

## REFERENCES

- [1] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1984.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications Series)*. New York: Wiley, 1991.
- [3] N. C. Martins and M. A. Dahleh, "Fundamental limitations of performance in the presence of finite capacity feedback," in *Proc. Amer. Control Conf.*, 2005, vol. 1, pp. 79–86.
- [4] G. Zang and P. A. Iglesias, "Nonlinear extension of Bode's integral based on an information theoretic interpretation," *Syst. Control Lett.*, vol. 50, pp. 11–29, 2003.
- [5] N. Elia, "When Bode meets Shannon: Control-oriented feedback communication schemes," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1477–1488, Sep. 2004.
- [6] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*. Cambridge, U.K.: Cambridge Univ. Press, 1995.
- [7] G. N. Nair, R. J. Evans, and I. M. Y. Mareels, "Topological feedback entropy and nonlinear stabilization," *IEEE Trans. Autom. Control*, vol. 49, no. 9, pp. 1585–1597, Sep. 2004.
- [8] A. Lasota and M. C. Mackey, *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. New York: Springer-Verlag, 1994.
- [9] F. Fagnani and S. Zampieri, "Stability analysis and synthesis for scalar linear systems with quantized feedback," *IEEE Trans. Autom. Control*, vol. 48, no. 9, pp. 1569–1584, Sep. 2003.
- [10] G. Froyland, "Extracting dynamical behaviour via Markov models," in *Nonlinear Dynamics and Statistics: Proceedings, Newton Institute, Cambridge, 1998*. Boston, MA: Birkhauser, 2001, pp. 283–324.
- [11] Y. Kifer, *Ergodic Theory of Random Transformations (Progress of Probability and Statistics Series 10)*. Boston, MA: Birkhauser, 1986.
- [12] U. Vaidya and P. G. Mehta, "Lyapunov measure for almost everywhere stability," *IEEE Trans. Autom. Control*, to be published.
- [13] M. Dellnitz and O. Junge, "On the approximation of complicated dynamical behavior," *SIAM J. Numer. Anal.*, vol. 36, pp. 491–515, 1999.

## A Necessary and Sufficient Condition for Consensus Over Random Networks

Alireza Tahbaz-Salehi, *Student Member, IEEE*,  
and Ali Jadbabaie, *Senior Member, IEEE*

**Abstract**—We consider the consensus problem for stochastic discrete-time linear dynamical systems. The underlying graph of such systems at a given time instance is derived from a random graph process, independent of other time instances. For such a framework, we present a necessary and sufficient condition for almost sure asymptotic consensus using simple ergodicity and probabilistic arguments. This easily verifiable condition uses the spectrum of the average weight matrix. Finally, we investigate a special case for which the linear dynamical system converges to a fixed vector with probability 1.

**Index Terms**—Consensus problem, random graphs, tail events, weak ergodicity.

### I. INTRODUCTION

Decentralized iterative schemes such as agreement and consensus problems have an old history [1]–[4]. Over the past few years, they have attracted a significant amount of attention in various contexts such as motion coordination of autonomous agents [5], [6], distributed computation of averages and least squares among sensors [7]–[9], and rendezvous problems [10]. In all these cases, the dynamical system under study is deterministic. More recently, there has been some interest in the stochastic variants of the problem. In [11], the authors study the linear dynamical system  $x(k) = W_k x(k-1)$ , where the weight matrices  $W_k$  are independent, identically distributed (i.i.d.) stochastic matrices. It is shown that all the entries of  $x(k)$  converge to a common value almost surely (with probability 1), if each edge of  $\mathbf{G}(W_k)$ , the graph corresponding to matrix  $W_k$ , is chosen independently with the same probability (Erdős–Rényi random graph model). A more general model is studied in [12], where the edges of  $\mathbf{G}(W_k)$  are directed and not necessarily independent. However, the author proves only convergence to a consensus in probability, rather than the more general notion of almost sure convergence. Moreover, the assumption in [12] is the occurrence of scrambling matrices with positive probability, which can be weakened.

