1}^T E \left| E[\tilde{v}_i \tilde{v}_i^T | \tilde{v}_0, \cdots, \tilde{v}_{i-1}] - Q_i^{\tilde{v}} \right| = 0, \tag{15}$$

$$\lim_{K \to \infty} \sup_{t \geq 0} E |\tilde{v}_t|^2 1_{(|\tilde{v}_t| \geq K)} = 0. \tag{16}$$

Letting $\sigma_i^2 = E|v_i^1|^2$, then

$$\lim_{T \to \infty} (1/T) \sum_{i=k}^{k+T-1} \sigma_i^2 = \bar{\sigma}^2, \tag{17}$$

$$\lim_{T \to \infty} (1/T) \sum_{i=k}^{k+T-1} E \left| E[|v_i^1|^2 | v_0^1, \cdots, v_{i-1}^1] - \sigma_i^2 \right| = 0, \tag{18}$$

both uniformly with respect to k. Finally

$$\lim_{K \to \infty} \sup_{t \geq 0} E |v_t^1|^2 1_{(|v_t^1| \geq K)} = 0. \tag{19}$$

□

For the special case $\{\mathbf{w}_t, t \geq 0\}$ being an i.i.d. sequence with zero mean and finite covariance, (14)-(19) are satisfied. It must be noted that the validity of (**A4**) does not depend on the choice of $\bar{\phi}$. More specifically, when a different $\hat{\phi}$ is used, (**A4**) is still true as long as all the associated variances and conditional variances correspond to the new $\hat{\phi}$.

(**A3'**) The sequence $\{a_t, t \geq 0\}$ satisfies i) $a_t > 0$, ii) $\lim_{t \to \infty}(a_{t+1}^{-1} - a_t^{-1}) = \alpha > 0$, iii) $\hat{B} \triangleq \tilde{B} + \alpha I/2$ is Hurwitz. □

It is evident that (**A3'**) implies (**A3**). If a sequence of random variables $\{\xi_t, t \geq 0\}$ converges in distribution to a normal random variable ξ_∞ with mean μ and covariance Σ, we denote $\xi \xrightarrow{d} N(\mu, \Sigma)$ and $\xi \xrightarrow{d} \xi_\infty$.

Denote $D = \int_0^\infty e^{\hat{B}t} \overline{Q}^{\tilde{v}} e^{\hat{B}^T t} dt$, and

$$D_a = \alpha^{-2} \bar{\sigma}^2 1_n 1_n^T, \quad D_b = \alpha^{-1} \bar{\phi} D \bar{\phi}. \tag{20}$$

Theorem 5: Assuming (**A1**), (**A2**), (**A3'**) and (**A4**) hold, we have i) $\sqrt{t} x_t^{e,a} \xrightarrow{d} N(0, D_a)$ and $\sqrt{t} x_t^{e,b} \xrightarrow{d} N(0, D_b)$ and ii) $\sqrt{t}(x_t^{e,a} + x_t^{e,b}) \xrightarrow{d} N(0, D_a + D_b)$.   □

To prove Theorem 5-i), it suffices to establish the two lemmas below. Their proofs are quite technical and are given in Appendix B. Theorem 5-ii) may be proved by first approximating $\sqrt{t}(x_t^{e,a} + x_t^{e,b})$ by the sum of finite terms of martingale differences (similar to the treatment in proving Lemma 7), and next carry out the asymptotic characteristic function estimation with $t \to \infty$.

