Stochastic Approximation for Consensus with General Time-Varying Weight Matrices

Minyi Huang

Abstract— **This paper considers consensus problems with delayed noisy measurements, and stochastic approximation is used to achieve mean square consensus. For stochastic approximation based consensus algorithms with switching topologies, the existing convergence analysis heavily relies on quadratic Lyapunov functions, whose existence may be difficult to guarantee for switching digraphs. The main contribution of this paper is to introduce a new approach for proving convergence. This is achieved by obtaining ergodicity results for backward products of degenerating stochastic matrices via a discrete time dynamical system approach. Our approach does not require the double stochasticity condition typically assumed for the existence of a quadratic Lyapunov function.**

I. Introduction

In the past decade consensus problems and various closely related formulations have been intensively investigated for multi-agent systems [13], [19], [22]. A comprehensive survey can be found in [18], [24]. In the recent years, consensus algorithms with imperfect information exchange have attracted increasing attention, addressing measurement noise or quantization effect [2], [23], [30], [3], [25], [1], [4], [15], [23], [25], [30]. The work [28] made early effort introducing stochastic gradient based consensus algorithms.

In consensus models with noisy measurements, stochastic approximation with decreasing step sizes may be applied such that each agent can extract state information from its neighbors while reducing the detrimental noise effect [11], [12], [14], [16], [21], [27]. A popular tool for proving convergence is to use quadratic Lyapunov functions. For fixed network topologies, the existence of such functions is guaranteed. When the network topologies experience switches, the quadratic Lyapunov function approach is still applicable in balanced graph models which give doubly stochastic matrices in averaging [19]. In undirected graphs, doubly stochastic weight matrices may be easily constructed by using the well known Metropolis weights [30]. Distributed iterative algorithms have been developed for constructing doubly stochastic matrices over digraphs [8].

However, in randomly varying digraph models it is difficult to construct doubly stochastic matrices for averaging unless global information is available about the instantaneous network topology. Iterative algorithms as those in [8] are not applicable since the network condition may have changed before the iterates can converge. It is of practical importance to consider models without the double stochasticity property.

This work was partially supported by NSERC of Canada. IEEE CDC'10.

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

In this paper, we use general time-varying weight matrices for stochastic approximation in noisy models and also deal with random delays in the reception of signals. This problem formulation presents new challenges in analysis. Firstly, the traditional methodologies in stochastic approximation are not applicable since the coefficient matrix has no asymptotic stability and changes too rapidly. Secondly, the Lyapunov approach is hardly applicable. When the coefficient matrix at each step is doubly stochastic, one may construct a quadratic Lyapunov function via the so called disagreement function [19]. From the point of view of switched systems, the disagreement function essentially defines multiple Lyapunov functions decreasing along the trajectory of the iteration. Now for general coefficient matrices, the use of quadratic Lyapunov functions is no longer feasible. For general nonexistence results on quadratic Lyapunov functions, see [20].

Our convergence analysis relies on the backward products of degenerating stochastic matrices forming a sequence of in-homogeneous stochastic matrices. For analyzing the products of in-homogeneous stochastic matrices from a finite collection, Wolfowitz's ergodicity theorem [29], [13] and paracontraction [5], [6], [7] are well known powerful techniques. However, they cannot be applied to our stochastic approximation model due to the degenerating nature of the matrices. In this paper, we will develop a dynamical system approach to examine the backward products of degenerating stochastic matrices and establish their ergodicity, which further ensures the convergence of the stochastic approximation algorithm for consensus. Our approach can deal with both synchronous and asynchronous algorithms.

Due to limited space, this paper only describes the main steps to prove the ergodicity and consensus theorems, and more detailed analysis is available in [9].

The paper is organized as follows. Section II formulates the stochastic consensus problem. Section III shows a necessary and sufficient condition for mean square consensus via ergodic backward products. Section IV introduces the notion of compatible matrices and Section V proves ergodicity of degenerating stochastic matrices. Section VI shows mean square consensus. Section VII presents simulation results and Section VIII concludes the paper.

II. The Stochastic Consensus Algorithm

We introduce some standard preliminaries on graph modeling of the network topology. A directed graph (digraph) $G=(\mathscr{N},\mathscr{E})$ consists of a set of nodes $\mathscr{N}=\{1,\dots,n\}$ and a set of directed edges $\mathscr{E}$. A directed edge (simply called an edge) is denoted by an ordered pair $(i,j)\in\mathscr{N}\times\mathscr{N}$, where

$i \neq j$. A directed path (from node i_1 to node i_l) consists of a sequence of nodes $i_1, \ldots, i_l$, $l \geq 2$, such that $(i_k, i_{k+1}) \in \mathscr{E}$. The digraph G is strongly connected if from any node to any other node, there exists a directed path. A directed tree is a digraph where each node i, except the root, has exactly one parent node j so that $(j,i) \in \mathscr{E}$. We call $G' = (\mathscr{N}', \mathscr{E}')$ a subgraph of G if $\mathscr{N}' \subset \mathscr{N}$ and $\mathscr{E}' \subset \mathscr{E}$. The digraph G is said to contain a spanning tree if there exists a directed tree $G_{\mathrm{tr}} = (\mathscr{N}, \mathscr{E}_{\mathrm{tr}})$ as a subgraph of G. The adjacency matrix of G is an $n \times n$ matrix $A_G = (a_{ij})_{1 \leq i,j \leq n}$, where $a_{ij} = 1$ if $(i,j) \in \mathscr{E}$, and $a_{ij} = 0$ otherwise. If G is an undirected graph, each edge is denoted as an unordered pair (i,j), where $i \neq j$.