The purpose of this note is to provide a necessary and sufficient condition for an almost sure consensus in the linear dynamical system  $x(k) = W_k x(k-1)$ , when the weight matrices are general i.i.d. stochastic matrices. Our results contain the results of [11] and [12] as special cases. This necessary and sufficient condition is easily verifiable and only depends on the spectrum of the average weight matrix  $\mathbb{E}W_k$ . Finally, for a special case, we state a variant of a theorem in [2] that provides the asymptotic consensus value. Even though it is possible to derive our main theorem by combining results from ergodic theory of Markov chains in random environments [13]–[15], our proofs are self-contained and are only based on simple linear algebra machinery and the concept of *coefficients of ergodicity*, as introduced by Dobrushin [16].

Manuscript received September 8, 2006; revised January 30, 2007, June 28, 2007, and August 31, 2007. Recommended by Associate Editor F. Bullo.

This work was supported in part by the Department of Defense Multidisciplinary University Research Initiative (ARO/MURI) under Grant W911NF-05-1-0381, in part by the Office of Naval Research Young Investigator Program (ONR/YIP) under Grant N00014-04-1-0467, and in part by the National Science Foundation under Grant ECS-0347285.

The authors are with the General Robotics, Automation, Sensing and Perception (GRASP) Laboratory, Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104-6228 USA (e-mail: atahbaz@seas.upenn.edu; jadbabai@seas.upenn.edu).

Digital Object Identifier 10.1109/TAC.2008.917743

## II. PROBLEM SETUP

Let  $(\Omega_0, \mathcal{B}, \mu)$  be a probability space, where  $\Omega_0 = S_n = \{\text{set of stochastic matrices of order } n \text{ with strictly positive diagonal entries}\}$ ,  $\mathcal{B}$  is the Borel  $\sigma$ -algebra of  $\Omega_0$ , and  $\mu$  is a probability measure defined on  $\Omega_0$ . Define the product probability space as  $(\Omega, \mathcal{F}, \mathbb{P}) = \prod_{k=1}^{\infty} (\Omega_0, \mathcal{B}, \mu)$ . By definition, the elements of the product space have the following forms

$$\begin{aligned}\Omega &= \{(\omega_1, \omega_2, \dots) : \omega_k \in \Omega_0\} \\ \mathcal{F} &= \mathbb{B} \times \mathbb{B} \times \dots \\ \mathbb{P} &= \mu \times \mu \times \dots\end{aligned}$$

The above equations mean that the coordinates of the infinite dimensional vector  $\omega \in \Omega$  are i.i.d. stochastic matrices with positive diagonals.

Now consider the following random discrete-time dynamical system:

$$x(k) = W_k(\omega)x(k-1) \quad (1)$$

where  $k \in \{1, 2, \dots\}$  is the discrete time index,  $x(k) \in \mathbb{R}^n$  is the state vector at time  $k$ , and the mapping  $W_k : \Omega \rightarrow S_n$  is the  $k$ th coordinate function, which, for all  $\omega = (\omega_1, \omega_2, \dots) \in \Omega$ , is defined as

$$W_k(\omega) = \omega_k.$$

As a result, (1) defines a stochastic linear dynamical system in which the weight matrices are drawn independently from the common distribution  $\mu$ . For notational simplicity, we denote  $W_k(\omega)$  by  $W_k$  through the rest of the note.