Lemma 6: Under the assumptions of Theorem 5, $\sqrt{t} \tilde{z}_t \xrightarrow{d} N(0, \alpha^{-1} D)$.   □

Lemma 7: Under the assumptions of Theorem 5, $\sqrt{t}(z_t^1 - z_\infty^1) \xrightarrow{d} N(0, \alpha^{-2} \bar{\sigma}^2)$.   □

Corollary 8: Under the assumptions of Theorem 5,

$$\lim_{t \to \infty} t E\{x_t^{e,a}(x_t^{e,a})^T\} = D_a = (\alpha^{-2} \bar{\sigma}^2) 1_n 1_n^T, \tag{21}$$

$$\lim_{t \to \infty} t E\{x_t^{e,b}(x_t^{e,b})^T\} = D_b = \alpha^{-1} \bar{\phi} D \bar{\phi}^T, \tag{22}$$

$$\lim_{t \to \infty} t E\{(x_t^{e,a} + x_t^{e,b})(x_t^{e,a} + x_t^{e,b})^T\} = D_a + D_b. \tag{23}$$

Proof: We obtain (21) from a direct calculation of $\lim_{t \to \infty} t \sum_{k=t}^\infty a_k^2 \sigma_k^2$. By (B.3) and (B.7), we obtain (22). To show (23), we first take expectation to eliminate the two cross terms in the expansion by using the martingale difference property of the terms in the series representation of $x_t^{e,a}$ and $x_t^{e,b}$, and the right hand side follows from (21) and (22).   □

Again, we remark that both $(\alpha^{-2} \bar{\sigma}^2) 1_n 1_n^T$ and $\alpha^{-1} \bar{\phi} D \bar{\phi}^T$ are independent of the particular choice of $\bar{\phi} \in \mathscr{C}(B)$.

IV. SIMULATIONS

We 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 independent and satisfy: for $k = 0, 1, \cdots$, (i) if $20k \leq t \leq 20k + 17$, each has a uniform distribution on $[-0.2, 0.2]$ (with $\sigma_w^2 = 0.01333$); (ii) if $20k + 18 \leq t < 20(k+1)$, each has a uniform distribution on $[-0.7, 0.7]$ (with $\sigma_w^2 = 0.1633$). Thus, the noise variances periodically reach a much higher level, which models burst receiver noises. The initial state vector is $x_t|_{t=0} = [5, 4, 2]^T$.
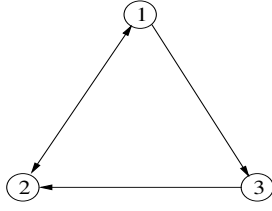
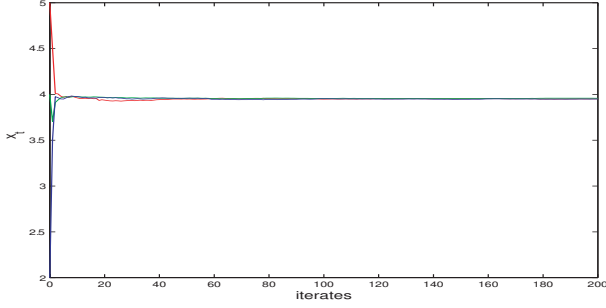Fig. 2.   The digraph with 3 nodes.

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

For algorithm (6), we take

$$B = \begin{bmatrix} -1 & 1 & 0 \\ 0.5 & -1 & 0.5 \\ 1 & 0 & -1 \end{bmatrix}, \quad \Phi = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0.5 & -1 \\ 1 & 1 & 0 \end{bmatrix} \quad (24)$$

and $a_0 = 0.5$, $a_t = \frac{0.5}{t}$ for $t \geq 1$. The 3 eigenvalues of B are 0, $-1.5 \pm 0.5\mathbf{i}$. The first two columns in B are used to construct Φ. We express $x_t^{e,b}$ by $\tilde{z}_t$ and consequently by x_t. The asymptotic normality conclusion of Theorem 5 holds for this example since (**A4**) is satisfied and $\tilde{B} + \alpha I/2$ is Hurwitz with eigenvalues $-0.5 \pm 0.5\mathbf{i}$ when the associated B is given by (24) and $\alpha = 2$. The convergence of x_t is shown in Fig. 3 which displays the first 200 iterates, and $\{\sqrt{t}x_t^{e,b}, t \geq 0\}$ is displayed in Fig. 4.