The dynamic network topology to specify the signal reception is modeled by a sequence of digraphs $\{G_t = (\mathscr{N}, \mathscr{E}_t), t \geq 0\}$, where $\mathscr{N} = \{1, \ldots, n\}$ and $\mathscr{E}_t$ randomly changes with time. The adjacency matrix A_{G_t} is a matrix-valued random variable and completely determines $\mathscr{E}_t$. If $(j,i) \in \mathscr{E}_t$, node i receives information from node j which is called a neighbor of node i. The neighbor set of node i is $\mathscr{N}_{i,t} = \{j | (j,i) \in \mathscr{E}_t\}$.

We make some convention about notation. The node index is often used as a superscript but not an exponent in different variables (x_t^i, z_t^i, etc.). For a matrix M, the element at the ith row and the jth column is called the (i,j)-th element and denoted by $M(i,j)$. For a vector or matrix M, denote the Frobenius norm $|M| = [\mathrm{Tr}(M^T M)]^{1/2}$. We use $\mathbf{1}_k \in \mathbb{R}^k$ to denote a column vector of k ones. For column vectors $Z_1, \ldots, Z_l$, $[Z_1; \ldots; Z_l]$ denotes the column vector obtained by vertical concatenation of the l vectors. We use C to denote a generic positive constant which may vary at different places.

A. The Stochastic Approximation Algorithm

The underlying probability space is denoted by $(\Omega, \mathscr{F}, P)$, corresponding to the sample space, the collection of all events, and the probability measure, respectively. At time $t \in \{0, 1, 2, \ldots\}$, node i is associated with a real-valued state x_t^i. Define the state vector

$$X_t = [x_t^1, \ldots, x_t^n]^T, \quad t \geq 0.$$

At time t, if $\mathscr{N}_{i,t} \neq \emptyset$ (the empty set), node i receives possibly outdated information from its neighbors modeled by

$$y_t^{ik} = x_{t-d_t^{ik}}^k + w_t^{ik}, \quad k \in \mathscr{N}_{i,t}, \tag{1}$$

where w_t^{ik} is the noise and $d_t^{ik} \geq 0$ is an integer-valued random delay. Since the system starts at $t = 0$, the implicit requirement for the neighbor set is that

$$k \in \mathscr{N}_{i,t} \text{ implies } t - d_t^{ik} \geq 0. \tag{2}$$

Each node uses its own state and its noisy measurements to form a weighted average. Define the matrix $B_t = (b_{ik}(t))_{1 \leq i,k \leq n}$ as follows. If $\mathscr{N}_{i,t} = \emptyset$, define

$$b_{ik}(t) = 0 \quad \text{for all } k \in \mathscr{N}. \tag{3}$$

If $\mathscr{N}_{i,t} \neq \emptyset$, define

$$\begin{cases} b_{ik}(t) \in [\underline{b},\ \overline{b}], & \text{if } k \in \mathscr{N}_{i,t} \\ b_{ik}(t) = 0, & \text{if } k \notin \mathscr{N}_{i,t} \cup \{i\} \\ b_{ii}(t) = -\sum_{k \in \mathscr{N}_{i,t}} b_{ik}(t), & \end{cases} \tag{4}$$

where $0 < \underline{b} \leq \overline{b} < \infty$ are two deterministic constants. Since the sequence $\{G_t, t \geq 0\}$ changes randomly, $\{B_t, t \geq 0\}$ is a matrix-valued random process.

At time $t \geq 0$, if $\mathscr{N}_{i,t} = \emptyset$, set $x_{t+1}^i \equiv x_t^i$. If $\mathscr{N}_{i,t} \neq \emptyset$, node i updates its state by the rule

$$x_{t+1}^i = [1 + a_t b_{ii}(t)] x_t^i + a_t \sum_{k \in \mathscr{N}_{i,t}} b_{ik}(t) y_t^{ik}, \quad t \geq 0, \tag{5}$$

where $\{a_t, t \geq 0\}$ is a sequence of positive step sizes. We call $I + a_t B_t$ the weight matrix. Since all nodes update their state simultaneously, (5) is a synchronous algorithm.

Denote the maximal set of communication links $\mathscr{E}_{\max} = \{(k,i) | \sup_{t \geq 0} P((k,i) \in \mathscr{E}_t) > 0\}$. For convenience of statistical modeling, we make the convention: w_t^{ik} and d_t^{ik} are defined for all $(k,i) \in \mathscr{E}_{\max}$. If (k,i) does not appear in $\mathscr{E}_t$ so that (1) does not physically occur, we still introduce w_t^{ik} and d_t^{ik} as dummy random variables. If $(k,i) \notin \mathscr{E}_t$, we set $d_t^{ik} = 0$. Let $\{w_t^{ik} | (k,i) \in \mathscr{E}_{\max}\}$ be listed by a fixed ordering of (k,i) to obtain a noise vector W_t of n_1 dimension.

Definition 1: The n nodes are said to achieve mean square consensus if $E|x_t^i|^2 < \infty$, $t \geq 0$, $1 \leq i \leq n$, and there exists a random variable x^* such that $\lim_{t \to \infty} E|x_t^i - x^*|^2 = 0$ for $1 \leq i \leq n$. ♢

B. Main Assumptions

(**A1**) The deterministic sequence $\{a_t, t \geq 0\}$ satisfies

$$a_0 > 0, \quad \alpha t^{-\gamma} \leq a_t \leq \beta t^{-\gamma}, \qquad t \geq 1, \tag{6}$$

where $\gamma \in (1/2,\ 1]$ and $0 < \alpha \leq \beta < \infty$. ♢

So (**A1**) implies $\sum_{t=0}^{\infty} a_t = \infty$ and $\sum_{t=0}^{\infty} a_t^2 < \infty$.

For $0 \leq t_1 < t_2$, define the digraph $G_{[t_1,t_2)} = (\mathscr{N}, \cup_{t_1 \leq t < t_2} \mathscr{E}_t)$, which is called the union of the digraphs $\{G_t | t_1 \leq t < t_2\}$. Clearly, $G_{[t_1,t_2)}$ depends on the sample ω.