For a general weight matrix  $W$ , one can define the corresponding graph  $\mathbf{G}(W)$  as a weighted directed graph with an edge  $(i, j)$  from vertex  $i$  to vertex  $j$  with weight  $W_{ji}$  if and only if  $W_{ji} \neq 0$ . In this case, we say vertex  $j$  has *access* to vertex  $i$ . We say vertices  $i$  and  $j$  *communicate* if both  $(i, j)$  and  $(j, i)$  are edges of  $\mathbf{G}(W)$ . Note that the communication relation is an equivalence relation and defines equivalence classes on the set of vertices. If no vertex in a specific communication class has access to any vertex outside that class, such a class is called *initial*. One important observation is the one-to-one correspondence between  $W$  and  $\mathbf{G}(W)$ . Also, note that because of the way  $\mathbf{G}(W)$  is defined, the relation  $x(k) = Wx(k-1)$  represents a distributed update scheme over the vertices of  $\mathbf{G}(W)$ . More precisely, the value of  $x_i(k)$  only depends on the elements of the set  $\{x_j(k-1) : W_{ij} \neq 0\} = \{x_j(k-1) : (j, i) \text{ is an edge of } \mathbf{G}(W)\}$ .

For the given dynamical system, we now define the notions of reaching state consensus in probability and almost surely.

*Definition 1:* Dynamical system (1) reaches consensus in probability, if for any initial state value  $x(0)$  and any  $\epsilon > 0$

$$\mathbb{P}(|x_i(k) - x_j(k)| > \epsilon) \rightarrow 0$$

as  $k \rightarrow \infty$  for all  $i, j = 1, \dots, n$ .

This notion of reaching state agreement asymptotically, which is addressed in [12], is a special case of reaching consensus almost surely, defined next.

*Definition 2:* Dynamical system (1) reaches consensus almost surely, if for any initial state value  $x(0)$

$$|x_i(k) - x_j(k)| \rightarrow 0 \quad \text{a.s.}$$

as  $k \rightarrow \infty$  for all  $i, j = 1, \dots, n$ .

Note that reaching almost sure consensus is stronger than reaching consensus in probability. In this case, not only the probability of the events  $\{|x_i(k) - x_j(k)| > \epsilon\}$  goes to zero for an arbitrary  $\epsilon > 0$  as  $k \rightarrow \infty$ , but also such events occur only finitely many times [17].

## III. ERGODICITY

Given (1), if  $x(0)$  is the initial state value, one can write the state vector at time  $k$  as

$$x(k) = W_k \cdots W_2 W_1 x(0). \quad (2)$$

As it is evident from (2), one needs to investigate the behavior of infinite products of stochastic matrices in order to check for an asymptotic consensus. This motivates us to borrow the concept of *weak ergodicity* of a sequence of stochastic matrices from the theory of Markov chains.

*Definition 3:* The sequence  $\{W_k\}_{k=1}^{\infty} = W_1, W_2, \dots$ , of  $n \times n$  stochastic matrices is weakly ergodic, if for all  $i, j, s = 1, \dots, n$  and all integer  $p \geq 0$

$$(U_{i,s}^{(k,p)} - U_{j,s}^{(k,p)}) \rightarrow 0$$

as  $k \rightarrow \infty$ , where  $U^{(k,p)} = W_{p+k} \cdots W_{p+2} W_{p+1}$  is the left product of the matrices in the sequence.

As the definition suggests, a sequence of stochastic matrices is weakly ergodic if the rows of the product matrix converge to each other, as the number of terms in the product grows. A closely related concept is *strong ergodicity* of a matrix sequence.

*Definition 4:* A sequence of  $n \times n$  stochastic matrices  $\{W_k\}_{k=1}^{\infty}$  is strongly ergodic, if for all  $i, s = 1, \dots, n$  and all integer  $p \geq 0$

$$U_{i,s}^{(k,p)} \rightarrow d_s^p$$

as  $k \rightarrow \infty$ , where  $U^{(k,p)}$  is the left product and  $d_s^p$  is a constant not depending on  $i$ .

One can easily see that weak and strong ergodicity<sup>1</sup> both describe a tendency to consensus. If either type of ergodicity (weak or strong) holds for the matrix sequence  $\{W_k\}_{k=1}^{\infty} = W_1, W_2, \dots$ , the pairwise differences between rows of the product matrix  $U^{(k,p)}$  converge to zero. Note that the converse of this statement is not true in general. In other words, the event of weak ergodicity of the sequence of matrices is a subset of the event that the linear dynamical system (1) reaches consensus asymptotically for all initial state values  $x(0)$ . For instance, the existence of a rank one matrix in the sequence implies asymptotic consensus, while it does not guarantee weak ergodicity.