V. CONCLUSIONS

We have presented asymptotic normality results for the scaled error terms in stochastic consensus algorithms. Our analysis is applicable to average consensus based algorithms with additive noises, which amounts to imposing additional conditions (i.e. B has zero row and column sums, corresponding to averaging with balanced graphs [17]).

APPENDIX A: PRELIMINARY LEMMAS

We need some preliminary lemmas before proving Lemmas 6 and 7 in Appendix B.

Lemma 9: [23], [16], [5] Suppose $\{\xi_{tk}, t \geq 1, 1 \leq k \leq t\}$ forms an array of martingale differences, i.e.,

$$E[\xi_{tk}|\xi_{t1}, \cdots, \xi_{t(k-1)}] = 0.$$
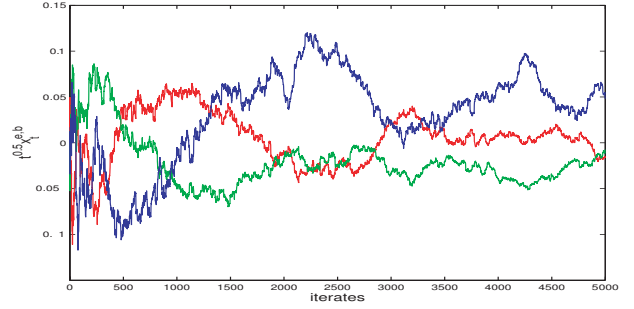
Fig. 4.   The scaled error component $\sqrt{t}x_t^{e,b}$.

Denote $S_{tk} = E\xi_{tk}\xi_{tk}^T$, $R_{tk} = E(\xi_{tk}\xi_{tk}^T|\xi_{t1}, \cdots, \xi_{t(k-1)})$, $S_t = \sum_{k=1}^t S_{tk}$, and $\zeta_t = \sum_{k=1}^t \xi_{tk}$. Assume

$$\sup_{t \geq 1} \sum_{k=1}^t E|\xi_{tk}|^2 < \infty, \quad (A.1)$$

$$\lim_{t \to \infty} \sum_{k=1}^t E|S_{tk} - R_{tk}| = 0, \quad (A.2)$$

$$\lim_{t \to \infty} S_t = S, \quad (A.3)$$

$$\lim_{t \to \infty} \sum_{k=1}^t E|\xi_{tk}|^2 1_{(|\xi_{tk}| > \varepsilon)}, \quad \forall \varepsilon > 0. \quad (A.4)$$

Then $\zeta_t \xrightarrow{d} N(0, S)$. □

We give two lemmas without proof for reasons of space.

Lemma 10: Let $\delta \in (0, 1)$ and suppose the sequence of nonnegative numbers $b_i, i \geq 1$ satisfies $s_t = \sum_{i=1}^t b_i = o(t)$. Then $\lim_{t \to \infty} t^{-\delta} \sum_{i=1}^t i^{-(1-\delta)} b_i = 0$. □

Lemma 11: Suppose the sequence of nonnegative numbers $b_i, i \geq 1$ satisfies $\lim_{T \to \infty} (1/T) \sum_{i=k}^{k+T-1} b_i = 0$ uniformly w.r.t. k. Then $\lim_{k \to \infty} k \sum_{j=k}^{\infty} j^{-2} b_i = 0$. □

APPENDIX B: PROOF OF ASYMPTOTIC NORMALITY

We shall use $C > 0$ to denote a generic constant which may vary from place to place.

Proof of Lemma 6: We have the recursion

$$\tilde{z}_{t+1}/\sqrt{a_{t+1}} = (\sqrt{a_t}/\sqrt{a_{t+1}})(I + a_t\tilde{B})(\tilde{z}_t/\sqrt{a_t}) + (a_t/\sqrt{a_{t+1}})\tilde{v}_t. \quad (B.1)$$

It is straightforward to show [5]

$$(\sqrt{a_t}/\sqrt{a_{t+1}})(I + a_t\tilde{B}) = I + a_t(\tilde{B} + \alpha I/2) + o(a_t),$$
$$\triangleq I + a_t\hat{B}_t. \quad (B.2)$$