(**A2**) There exist integer-valued random variables $0 \equiv T_0 < T_1 < \ldots < T_{l-1} < T_l < \ldots$ such that after excluding a null set $N_0 \subset \Omega$ (i.e., $P(N_0) = 0$), the two conditions hold:

(i) $G_{[T_l, T_{l+1})}$ is strongly connected for $l \geq 0$ and $\omega \in \Omega \backslash N_0$.

(ii) $\alpha_2 := \sup_{l \geq 1}(T_l - T_{l-1}) < \infty$, for $\omega \in \Omega \backslash N_0$. ♢

(**A3**) $\{W_t, t \geq 0\}$ is a sequence of independent random vectors of zero mean and is independent of $\{(B_t, A_{G_t}, \{d_t^{ik} | (k,i) \in \mathscr{E}_{\max}\}), t \geq 0\}$, where $0 \leq d_t^{ik} \leq d^*$ for a fixed integer $d^* \geq 0$. In addition, $E|X_0|^2 < \infty$ and $\sup_{t \geq 0} E|W_t|^2 < \infty$. ♢

For leader following we introduce another connectivity condition.

(**A2′**) There is a fixed leader node i_L which has no neighbor in each G_t. There exist integer-valued random variables $0 \equiv T_0 < T_1 < \ldots < T_l < \ldots$ such that after excluding a null set $N_0 \subset \Omega$, $G_{[T_l, T_{l+1})}$ contains a spanning tree with root i_L for $l \geq 0$ and $\omega \in \Omega \backslash N_0$. In addition, (**A2**)-(ii) is satisfied. ♢

C. A Vector Form of the Algorithm

Denote the set of $n \times n$ random matrices

$$B_{d,t} = (B_{d,t}(i,k))_{1 \leq i,k \leq n}, \quad d = 0, 1, \ldots, d^*.$$

For their diagonal elements, we take $B_{0,t}(i,i) = b_{ii}(t)$ and $B_{d,t}(i,i) = 0$, $d = 1, \ldots, d^*$, for all i. For $d = 0, 1, \ldots, d^*$, the off-diagonal element $B_{d,t}(i,k)$ is nonzero and further taken as $b_{ik}(t)$ if and only if $b_{ik}(t) > 0$ and $d_t^{ik} = d$. Denote

$$\mathbf{B}_t^0 = [B_{0,t}, B_{1,t}, \ldots, B_{d^*,t}]. \quad (7)$$

The ith row of $\mathbf{B}_t^0$ contains the same set of nonzero elements as the ith row of B_t does. Due to (2), if $t < d^*$, we necessarily have $B_{d,t} = 0$ for all $d = t+1, \ldots, d^*$.

We write (5) in the equivalent form

$$X_{t+1} = X_t + a_t \mathbf{B}_t^0 [X_t; X_{t-1}; \ldots; X_{t-d^*}] + a_t D_t W_t, \quad t \geq 0, \quad (8)$$

where D_t is an $n \times n_1$ random coefficient matrix determined by B_t and we set $X_t \equiv 0$ for $-d^* \leq t < 0$. If $d^* = 0$, (8) reduces to

$$X_{t+1} = (I + a_t B_t) X_t + a_t D_t W_t, \quad t \geq 0.$$

Denote the $n(d^*+1) \times n(d^*+1)$ matrix

$$\mathbf{U} = \begin{bmatrix} I & 0 & 0 & \cdots & 0 \\ I & 0 & 0 & \cdots & 0 \\ 0 & I & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & I & 0 \end{bmatrix}, \quad (9)$$

where each identity matrix is $n \times n$, and denote

$$\mathbf{B}_t = \begin{bmatrix} \mathbf{B_t^0} \\ 0_{nd^* \times n(d^*+1)} \end{bmatrix}, \quad \mathbf{D}_t = \begin{bmatrix} D_t \\ 0_{nd^* \times n_1} \end{bmatrix}. \quad (10)$$

It is clear that $\mathbf{B}_t$ is determined by $(B_t, \{d_t^{ik} | (k,i) \in \mathscr{E}_{\max}\})$.

Denote $\mathbf{X}_t = [X_t; X_{t-1}; \ldots; X_{t-d^*}] \in \mathbb{R}^{n(d^*+1)}$. We have

$$\mathbf{X}_{t+1} = (\mathbf{U} + a_t \mathbf{B}_t) \mathbf{X}_t + a_t \mathbf{D}_t W_t, \quad t \geq 0, \quad (11)$$

where $X_t \equiv 0$ for $-d^* \leq t < 0$.

If $a_t \to 0$ as $t \to \infty$, for some $t_0(\omega)$ depending on $\omega \in \Omega$, $\{\mathbf{U} + a_t \mathbf{B}_t(\omega), t \geq t_0(\omega)\}$ is a sequence of stochastic matrices converging to the 0-1 stochastic matrix $\mathbf{U}$ and will be called a sequence of degenerating stochastic matrices.

III. A Necessary and Sufficient Condition for Consensus

We use a general algorithm to reveal a fundamental relationship between mean square consensus and ergodicity of backward matrix products. Consider the system

$$Y_{t+1} = A_t Y_t + H_t V_t, \quad t \geq 0, \quad (12)$$

where $Y_t \in \mathbb{R}^{m_1}$ denotes the states of m_1 agents, $V_t \in \mathbb{R}^{m_2}$ is the noise vector, and the initial condition is Y_0. Here $\{A_t, t \geq 0\}$ and $\{H_t, t \geq 0\}$ are two sequences of random matrices of compatible dimensions. For each fixed ω, $A_t(\omega)$ is a stochastic matrix for all $t \geq 0$. The model (12) includes (11) as a special case if the coefficient matrices of $\mathbf{X}_t$ in (11) are nonnegative for all $t \geq 0$.