At the first glance, it may seem that there exists a difference between weak and strong ergodicity. In the case of weak ergodicity, any two entries of vector  $x(k)$  converge to each other, but each entry does not necessarily converge to a limit. On the other hand, in the presence of strong ergodicity, not only the difference between any two entries converges to zero, but also all entries enjoy a common limit. Although one may consider the difference to be important, as the following theorem suggests, that is not the case [2], [18].

*Theorem 1:* Given a sequence of stochastic matrices  $\{W_k\}_{k=0}^{\infty}$  and their left products  $U^{(k,p)} = W_{k+p} \cdots W_{p+1}$ , weak and strong ergodicity are equivalent.

*Proof:* We only need to prove that weak ergodicity implies strong ergodicity. For any  $\epsilon > 0$ , weak ergodicity implies that for large enough  $k$ , we have  $-\epsilon \leq U_{i,s}^{(k,p)} - U_{j,s}^{(k,p)} \leq \epsilon$  uniformly for all  $i, j, s = 1, \dots, n$ . Since  $U^{(k+1,p)} = W_{k+p+1} U^{(k,p)}$ , we have

$$U_{i,s}^{(k,p)} - \epsilon \leq U_{h,s}^{(k+1,p)} \leq U_{i,s}^{(k,p)} + \epsilon$$

which, by induction, implies that

$$U_{i,s}^{(k,p)} - \epsilon \leq U_{h,s}^{(k+r,p)} \leq U_{i,s}^{(k,p)} + \epsilon$$

<sup>1</sup>To be more precise, we have stated the definitions of weak and strong ergodicity *in the backward direction*. Since this is the only type of ergodicity that we deal with, we simply refer to these properties as ergodicity.

for all  $i, s, h = 1, \dots, n$  and  $r \geq 0$ . By setting  $h = i$ , it is evident that  $U_{i,s}^{(k,p)}$  is a Cauchy sequence, and therefore,  $\lim_{k \rightarrow \infty} U_{i,s}^{(k,p)}$  exists. ■

Therefore, weak ergodicity is equivalent to the existence of a vector  $d$  satisfying  $U^{(k,p)} \rightarrow \mathbf{1}d^T$ , in which  $\mathbf{1}$  is a vector with all entries equal to one. We now define the *coefficient of ergodicity*, which is a key concept in proving weak ergodicity results.

**Definition 5:** The scalar continuous function  $\tau(\cdot)$  defined on the set of  $n \times n$  stochastic matrices is called a coefficient of ergodicity if it satisfies  $0 \leq \tau(\cdot) \leq 1$ . A coefficient of ergodicity is said to be *proper* if

$$\tau(W) = 0, \quad \text{if and only if } W = \mathbf{1}d^T$$

where  $d$  is a vector of size  $n$  satisfying  $d^T \mathbf{1} = 1$ .

Two examples of coefficients of ergodicity used in this note are

$$\begin{aligned} \kappa(W) &= \frac{1}{2} \max_{i,j} \sum_{s=1}^n |W_{is} - W_{js}| \\ \nu(W) &= 1 - \max_i (\min_j W_{ij}). \end{aligned}$$

Note that  $\nu(\cdot)$  is an improper coefficient of ergodicity, while  $\kappa(\cdot)$  is proper, and for any stochastic matrix  $W$ , they satisfy

$$\kappa(W) \leq \nu(W). \quad (3)$$

Given the above definitions, it is straightforward to show that weak ergodicity is equivalent to

$$\tau(U^{(k,p)}) \rightarrow 0 \quad \forall p \in \mathbb{N} \cup \{0\}$$

as  $k \rightarrow \infty$  for a proper coefficient of ergodicity  $\tau$ . Therefore, we can state the following theorem.