Denote $\Pi_{l,i} = \prod_{k=i+1}^l (I + a_k\hat{B}_k)$, $l \geq i$, where $(I + a_{i+1}\hat{B}_{i+1})$ appears as the most right term in the successive matrix product. Denote $\Pi_{i,i} \triangleq I$. By elementary estimates (see, e.g. [5]), it can be shown that $\left|\Pi_{l,i} - \exp\left\{(\sum_{k=i+1}^l a_k)(\tilde{B} + \alpha I/2)\right\}\right| \to 0$, uniformly with respect to l when $i \to \infty$, and $|\Pi_{l,i}| \leq C_\Pi \exp(-\eta(\sum_{k=i+1}^l a_k))$, where C_Π and η are fixed constants.

By (B.1)-(B.2), we have $\tilde{z}_{t+1}/\sqrt{a_{t+1}} = (I + a_t\hat{B}_t)(\tilde{z}_t/\sqrt{a_t}) + (a_t/\sqrt{a_{t+1}})\tilde{v}_t$. Denote $\hat{z}_t = \tilde{z}_t/\sqrt{a_t}$. Then

$$\hat{z}_{t+1} = \prod_{i=0}^t (1 + a_i\hat{B}_i)\hat{z}_0 + \sum_{i=0}^t \Pi_{t,i}(a_i/\sqrt{a_{i+1}})\tilde{v}_i. \quad (B.3)$$

Since $\lim_{t\to\infty}\prod_{i=0}^{t}(1+a_i\hat{B}_i)=0$, it suffices to show asymptotic normality of $\sum_{i=0}^{t}\Pi_{t,i}(a_i/\sqrt{a_{i+1}})\tilde{v}_i$. Define $\xi_{tk}=\Pi_{t,k}(a_k/\sqrt{a_{k+1}})\tilde{v}_k$. We verify conditions (A.2)-(A.3) in Lemma 9 with $0\le k\le t$; then (A.1) follows from (A.3).

Denote the σ-algebra $\tilde{\mathscr{F}}_i=\mathscr{F}(\tilde{v}_0,\cdots,\tilde{v}_i)$. In fact,

$$\Delta_t\triangleq\sum_{k=0}^{t}E|S_{tk}-R_{tk}|\le C\sum_{i=0}^{t}a_i|\Pi_{t,i}|^2E\left|E[\tilde{v}_i\tilde{v}_i^T|\tilde{\mathscr{F}}_{i-1}]-E\tilde{v}_i\tilde{v}_i^T\right|.$$

Since there exist $0<\kappa\le\beta$ such that $\kappa(i+1)^{-1}\le a_i\le\beta(i+1)^{-1}$, $i\ge 0$, we have $a_i|\Pi_{t,i}|^2\le C(i+1)^{-1}\exp(-2\eta\kappa\sum_{k=i+1}^{t}k^{-1})\le C(i+1)^{-1+\delta}(t+1)^{-\delta}$, where $\delta=2\eta\kappa$; we take a small κ so that $\delta<1$. Then the relation

$$\Delta_t\le C(t+1)^{-\delta}\sum_{k=0}^{t}(k+1)^{-1+\delta}E|E[\tilde{v}_k\tilde{v}_k^T|\tilde{\mathscr{F}}_{k-1}]-E\tilde{v}_k\tilde{v}_k^T|$$

combined with Lemma 10 gives $\lim_{t\to\infty}\sum_{k=0}^{t}E|S_{tk}-R_{tk}|=0$.