A. Ergodicity of Backward Products

Let $\{\tilde{A}_t, t \geq 0\}$ be a sequence of deterministic nonnegative matrices, where each $\tilde{A}_t$ is a stochastic matrix. Define the so-called backward product $\Phi_{t,s} = \tilde{A}_{t-1} \ldots \tilde{A}_s$ for $t \geq s \geq 0$, where $\Phi_{s,s} := I$. The product $\Phi_{t,s}$ is still a stochastic matrix. Let $\Phi_{t,s}(i,j)$ denote its (i,j)-th element.

Definition 2: [26] We say weak ergodicity holds for backward products of the sequence $\{\tilde{A}_t, t \geq 0\}$ of stochastic matrices if

$$\lim_{t \to \infty} |\Phi_{t,s}(i_1, j) - \Phi_{t,s}(i_2, j)| = 0$$

for any given $s \geq 0$ and i_1, i_2, j. If in addition to weak ergodicity, $\Phi_{t,s}(i,j)$ converges as $t \to \infty$, for any s, i, j, we say strong ergodicity holds. $\diamondsuit$

By [26, p. 154, Theorem 4.17], weak and strong ergodicity are equivalent for backward products of any sequences of stochastic matrices. Hence, in the following we only speak of ergodicity of backward products.

B. The Necessary and Sufficient Condition for Consensus

For the theorem below, we run algorithm (12) with any initial time-state pair (t_0, Y_{t_0}). Denote $Y_t = [Y_t^1, \ldots, Y_t^{m_1}]^T$ and $\Psi_{t,s} = A_{t-1} \ldots A_s$ for $t \geq s$, where $\Psi_{s,s} := I$.

Theorem 3: [9] Assume

(i) $\{V_t, t \geq 0\}$ is a sequence of random vectors of zero mean, independent of $\{(A_t, H_t), t \geq 0\}$;

(ii) $\sum_{t=0}^{\infty} E|H_t|^2 E|V_t|^2 < \infty$;

(iii) there exists a sequence of nonnegative numbers $\{\phi_k, k \geq 0\}$ such that

$$|E[V_k V_{k'}^T]| \leq \phi_{|k-k'|} [E|V_k|^2 E|V_{k'}|^2]^{1/2}, \quad \sum_{k=0}^{\infty} \phi_k < \infty.$$

Then (12) ensures mean square consensus for any initial time-state pair (t_0, Y_{t_0}) with $E|Y_{t_0}|^2 < \infty$ if and only if $\{A_t, t \geq 0\}$ has ergodic backward products with probability one. $\square$

Remark: If $\{V_t, t \geq 0\}$ are independent with $EV_t = 0$ and $E|V_t|^2 < \infty$, (iii) holds. $\diamondsuit$

IV. Compatible Nonnegative Matrices

This section develops some basic tools for analyzing sequences of degenerating stochastic matrices. To avoid introducing too many variables, the vectors X_t and $\mathbf{X_t}$ appearing in Section II will be reused in new models.

A. Compatible Matrices

Let $\{\tilde{\mathbf{A}}_t, t \geq 0\}$ be $n(d^*+1) \times n(d^*+1)$ deterministic nonnegative matrices. Consider the weighted averaging model

$$\mathbf{X}_{t+1} = \tilde{\mathbf{A}}_t \mathbf{X}_t, \qquad t \geq 0, \quad (13)$$

where $\mathbf{X}_0$ is deterministic and $\tilde{\mathbf{A}}_t$ is a stochastic matrix. Denote the partition

$$\tilde{\mathbf{A}}_t = \begin{bmatrix} \tilde{A}_{00,t} & \ldots & \tilde{A}_{0d^*,t} \\ \vdots & \ddots & \vdots \\ \tilde{A}_{d^*0,t} & \ldots & \tilde{A}_{d^*d^*,t} \end{bmatrix},$$

where each matrix $\tilde{A}_{ik,t}$ is $n\times n$. Let $\tilde{\mathbf{A}}_t(i,j)$ be the (i,j)-th element of $\tilde{\mathbf{A}}_t$.

We impose some structural restrictions on $\{\tilde{\mathbf{A}}_t, t\ge 0\}$, which are motivated by backward products of the coefficient matrices in (11). Roughly, each $\tilde{A}_{d0,t}$, $0\le d\le d^*$, is an identity matrix subject to small perturbations. So after excluding the diagonal elements of $\{\tilde{A}_{d0,t}, 0\le d\le d^*\}$, all the remaining elements of $\tilde{\mathbf{A}}_t$ are necessarily small. However, they are important in affecting the asymptotic behavior of the system (13). For further analysis, we isolate a set of relatively large elements in the first n rows of $\tilde{\mathbf{A}}_t$ excluding the diagonal elements of $\tilde{A}_{00,t}$ by associating $\tilde{\mathbf{A}}_t$ with a digraph of n nodes. We introduce the following definition for both square and non-square matrices.