**Theorem 2:** Suppose  $\tau(\cdot)$  is a proper coefficient of ergodicity that for any  $m \geq 1$  stochastic matrices  $W_k, k = 1, 2, \dots, m$  satisfies

$$\tau(W_m \cdots W_2 W_1) \leq \prod_{k=1}^m \tau(W_k). \quad (4)$$

Then, the sequence  $\{W_k\}_{k=1}^\infty$  is weakly ergodic if and only if there exists a strictly increasing sequence of integers  $k_r, r = 1, 2, \dots$ , such that

$$\sum_{r=1}^{\infty} (1 - \tau(W_{k_{r+1}} \cdots W_{k_r})) = \infty. \quad (5)$$

*Proof:* Since only the sufficiency part of this theorem will be used in this note, we only prove that (5) implies weak ergodicity of the sequence. A proof for the reverse implication can be found in [18, Th. 4.18].

Suppose that there exists a strictly increasing sequence of positive integers  $k_r$  such that (5) holds. Then, the inequality  $\log x \leq x - 1$  implies that

$$\sum_{r=1}^{\infty} \log(\tau(W_{k_{r+1}} \cdots W_{k_r})) = -\infty$$

and, as a result,  $\prod_{r=1}^{\infty} \tau(W_{k_{r+1}} \cdots W_{k_r}) = 0$ . Because we assumed that  $\tau$  is proper, (4) guarantees weak ergodicity of the sequence. ■

#### IV. MAIN RESULTS: NECESSARY AND SUFFICIENT CONDITIONS FOR ERGODICITY

In this section, we study the necessary and sufficient conditions for ergodicity of an i.i.d. sequence of stochastic matrices based on the framework presented in Section II.

**Lemma 1:** The weak ergodicity of the sequence  $W_1, W_2, \dots$  is a trivial event.

*Proof:* Let  $k$  be a positive integer. Define the event

$$A_k = \{\text{The sequence } W_k, W_{k+1}, \dots \text{ is weakly ergodic.}\}$$

which is an event in  $\mathcal{F}'_k = \sigma(W_k, W_{k+1}, \dots)$ . These events form a *decreasing sequence* of events, satisfying  $A_1 \supseteq A_2 \supseteq \dots$ . As a result,

$$\bigcap_{k=1}^{\infty} A_k \in \mathcal{T}$$

where  $\mathcal{T}$  is the tail  $\sigma$ -field of the sequence of stochastic matrices defined as  $\mathcal{T} = \bigcap_{k=1}^{\infty} \mathcal{F}'_k$ . Therefore, by Kolmogorov's 0-1 law [17],  $\mathbb{P}(\bigcap_k A_k) = 0$  or 1. Now we have

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{j \rightarrow \infty} \mathbb{P}\left(\bigcap_{k=1}^j A_k\right) = \lim_{j \rightarrow \infty} \mathbb{P}(A_j) = \mathbb{P}(A_1)$$

where the first equality is due to continuity of the probability measure. The last equality holds because the distribution of  $W_k$  does not depend on  $k$ . Therefore, weak ergodicity of  $W_1, W_2, \dots$  is trivial. ■

This lemma indicates that weak ergodicity of random i.i.d. weight matrices obeys a 0-1 law. In other words, the sequence of stochastic matrices is weakly ergodic either almost surely or almost never, indicating a discontinuous behavior. In order to find a criterion to distinguish between these two cases, we need another lemma, the proof of which can be found in [19].

**Lemma 2:** Suppose that  $W$  is a stochastic matrix for which its corresponding graph has  $s$  communication classes named  $\alpha_1, \dots, \alpha_s$ . Class  $\alpha_r$  is initial, if and only if the spectral radius of  $\alpha_r[W]$  equals to 1, where  $\alpha_r[W]$  is the submatrix of  $W$  corresponding to the vertices in the class  $\alpha_r$ .