Below we simply write $Q_i=Q_i^{\tilde{v}}$. Next, we verify (A.3) in Lemma 9. We have

$$S_t=\sum_{k=0}^{t}S_{tk}=\sum_{i=0}^{t}a_i^2a_{i+1}^{-1}\Pi_{t,i}Q_i\Pi_{t,i}^T$$
$$=\sum_{i=0}^{t}a_i\Pi_{t,i}Q_i\Pi_{t,i}^T+\sum_{i=0}^{t}(\alpha a_i^2+o(a_i^2))\Pi_{t,i}Q_i\Pi_{t,i}^T$$
$$\triangleq S_t^{(1)}+S_t^{(2)}.\tag{B.4}$$

Since Q_i is a bounded sequence implied by (14), we have

$$|S_t^{(2)}|\le C\sum_{i=0}^{T_1}a_i^2|\Pi_{t,i}|^2+C\sum_{i=T_1+1}^{t}a_i^2|\Pi_{t,i}|^2\triangleq S_t^{(2,a)}+S_2^{(2,b)}.$$

Since $\Pi_{t,i}$ is bounded, for any given $\varepsilon>0$, we can take a sufficiently large T_1 such that $\sup_{t\ge T_1}S_t^{(2,b)}\le\varepsilon$. On the other hand, $\lim_{t\to\infty}\Pi_{t,i}=0$. Hence, $\lim_{t\to\infty}S_t^{(2)}=0$.

We continue to determine the limit for $S_t^{(1)}$. By extending the treatment in [5] (pp. 125) to a sequence of time-varying covariances, we may use (B.4) to show that

$$\lim_{t\to\infty}|S_t^{(1)}-\sum_{i=0}^{t}a_i\exp(\hat{B}\sum_{k=i+1}^{t}a_k)Q_i\exp(\hat{B}^T\sum_{k=i+1}^{t}a_k)|=0,$$

where $\hat{B}=\tilde{B}+\alpha I/2$. Subsequently, we need to show that the second term in the difference has a limit as $t\to\infty$.

Fix any $\varepsilon>0$. First, by **(A4)** we may take a large T_0 such that $\sup_{k\ge 0}|(1/T_0)\sum_{i=k}^{k+T_0}Q_i-\overline{Q}^{\tilde{v}}|\le\varepsilon$. For the proof below, once T_0 is selected, it suffices to consider $t=kT_0$. Denote $M_{t,i}=\exp(\hat{B}\sum_{k=i+1}^{t}a_k)$. The estimates below appear a bit technical. However, the basic idea is relatively simple. Intuitively, when j is large, on the time window $[jT_0+1,(j+1)T_0]$, since a_i varies slowly, the pair $(a_i,M_{t,i})$ appearing in $a_iM_{t,i}Q_iM_{t,i}^T$ may be frozen as its value at the starting time jT_0+1, incurring only a small error. Let K_0 be fixed, we may verify that for all $K_0T_0<i<j\le i+T_0$, $|a_iM_{t,i}M_{t,i}^T-a_jM_{t,j}M_{t,j}^T|\le CT_0a_i^2$ where C does not depend on (T_0,K_0).

For $k\ge K_0+1$, we have

$$\sum_{i=0}^{kT_0}a_iM_{t,i}Q_iM_{t,i}^T=\sum_{i=0}^{K_0T_0}a_iM_{t,i}Q_iM_{t,i}^T+\sum_{j=K_0}^{k-1}\sum_{i=jT_0+1}^{(j+1)T_0}a_iM_{t,i}Q_iM_{t,i}^T$$

where $t=kT_0$. Then we set $\tilde{S}_t=\sum_{j=K_0}^{k-1}\sum_{i=jT_0+1}^{(j+1)T_0}a_iM_{t,i}Q_iM_{t,i}^T$, and it is easy to show that

$$\left|\tilde{S}_t-\sum_{j=K_0}^{k-1}\sum_{i=jT_0+1}^{(j+1)T_0}a_{jT_0+1}M_{t,jT_0+1}Q_iM_{t,jT_0+1}^T\right|\le CT_0\sum_{i=K_0T_0+1}^{kT_0}a_i^2.$$

Hence

$$\left|\tilde{S}_t-\left\{\sum_{j=K_0}^{k-1}\sum_{i=jT_0+1}^{(j+1)T_0}a_{jT_0+1}M_{t,jT_0+1}\overline{Q}^{\tilde{v}}M_{t,jT_0+1}^T\right|\right.\tag{B.5}$$
$$\le CT_0\sum_{i=K_0T_0+1}^{kT_0}a_i^2+C\varepsilon,$$

where C does not depend on T_0 and we obtain the term $C\varepsilon$ by using (14) and the fact $\sup_{k\ge 0}\sum_{i=0}^{k}a_i|M_{k,i}M_{k,i}^T|<\infty$.