Definition 4: Let $\{\tilde{\mathbf{A}}_t, t\ge 0\}$ be a sequence of $n(d^*+1)\times n(d^*+1)$ stochastic matrices, $\{d_t, t\ge 0\}$ be a sequence of nonnegative numbers converging to zero as $t\to\infty$, and $\{\tilde{G}_t=(\mathscr{N},\tilde{\mathscr{E}}_t), t\ge 0\}$ be a sequence of digraphs. If there exist constants t_c and $0<\underline{c}\le\bar{c}$ such that for all $t\ge t_c$,

i) the first n rows of $\tilde{\mathbf{A}}_t$ satisfy

$$\tilde{\mathbf{A}}_t(i,j)\le \bar{c}d_t,\quad \forall 1\le i\le n,\ 1\le j\le n(d^*+1),\ j\ne i, \tag{14}$$

$$\max_{0\le d\le d^*}\tilde{\mathbf{A}}_t(i,j+dn)\ge \underline{c}d_t,\quad \forall (j,i)\in\tilde{\mathscr{E}}_t; \tag{15}$$

ii) the (i,i)-th element of $\tilde{A}_{d0,t}$ satisfies

$$\tilde{A}_{d0,t}(i,i)\ge 1-\bar{c}d_t,\quad \forall 1\le i\le n,\ 1\le d\le d^*, \tag{16}$$

then $\{\tilde{\mathbf{A}}_t, t\ge 0\}$ is said to be (d_t)-compatible with $\{\tilde{G}_t, t\ge 0\}$. Similarly, if $\{\bar{\mathbf{A}}_t, t\ge 0\}$ is a sequence of $n\times n(d^*+1)$ nonnegative matrices with unit row sums and if (14)-(15) hold for the corresponding elements $\bar{\mathbf{A}}_t(i,j)$ and $\bar{\mathbf{A}}_t(i,j+dn)$, $\{\bar{\mathbf{A}}_t, t\ge 0\}$ is said to be (d_t)-compatible with $\{\tilde{G}_t, t\ge 0\}$. We may further define (d_t)-compatibility with any positive initial time s_0 in an obvious manner. ◇

By utilizing $\{\tilde{G}_t, t\ge 0\}$ in this manner, we may obtain useful information concerning the interaction of the components of $\mathbf{X}_t$ by the connectivity properties of the digraphs. The class of compatible nonnegative matrices will serve as the basis for analyzing general degenerating stochastic matrices arising in stochastic approximation.

If $\{\tilde{\mathbf{A}}_t, t\ge 0\}$ is (d_t)-compatible with $\{\tilde{G}_t, t\ge 0\}$, this property still holds if d_t is replaced by cd_t for some $c>0$. At some places where it is unnecessary to explicitly indicate d_t, we simply say that $\{\tilde{\mathbf{A}}_t, t\ge 0\}$ is compatible with $\{\tilde{G}_t, t\ge 0\}$.

Example: Let $\{G_t(\omega), t\ge 0\}$ be given in Section II and $\{B_t(\omega), t\ge 0\}$ be specified by (3)-(4). By using (7), define the $n\times n(d^*+1)$ matrices $M_t(\omega)=[I_n, 0_{n\times nd^*}]+a_t\mathbf{B}_t^0(\omega)$, $t\ge 0$, where $\{a_t, t\ge 0\}$ satisfies (6). Select t_0 such that $M_t(\omega)$ is a nonnegative matrix for $t\ge t_0$. It may be verified that $\{M_t(\omega), t\ge t_0\}$ is (a_t)-compatible with $\{G_t(\omega), t\ge t_0\}$. ◇

The following lemma holds since $\tilde{\mathbf{A}}_t$ has unit row sums.

Lemma 5: Assume that $\{\tilde{\mathbf{A}}_t, t\ge 0\}$ in (13) is (d_t)-compatible with $\{\tilde{G}_t, t\ge 0\}$. We have

$$\tilde{\mathbf{A}}_t(i,i)\ge 1-[n(d^*+1)-1]\bar{c}d_t,$$
$$\tilde{\mathbf{A}}_t(i+nd,j)\le \bar{c}d_t,\quad 1\le j\le n(d^*+1),\ j\ne i,$$

where $t\ge t_c$ for some constant t_c, $1\le i\le n$ and $1\le d\le d^*$. □

To distinguish from the notation in (11), we denote

$$\mathbf{X}_t=[X_t; X_t^{(-1)};\dots;X_t^{(-d^*)}]=[x_t^1,\dots,x_t^{n(d^*+1)}]^T \tag{17}$$

for (13), where X_t and $X_t^{(-d)}$, $1\le d\le d^*$, are n dimensional. If $\{\tilde{\mathbf{A}}_t, t\ge 0\}$ is (d_t)-compatible with $\{\tilde{G}_t, t\ge 0\}$, for convenience of exposition we call X_t the states of the n nodes in $\tilde{G}_t$ and may think of $X_t^{(-d)}$, $1\le d\le d^*$, as d^* copies of X_t with one step delay and different perturbations by the magnitude of d_t. We introduce the assumption for (13).

(**H1**) (i) $\{\tilde{\mathbf{A}}_t, t\ge 0\}$ is (a_t)-compatible with a sequence $\{\tilde{G}_t, t\ge 0\}$ of digraphs, where $\{a_t, t\ge 0\}$ satisfies (**A1**); (ii) each $\tilde{G}_t$, $t\ge 0$, is strongly connected. ◇

B. State Reordering and Mutual Attraction of Trajectories

A direct convergence analysis of $\mathbf{X}_t$ in (13) is quite difficult since its components undergo very complex evolution. We construct a new vector Z_t based on $X_t=[x_t^1,\dots,x_t^n]^T$. Let the n entries of X_t be listed in descending order

$$x_t^{i_1}\ge x_t^{i_2}\ge\dots\ge x_t^{i_n},$$

where $\{i_1,\dots,i_n\}$ is a permutation of $\{1,\dots,n\}$ and in general changes with time. We interpret x_t^i as the state of node i, $1\le i\le n$. Define

$$Z_t=[z_t^1,\dots,z_t^n]^T := [x_t^{i_1},\dots,x_t^{i_n}]^T. \tag{18}$$

Define the n scalar sequences

$$\{z_t^k, t\ge 0\},\quad 1\le k\le n. \tag{19}$$

We call $\{z_t^k, t\ge 0\}$ the level-k trajectory. By Lemma 5, we may estimate the difference between X_t and $X_t^{(-d)}$ to yield the next Lemma.