Finally, suppose that the average weight matrix  $\mathbb{E}W_k$  has  $n$  eigenvalues satisfying

$$0 \leq |\lambda_n(\mathbb{E}W_k)| \leq \dots \leq |\lambda_2(\mathbb{E}W_k)| \leq |\lambda_1(\mathbb{E}W_k)| = 1.$$

At this point, we state our main theorem.

**Theorem 3:** For a given random i.i.d. sequence  $\{W_k\}_{k=0}^\infty = W_1, W_2, \dots$  of stochastic matrices with positive diagonals, the following three statements are equivalent.

- The random sequence  $\{W_k\}_{k=0}^\infty$  is (weakly) ergodic almost surely.
- The deterministic discrete-time linear dynamical system  $x(k) = (\mathbb{E}W_k)x(k-1)$  reaches a consensus asymptotically.
- $|\lambda_2(\mathbb{E}W_k)| < 1$ .

*Proof:* First, we show that (a) implies (b). If the random sequence  $W_1, W_2, \dots$  is weakly ergodic with probability 1, we have

$$(U_{i,s}^{(k,p)} - U_{j,s}^{(k,p)}) \rightarrow 0 \quad \text{a.s.}$$

Therefore, the dominated convergence theorem [17] implies  $(\mathbb{E}W)_{i,s}^k - (\mathbb{E}W)_{j,s}^k \rightarrow 0$ , which is another way of stating (b).

In order to show that (b) implies (c), we assume that  $|\lambda_2(\mathbb{E}W_k)| = 1$ . Since all  $\omega_k \in \Omega_0$  have positive diagonals,  $\mathbb{E}W_k$  has strictly positive diagonal entries as well. Hence, if  $\mathbb{E}W_k$  is irreducible, then it is primitive, and, as a result of the Perron–Frobenius theorem [19],  $|\lambda_2(\mathbb{E}W_k)| < 1$ , which is in contradiction with our assumption. Therefore,  $|\lambda_2(\mathbb{E}W_k)| = 1$  implies reducibility of  $\mathbb{E}W_k$ . As a result, without the loss of generality, one can label the vertices such that  $\mathbb{E}W_k$  gets the

following block triangular form

$$\mathbb{E}W_k = \begin{bmatrix} Q_{11} & 0 & \cdots & 0 \\ Q_{21} & Q_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ Q_{s1} & Q_{s2} & \cdots & Q_{ss} \end{bmatrix} \quad (6)$$

where each  $Q_{ii}$  is an irreducible matrix and represents the vertices in the equivalence class  $\alpha_i$ . Since  $|\lambda_2(\mathbb{E}W_k)| = 1$ , submatrices corresponding to at least two of the classes have unit spectral radii (note that because of irreducibility and aperiodicity of  $Q_{ii}$ 's, the multiplicity of the unit-modulus eigenvalue of each one of them cannot be more than 1). Therefore, lemma 2 implies

$$\exists i \neq j \text{ s.t. } \alpha_i \text{ and } \alpha_j \text{ are both initial classes}$$

or equivalently,  $Q_{ir} = 0$  for all  $r \neq i$  and  $Q_{jl} = 0$  for all  $l \neq j$ . In other words, the matrix  $\mathbb{E}W_k$  has two orthogonal rows, and, as a result, (b) cannot hold.

The last implication can be proved by combining results from the ergodic theory of Markov chains in random environments, more specifically by using [13]–[15]. However, here we provide a simpler proof based on theorem 2. In order to do so, we assume that (c) holds. Since  $|\lambda_2(\mathbb{E}W_k)|$  is subunit, lemma 2 implies that  $\mathbf{G}(\mathbb{E}W_k)$  has exactly one initial class. We investigate the two cases of  $\mathbb{E}W_k$  being irreducible and reducible separately.