By switching the index jT_0+1 back to i in (B.5) and taking into account the error incurred, we obtain the estimate

$$\left|\tilde{S}_t-\sum_{j=K_0}^{k-1}\sum_{i=jT_0+1}^{(j+1)T_0}a_iM_{t,i}\overline{Q}^{\tilde{v}}M_{t,i}^T\right|\le CT_0\sum_{i=K_0T_0+1}^{kT_0}a_i^2+C\varepsilon$$

where $t=kT_0$ and C does not depend on (T_0,K_0). This gives

$$|\tilde{S}_t-\sum_{i=0}^{kT_0}a_iM_{t,i}\overline{Q}^{\tilde{v}}M_{t,i}^T|\le CT_0\sum_{i=K_0T_0}^{kT_0}a_i^2+C\varepsilon+o(1),\tag{B.6}$$

where $t=kT_0$ and we have used $\sum_{i=0}^{K_0T_0}a_i|M_{t,i}\overline{Q}^{\tilde{v}}M_{t,i}^T|=o(1)$. Now, after T_0 is selected, we may fix a large K_0 so that $CT_0\sum_{i=K_0T_0}^{kT_0}a_i^2\le\varepsilon$. Subsequently, we pick up T_0' sufficiently large such that for all $t\ge T_0'$, we have $\sum_{i=0}^{K_0T_0}a_i|M_{t,i}\overline{Q}^{\tilde{v}}M_{t,i}^T|\le\varepsilon$. Hence the right hand side of (B.6) is upper bounded by $(C+2)\varepsilon$ for all $t\ge T_0'$. Note that for any given T_0,K_0, $|S_t-\tilde{S}_t|\to 0$ as $t=kT_0\to\infty$, and therefore, it follows that

$$\lim_{t\to\infty}|S_t-\sum_{i=0}^{t}a_i\exp(\hat{B}\sum_{k=i+1}^{t}a_k)\overline{Q}^{\tilde{v}}\exp(\hat{B}^T\sum_{k=i+1}^{t}a_k)|=0,\tag{B.7}$$

where the second term in the difference has a limit with a standard integral representation $\int_0^{\infty}e^{\hat{B}t}\overline{Q}^{\tilde{v}}e^{\hat{B}^Tt}dt$ (see [16], [5]). Finally, we may verify (A.4) by elementary estimates. $\square$

Proof of Lemma 7: We write $z_k^1-z_\infty^1=-\sum_{i=k+1}^{\infty}a_iv_i^1$. Let $\delta_k=\sqrt{k}\sum_{i=k+k^2+1}^{\infty}a_iv_i^1$. Then

$$E|\delta_k|^2=O(1/k)\tag{B.8}$$

as $k\to\infty$. It suffices to show that $\sqrt{k}\sum_{i=k}^{k+k^2}a_iv_i^1$ converges in distribution to a normal random variable.

Denote $\Sigma_{k,j}=\sqrt{k}\sum_{t=k}^{k+j}a_tv_t^1$ and $s_{k,j}=k\sum_{i=k}^{k+j}a_i^2\sigma_i^2$ for $j\ge 0$. Let **i** be the imaginary unit, and for all the estimates below, t is interpreted as a real number. We can show that

$$\Delta_{k,N}\triangleq|E\exp\{\mathbf{i}t\Sigma_{k,N}\}-\exp(-s_{k,N}t^2/2)|$$
$$\le\exp(-s_{k,N}t^2/2)\sum_{j=0}^{N}\left|E\exp\{\mathbf{i}t\Sigma_{k,j}+s_{k,j}t^2/2\}\right.$$
$$\left.-E\exp\{\mathbf{i}t\Sigma_{k,j-1}+s_{k,j-1}t^2/2\}\right|$$