Lemma 6: Let $\mathbf{X}_t=[X_t; X_t^{(-1)};\dots;X_t^{(-d^*)}]$ be given by (13), and assume (**H1**)-(i). Then for some fixed constant $\tilde{c}$,

$$\max_{1\le d\le d^*}|X_t-X_t^{(-d)}|\le \tilde{c}a_{t-1},\quad t\ge 1. \tag{20}$$

□

Lemma 7: Assume (**H1**)-(i) and let Z_t be defined by (13) and (18). Then both $\{z_t^1, t\ge 0\}$ and $\{z_t^n, t\ge 0\}$ converge. □

The asymptotic behavior of the other sequences $\{z_t^k, t\ge 0\}$, $2\le k\le n-1$, is less obvious. The following theorem is a key result for establishing ergodicity of backward products of degenerating stochastic matrices. The basic idea of its proof is to use induction. First, Lemma 7 shows that the level-1 trajectory converges. Next, we show that each level-$(k+1)$ trajectory converges to the same limit as the level-k trajectory. Then by Lemma 6, the convergence of $\mathbf{X}_t$ follows.

Theorem 8: [9] Let $(\mathbf{X}_t, Z_t)$ be defined by (13) and (18) with any initial condition $\mathbf{X}_{t_0}$, $t_0\ge 0$. Assume (**H1**). Then there exists a number c depending on $\mathbf{X}_{t_0}$ such that

(i) $\lim_{t\to\infty} Z_t=c\mathbf{1}_n$;

(ii) $\lim_{t\to\infty}\mathbf{X}_t=c\mathbf{1}_{n(d^*+1)}$. □

V. Ergodicity of Degenerating Stochastic Matrices

Throughout subsections V-A and V-B, all matrices and graphs involved are deterministic.

A. Ergodicity of Backward Products

For the sequence $\{\tilde{\mathbf{A}}_t, t \geq 0\}$ of stochastic matrices specified in (13), define the backward product $\Phi_{t,s} = \tilde{\mathbf{A}}_{t-1} \dots \tilde{\mathbf{A}}_s$, $t \geq s \geq 0$, where $\Phi_{s,s} := I$. The following ergodicity theorem is an easy consequence of Theorem 8.

Theorem 9: [9] Assuming (**H1**), ergodicity holds for the backward products of $\{\tilde{\mathbf{A}}_t, t \geq 0\}$. □

B. Ergodicity over Jointly Strongly Connected Digraphs

Let $\{\bar{\mathbf{A}}_t, t \geq 0\}$ be a sequence of $n \times n(d^*+1)$ nonnegative matrices with unit row sums, and let $\{\tilde{G}_t = (\mathcal{N}, \tilde{\mathcal{E}}_t), t \geq 0\}$ be a sequence of digraphs. Define the square stochastic matrix

$$\widetilde{\mathbf{A}}_t = \begin{bmatrix} & \bar{\mathbf{A}}_t & \\ I_{nd^*} & & 0_{nd^* \times n} \end{bmatrix}. \qquad (21)$$

To analyze the backward products of $\{\widetilde{\mathbf{A}}_t, t \geq 0\}$, we introduce the following assumption for non-square nonnegative matrices as the counterpart of (**H1**)-(i).

(**H2**) $\{\bar{\mathbf{A}}_t, t \geq 0\}$ is (a_t)-compatible with $\{\tilde{G}_t, t \geq 0\}$, where $\{a_t, t \geq 0\}$ satisfies (**A1**). ♢

Denote $\mathbf{M}^{t,s} = \widetilde{\mathbf{A}}_{t+s-1} \dots \widetilde{\mathbf{A}}_t$, $s \geq 1$, in the form

$$\mathbf{M}^{t,s} = \begin{bmatrix} M_{00}^{t,s} & \dots & M_{0d^*}^{t,s} \\ \vdots & \ddots & \vdots \\ M_{d^*0}^{t,s} & \dots & M_{d^*d^*}^{t,s} \end{bmatrix},$$

where each $\mathbf{M}_{ij}^{t,s}$, $0 \leq i, j \leq d^*$, is an $n \times n$ matrix.

In the lemma below, we follow the rule in Section II to define the union of digraphs.

Lemma 10: Assume (**H2**). Let $h \geq \max\{d^*, 1\}$ be a fixed integer and denote $\tilde{G}_{[t,t+s)} = (\mathcal{N}, \tilde{\mathcal{E}}_{[t,t+s)})$ for $t \geq 0$ and $\max\{d^*, 1\} \leq s \leq h$. Then there exist constants $t_c' \geq 1$ and $0 < \underline{c}' \leq \overline{c}'$, all independent of (t,s), such that for all $t \geq t_c'$,

(i) the first n rows of $\mathbf{M}^{t,s}$ satisfy

$$\mathbf{M}^{t,s}(i,j) \leq \overline{c}' t^{-\gamma}, \quad \forall 1 \leq i \leq n, \quad 1 \leq j \leq n(d^*+1), \; j \neq i, \qquad (22)$$

$$\max_{0 \leq d \leq d^*} \mathbf{M}^{t,s}(i, j+dn) \geq \underline{c}' t^{-\gamma}, \quad \forall (j,i) \in \tilde{\mathcal{E}}_{[t,t+s)}; \qquad (23)$$

(ii) there exists a fixed constant c_0 independent of (t,s) such that $|\mathbf{M}_{d0}^{t,s} - I| \leq c_0 t^{-\gamma}$ for $1 \leq d \leq d^*$. □

Lemma 10 may be proved by matrix product estimates. It further implies the following compatibility result.