1) *Irreducible Case:* Suppose  $\mathbb{E}W_k$  is irreducible. Since it has only one unit-modulus eigenvalue,  $\mathbb{E}W_k$  is primitive [19]. Hence,

$$\exists m \text{ s.t. } [\mathbb{E}W_k]^m > 0$$

where by  $> 0$  for a matrix, we mean entrywise positivity. Independence over time implies

$$\mathbb{P}(W_m \cdots W_1) = [\mathbb{E}W_k]^m > 0.$$

As a result, for all  $i, j = 1, \dots, n$ , the  $(i, j)$  entry of  $U^{(m,0)} = W_m \cdots W_2 W_1$  is positive with nonzero probability, say  $p_{ij} > 0$ . Therefore, since the weight matrices are i.i.d. with positive diagonals, the matrix  $W_{n^2 m} \cdots W_2 W_1$  is completely entrywise positive with at least probability  $\prod_{i,j} p_{ij} > 0$ , i.e., the event  $\{W_{n^2 m} \cdots W_2 W_1 > 0\}$  has nonzero probability.<sup>2</sup> As a result, if we define  $\delta(W) = 1 - \nu(W) = \max_j (\min_i W_{ij})$ , then there exists  $\epsilon > 0$  such that

$$\mathbb{P}(\delta(W_{n^2 m} \cdots W_2 W_1) > \epsilon) > 0.$$

Hence, by the second Borel–Cantelli lemma [17, p. 49], we have

$$\mathbb{P}(\delta(W_{(r+1)n^2 m} \cdots W_{rn^2 m+1}) > \epsilon \text{ for infinitely many } r) = 1.$$

Once we set  $k_r = rn^2 m$ , we have

$$\delta(W_{k_{r+1}} \cdots W_{k_r+1}) > \epsilon \quad \text{i.o. a.s.}$$

Also note that (4) hold for  $\tau(\cdot) = \kappa(\cdot)$ . Therefore, this together with (3) implies

$$\sum_{r=1}^{\infty} (1 - \kappa(W_{k_{r+1}} \cdots W_{k_r+1})) = \infty \quad \text{a.s.}$$

<sup>2</sup>Selecting  $n^2 m$  as the number of matrices in the product in order to make the product matrix entrywise positive is a conservative pick. The product matrix becomes positive with nonzero probability with less terms, but such a pick suffices our purpose.

which is exactly (5), the sufficient condition for weak ergodicity. Therefore, the sequence is weakly ergodic almost surely.

2) *Reducible Case:* When  $\mathbb{E}W_k$  is reducible, without loss of generality, it can be written as (6), where all  $Q_{ii}$  are irreducible matrices. Since  $\alpha_1$  (the class corresponding to submatrix  $Q_{11}$ ) is the only initial class of  $\mathbf{G}(\mathbb{E}W_k)$ , there exists a directed path from a vertex in  $\alpha_1$  (e.g., say, vertex labeled 1) to any vertex of  $\mathbf{G}(\mathbb{E}W_k)$ , such that the length of the path is at most some positive integer  $m$ . In other words, any vertex of  $\mathbf{G}(\mathbb{E}W_k)$  is at most an  $m$ -hop neighbor of vertex 1. This combined with the fact that  $\mathbb{E}W_k$  has strictly positive diagonals guarantees that the first column of  $[\mathbb{E}W_k]^m$  is strictly positive. Therefore, as in case 1, independence implies the positivity of the first column of  $\mathbb{E}(W_m \cdots W_1)$ .

As a result, for  $j = 1, \dots, n$  the  $(j, 1)$  entry of the matrix  $W_m \cdots W_1$  is nonzero with positive probability  $\hat{p}_{j1}$ . Hence, identical to the discussion of case 1, we have

$$\exists \epsilon > 0 \text{ s.t. } \mathbb{P}(\delta(W_{nm} \cdots W_2 W_1) > \epsilon) \geq \prod_{j=1}^n \hat{p}_{j1} > 0.$$