for any $N > k$. Now we have

$$\left| E\exp\{\mathbf{i}t\Sigma_{k,j} + s_{k,j}t^2/2\} - E\exp\{\mathbf{i}t\Sigma_{k,j-1} + s_{k,j-1}t^2/2\}\right|$$
$$\leq \exp\{s_{k,j}t^2/2\}|E\exp\{\mathbf{i}ta_jv_j^1\sqrt{k}\} - \exp\{-t^2ka_j^2\sigma_j^2/2\}|$$
$$\triangleq \exp\{s_{k,j}t^2/2\}D_k.$$

For random variable ξ, denote $H(t,\xi) = e^{\mathbf{i}t\xi} - 1 - \mathbf{i}t\xi + t^2\xi^2/2$. Let $H_k^{(1)} = H(ta_j\sqrt{k}, v_j^1)$, $H_k^{(2)} = \exp\{-t^2ka_j^2\sigma_j^2\} - 1 + t^2ka_j^2\sigma_j^2/2$. Denote σ-algebra $\mathscr{F}_i^1 = \mathscr{F}(v_0^1, \cdots, v_i^1)$. Then

$$D_k = |EH_k^{(1)} - H_k^{(2)} - (t^2ka_j^2/2)E(E[|v_j^1|^2|\mathscr{F}_{j-1}^1] - \sigma_j^2)|$$
$$\leq E|H_k^{(1)}| + |H_k^{(2)}| + (t^2ka_j^2/2)E\left|E[|v_j^1|^2|\mathscr{F}_{j-1}^1] - \sigma_j^2\right|.$$

By the elementary inequality $|H(t,\xi)| \leq \min\{t^2\xi^2, |t\xi|^3/6\}$ (see [6], pp. 292), we have

$$|H_k^{(1)}| \leq t^2a_j^2k|v_j^1|^21_{(|v_j^1|>\varepsilon k^{1/3})} + t^3a_j^3|v_j^1|^3k^{3/2}1_{(|v_j^1|\leq\varepsilon k^{1/3})}$$

for any given $\varepsilon > 0$, and furthermore, $|H_k^{(2)}| \leq t^4k^2a_j^4\sigma_j^4$. Let $c_s = \sup_{k,j\geq k}(s_{k,j}/2) < \infty$. We have

$$\Delta_{k,N} \leq e^{c_st^2}\left\{ \sum_{j=k}^{k+N} t^2a_j^2kE|v_j^1|^21_{(|v_j^1|>\varepsilon k^{1/3})} \; (\triangleq \Delta_{k,N}^{(1)}) \right.$$
$$+ \sum_{j=k}^{k+N} t^3a_j^3k^{3/2}E|v_j^1|^31_{(|v_j^1|\leq\varepsilon k^{1/3})} \; (\triangleq \Delta_{k,N}^{(2)})$$
$$+ \sum_{j=k}^{k+N} t^4k^2a_j^4\sigma_j^4 \; (\triangleq \Delta_{k,N}^{(3)})$$
$$\left. + \sum_{j=k}^{k+N} (t^2ka_j^2/2)E\left|E[v_j^1|^2|\mathscr{F}_{j-1}^1] - \sigma_j^2\right| \; (\triangleq \Delta_{k,N}^{(4)})\right\}.$$

We have

$$\Delta_{k,N}^{(1)} \leq \sup_j E|v_j^1|^21_{(|v_j^1|>\varepsilon k^{1/3})}k\sum_{j=k}^{\infty} a_j^2 \leq C\sup_j E|v_j^1|^21_{(|v_j^1|>\varepsilon k^{1/3})},$$