Lemma 11: Assume that (**H2**) holds and that there exists a sequence of integers $0 =: \tau_0 < \tau_1 < \dots$ such that $d^*+1 \leq \inf_{i \geq 0}(\tau_{i+1} - \tau_i) \leq \sup_{i \geq 0}(\tau_{i+1} - \tau_i) < \infty$. Then $\{\widehat{\mathbf{A}}_t, t \geq 0\}$ is (a_t)-compatible with $\{\widehat{G}_t, t \geq 0\}$, where $\widehat{\mathbf{A}}_t = \widetilde{\mathbf{A}}_{\tau_{t+1}-1} \dots \widetilde{\mathbf{A}}_{\tau_t}$ and $\widehat{G}_t = \tilde{G}_{[\tau_t, \tau_{t+1})}$. □

We state the ergodicity result for stochastic matrices associated with jointly strongly connected digraphs. The basic idea of its proof is to use $\{\widetilde{A}_t, t \geq 0\}$ to form a sequence of products on disjoint bounded time intervals, and next use Lemma 11 to check the compatibility condition and finally apply Theorem 9.

Theorem 12: [9] Assume (**H2**). In addition, there exists a sequence $0 =: \tau_0 < \tau_1 < \dots$ such that (i) $\sup_{i \geq 0}(\tau_{i+1} - \tau_i) < \infty$ and (ii) $\tilde{G}_{[\tau_i, \tau_{i+1})}$ is strongly connected for each $i \geq 0$. Then ergodicity holds for the backward products of $\{\widetilde{\mathbf{A}}_t, t \geq 0\}$. □

C. Application to Random Networks

By applying Theorem 12 to sample paths, we obtain the following corollary.

Corollary 13: Assume (i) $\{B_t, t \geq 1\}$ is given by (3)-(4) and (**A1**)-(**A2**) hold; (ii) $s_0(\omega)$ is an integer such that each $\mathbf{A}_t = \mathbf{U} + a_t \mathbf{B}_t$, $t \geq s_0(\omega)$, is a stochastic matrix. For each $\omega \in \Omega \backslash N_0$, ergodicity holds for the backward products of $\{\mathbf{A}_t(\omega), t \geq s_0(\omega)\}$. □

VI. Mean Square Consensus

Denote $\mathbf{A}_t = \mathbf{U} + a_t \mathbf{B}_t$. Theorem 14 below may be proved by using Theorem 3 and Corollary 13.

Theorem 14: [9] Under (**A1**)-(**A3**), mean square consensus holds for (5), i.e., $\lim_{t \to \infty} E|x_t^i - x^*|^2 = 0$ for some x^*. □

For leader following, convergence may be proved by a similar ergodicity approach.

Corollary 15: [9] In Theorem 14, if (**A2**) is replaced by (**A2′**) while other assumptions still hold, then $\lim_{t \to \infty} E|x_t^i - x_0^{i_L}|^2 = 0$ for all i. □

Remark: Theorem 14 and Corollary 15 may be generalized to correlated noises by using Theorem 3. ♢

Remark: If (1) is replaced by $y_t^{ik} = (x_s^k + w_s^{ik})|_{s=t-d_t^{ik}}$, Theorem 14 and Corollary 15 still hold. ♢

A. Generalization to Asynchronous Algorithms

We describe an asynchronous version of the consensus algorithm. Each node maintains a counter θ_t^i for generating a step size. Denote $\theta_t = [\theta_t^1, \dots, \theta_t^n]$.

(AU) Asynchronous update:

$$\theta_t^i = \sum_{s=1}^{t} 1_{\{|\mathcal{N}_{i,s}| > 0\}}, \quad i \in \mathcal{N}, \; t \geq 1, \qquad (24)$$

and $\theta_0^i = 0$, where $|\mathcal{N}_{i,s}|$ is the number of neighbors of node i at time s. So (24) means that the node increases its counter by one whenever it receives signals from its neighbors.

The algorithm for case (AU) is specified as follows:

$$x_{t+1}^i = [1 + a_{\theta_t^i} b_{ii}(t)] x_t^i + a_{\theta_t^i} \sum_{k \in \mathcal{N}_{i,t}} b_{ik}(t) y_t^{ik}, \quad t \geq 0, \qquad (25)$$

which is essentially driven by event times, i.e., the moments of receiving signals. Once initialized, this algorithm may be implemented without synchronized time slots although we use the pre-specified discrete times $0, 1, 2, \dots$ to describe (25). By combining (**A1**)-(**A3**) with some mild additional assumptions (mainly a moment condition for $\sup_{l \geq 1} |T_{l+1} - T_l|$), we may use the ergodicity approach to show mean square consensus for (25); see [9] for details.

VII. Simulation

Consider an undirected graph G of 4 nodes shown in Fig. 1. At time $t \geq 0$, node i transmits with probability $p_{i,t}$ with the exception that if up to time $t-1$ node i has not transmitted for L_i consecutive steps, it guarantees a transmission at t. The dependence of $p_{i,t}$ on time indicates non-stationarity of network switches. The delays are specified by $d_t^{12} = d_t^{21} = 0$, $d_t^{23} = d_t^{32} = 1$ and $d_t^{24} = d_t^{42} = 2$. The real-time network topology is determined according to signal receptions.

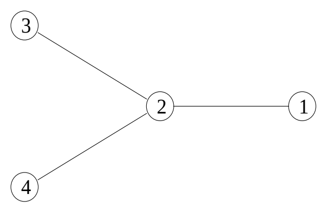

Fig. 1. The network.