where $C = \sup_{k\geq 1} k\sum_{j=k}^{\infty} a_j^2 < \infty$, and

$$\Delta_{k,N}^{(2)} \leq t^3k^{3/2+1/3}\varepsilon \sum_{j=k}^{\infty} a_j^3E|v_j^1|^21_{(|v_j^1|\leq\varepsilon k^{1/3})}$$
$$\leq \varepsilon t^3k^{3/2+1/3}(\sup_j E|v_j^1|^2)Ck^{-2} = \varepsilon Ct^3k^{-1/6}(\sup_j E|v_j^1|^2)$$

for all $k \geq 1$, by the fact that for all $k \geq 1$, $\sum_{j=k}^{\infty} a_j^3 \leq Ck^{-2}$. Similarly, $\Delta_{k,N}^{(3)} \leq Ct^4k^{-1}$ for all $k \geq 1$. Next,

$$\Delta_{k,N}^{(4)} \leq Ct^2k\sum_{j=k}^{\infty} j^{-2}E\left|E[|v_j^1|^2|\mathscr{F}_{j-1}^1] - \sigma_j^2\right|$$

where $\lim_{k\to\infty} k\sum_{j=k}^{\infty} j^{-2}E|E[|v_j^1|^2|\mathscr{F}_{j-1}^1] - \sigma_j^2| = 0$ by (18) and Lemma 11.

Now, we take $N = k^2$, and it follows that given any compact interval $[T_1, T_2]$, we have

$$\lim_{k\to\infty} \sup_{t\in[T_1,T_2]} |E\exp\{\mathbf{i}t\Sigma_{k,k^2}\} - \exp(-s_{k,k^2}t^2/2)| = 0.$$

On the other hand, we have $\lim_{k\to\infty} s_{k,k^2} = \alpha^{-2}\bar\sigma^2$. Hence

$$\lim_{k\to\infty} \sup_{t\in[T_1,T_2]} |E\exp\{\mathbf{i}t\Sigma_{k,k^2}\} - \exp(-\alpha^{-2}\bar\sigma^2t^2/2)| = 0.$$

Recalling (B.8), this completes the proof. □

REFERENCES

[1] P. Barooah and J. P. Hespanha. Estimation on graphs from relative measurements: distributed algorithms and fundamental limits. *IEEE Contr. Syst. Mag.*, vol. 27, pp. 57-74, Aug. 2007.

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

[3] R. Carli, F. Fagnani, A. Speranzon, and S. Zampieri. Communication constraints in coordinated consensus problems. *Proc. Amer. Contr. Conf.*, Minneapolis, MN, pp. 4189-4194, June 2006.

[4] R. Carli, F. Fagnani, P. Frasca, T. Taylor, and S. Zampieri. Average consensus on networks with transmission noise or quantization. *Proc. European Control Conference*, Kos, Greece, pp. 1852-1857, July 2007.

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

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

[7] P. Hall and C. C. Heyde. *Martingale Limit Theory and Its Application*. Academic Press, New York, 1980.

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

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

[10] 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.

[11] M. Huang and J. H. Manton. Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior. To appear in *SIAM J. Contr. Optim.*, submitted, Oct. 2006.

[12] M. Huang and J. H. Manton. Stochastic consensus seeking with measurement noise: convergence and asymptotic normality. *Proc. American Control Conference*, Seattle, WA, pp. 1337-1342, June 2008.

[13] M. Huang and J. H. Manton. Stochastic consensus seeking with noisy and directed inter-agent communication. Submitted to *IEEE Trans. Autom. Control* in June 2007, revised Dec. 2007, Aug. 2008.

[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, pp. 988-1000, June 2003.

[15] H. J. Kushner and G. G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, New York, 1997.

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

[17] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Trans. Autom. Control*, vol. 49, 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, Jan. 2007.

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

[20] 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.

[21] 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.

[22] 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.

[23] J. Sacks. Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statis.*, vol. 29, no. 2, pp. 373-405, 1958.

[24] I. D. Schizas, A. Ribeiro, and G. B. Giannakis. Consensus in ad hoc WSNs with noisy links–part I: distributed estimation of deterministic signals. *IEEE Trans. Signal Process.*, vol. 56, pp. 350-364, Jan. 2008.

[25] 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.

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

[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.