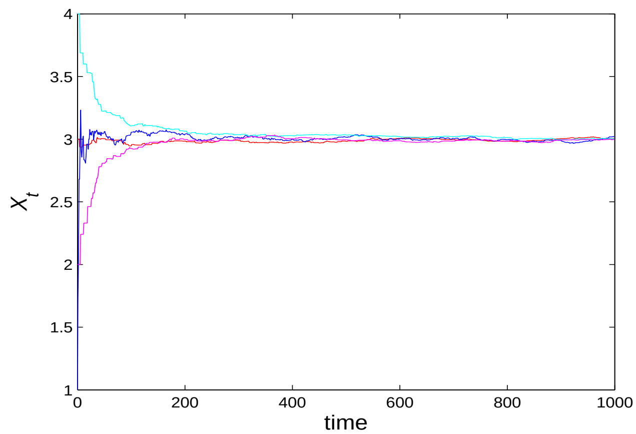

Fig. 2. Convergence of algorithm (5).

We take $p_{1,t} = 1/(2+2\sin^2(0.05t))$, $[p_{2,t}, p_{3,t}, p_{4,t}] \equiv [0.3, 0.4, 0.2]$, $b_{ik}(t) = 1$ if $k \in \mathcal{N}_{i,t}$, $a_t = (t+4)^{-0.7}$, $t \geq 0$, $[L_1, L_2, L_3, L_4] = [10, 8, 6, 10]$, and $X_0 = [3, 1, 4, 2]^T$. The i.i.d. Gaussian measurement noises have zero mean and variance 0.04. Fig. 2 shows the convergence of the states for algorithm (5).

VIII. Conclusion

We considered stochastic approximation for consensus seeking with delayed measurements in dynamic noisy environments. This paper developed a new approach by studying ergodicity of degenerating stochastic matrices and obtained convergence results without the usual double stochasticity condition. In future work, it will be of interest to further relax the bounded time interval condition for joint connectivity. A typical situation is Markovian switching networks [17], [10], where the bounded time interval condition in general fails.

References

[1] D. Acemoglu, A. Nedić, and A. Ozdaglar. Convergence of rule-of-thumb learning rules in social networks. *Proc. the 47th IEEE CDC*, Cancun, Mexico, pp. 1714-1720, Dec. 2008.

[2] T. C. Aysal and K. E. Barner. Convergence of consensus models with stochastic disturbances. *IEEE Trans. Info. Theory*, vol. 56, no. 8, pp. 4101-4113, 2010.

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

[4] F. Cucker and E. Mordecki. Flocking in noisy environments. *J. Math. Pures Appl.*, vol. 89, no. 3, pp. 278-296, 2008.

[5] L. Elsner, I. Koltracht, and M. Neumann. On the convergence of asynchronous paracontractions with applications to tomographic reconstruction from incomplete data. *Linear Algebra and Appl.*, vol. 130, pp. 65-82, 1990.

[6] L. Elsner, I. Koltracht, and M. Neumann. Convergence of sequential and asynchronous nonlinear paracontractions. *Numer. Math.*, vol. 62, pp. 305-319, 1992.

[7] L. Fang and P. J. Antsaklis. Asynchronous consensus protocols using nonlinear paracontraction theory. *IEEE Trans. Autom. Control*, vol. 53, no. 10, pp. 2351-2355, Nov. 2008.

[8] B. Gharesifard and J. Cortés. Distributed strategies for generating weight-balanced and doubly stochastic digraphs, submitted to *SIAM J. Contr. Optim.*, Oct. 2009. Available at http://arxiv.org/abs/0911.0232.

[9] M. Huang. Stochastic approximation for consensus: a new approach via ergodic backward products. Preprint, 2010.

[10] M. Huang, S. Dey, G. N. Nair, and J. H. Manton. Stochastic consensus over noisy networks with Markovian and arbitrary switches. *Automatica*, vol. 46, no. 10, pp. 1571-1583, 2010.

[11] M. Huang and J. H. Manton. Coordination and consensus of networked agents with noisy measurements: stochastic algorithms and asymptotic behavior. *SIAM J. Control Optim.*, vol. 48, no. 1, pp. 134-161, 2009.

[12] M. Huang and J. H. Manton. Stochastic consensus seeking with noisy and directed inter-agent communication: fixed and randomly varying topologies. *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 235-241, Jan. 2010.

[13] 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.

[14] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks with imperfect communication: link failures and channel noise. *IEEE Trans. Sig. Process.*, vol. 57, no. 1, pp. 355-369, 2009.

[15] A. Kashyap, T. Basar, and R. Srikant. Quantized consensus. *Automatica*, vol. 43, pp. 1192-1203, 2007

[16] T. Li and J.-F. Zhang. Mean square average-consensus under measurement noises and fixed topologies. *Automatica*, vol. 45, no. 8, pp. 1929-1936, 2009.

[17] I. Matei, N. Martins, and J. S. Baras. Almost sure convergence to consensus in Markovian random graphs. *Proc. 47th IEEE Conference on Decision and Control*, Cancun, Mexico, pp. 3535-3540, Dec. 2008.

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

[19] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Trans. Automatic Control*, vol. 49, pp. 1520-1533, Sep., 2004.

[20] A. Olshevsky and J. N. Tsitsiklis. On the nonexistence of quadratic Lyapunov functions for consensus algorithms. *IEEE Trans. Automatic Control*, vol. 53, no. 11, pp. 2642-2645, Dec. 2008.

[21] R. Rajagopal and M. J. Wainwright. Network-based consensus averaging with general noisy channels. arXiv:0805.0438v1, 2008.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] E. Seneta. *Non-negative Matrices and Markov Chains*. Revised 2nd ed., Springer, New York, 2006.

[27] S. S. Stankovic, M. S. Stankovic, and D. M. Stipanovic. Decentralized parameter estimation by consensus based stochastic approximation. *Proc. 46th IEEE CDC*, New Orleans, LA, pp. 1535-1540, Dec. 2007.

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

[29] J. Wolfowitz. Products of indecomposable, aperiodic, stochastic matrices. *Proc. American Math. Soc.*, vol. 14, no. 5, pp. 733-737, 1969.

[30] 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.