Now, if we set  $k_r = rnm$ , once again the second Borel–Cantelli lemma guarantees that

$$\mathbb{P}(\delta(W_{k_{r+1}} \cdots W_{k_r+1}) > \epsilon \text{ for infinitely many } r) = 1.$$

Therefore, the sum  $\sum_{r=1}^{\infty} (1 - \kappa(W_{k_{r+1}} \cdots W_{k_r+1}))$  diverges to infinity with probability 1, and theorem 2 implies that the random sequence  $\{W_k\}_{k=0}^{\infty}$  is weakly ergodic almost surely. This completes the proof. ■

Theorem 3 combined with lemma 1 provides a simple criterion to distinguish between the two cases of almost sure and almost never weak ergodicity. It suggests that the information in the average weight matrix  $\mathbb{E}W_k$  suffices to predict the long-run behavior of the left product matrices  $U^{(k,p)}$ . The following corollary states that the same information is sufficient to extract the asymptotic convergence properties of the linear dynamical system (1).

*Corollary 4:* The linear dynamical system (1) reaches consensus almost surely if and only if  $|\lambda_2(\mathbb{E}W_k)| < 1$ . Otherwise, it reaches asymptotic consensus almost never.

*Proof:* According to theorem 3,  $|\lambda_2(\mathbb{E}W_k)| < 1$  guarantees weak ergodicity with probability 1, and, as a result, the event of asymptotic consensus occurs with probability 1, since it is a superset of the weak ergodicity event. To prove the reverse implication, note that when  $\mathbb{E}W_k$  has more than one unit-modulus eigenvalues, as in the proof of theorem 3, its corresponding graph has more than one initial class, which implies that  $\mathbb{E}W_k$  has two orthogonal rows. Since  $\Omega_0$  is a subset of non-negative matrices,  $W_k$  has the same type (zero block pattern) as  $\mathbb{E}W_k$  for all time  $k$  with probability 1. Therefore,  $U^{(k,0)} = W_k \cdots W_2 W_1$  has two orthogonal rows almost surely for any  $k$ , which means that the random discrete-time dynamical system (1) reaches a consensus with probability 0. ■

Therefore,  $|\lambda_2(\mathbb{E}W_k)| < 1$  provides a necessary and sufficient condition for almost sure asymptotic consensus in (1). This should not come as a surprise to the reader. In fact, when  $|\lambda_2(\mathbb{E}W_k)|$  is subunit, there exists a sequence of integer numbers  $k_r$ ,  $r = 1, 2, \dots$  such that  $\kappa(W_{k_{r+1}} \cdots W_{k_r+1}) < 1 - \epsilon$  with probability 1, and therefore, the product of matrices  $\{W_{k_{r+1}}, \dots, W_{k_r+1}\}$  are scrambling. This also implies that the collection of graphs  $\{\mathbf{G}(W_{k_{r+1}}), \dots, \mathbf{G}(W_{k_r+1})\}$  is jointly connected (i.e., the graph constructed by forming the union of the edge sets of the graphs in the collection contains a spanning tree) [5] almost surely. This infinite often connectivity over time guarantees the possibility of information flow on the graph over time, and therefore, results in asymptotic consensus with probability 1. On the other hand,

when  $|\lambda_2(\mathbb{E}W_k)| = 1$ , no such sequence exists, and therefore, there are at least two classes of vertices in the graph such that they never have access to each other, and hence, no consensus.

Moreover, theorem 3 contains the results of [11] and [12] as special cases. Since in [11], the authors use Erdős–Rényi as their random graph model, the matrix  $\mathbb{E}W_k$  is completely entrywise positive, which results in  $|\lambda_2(\mathbb{E}W_k)| < 1$ , and hence, almost sure consensus. On the other hand, when the weight matrices are scrambling with positive probability, as in [12],  $\mathbb{E}W_k$  is also scrambling, and, as a result, its unit-modulus eigenvalue has multiplicity 1. Hence, (1) reaches an asymptotic consensus almost surely (and therefore, in probability).

## V. CONSENSUS VALUE

As shown in the previous section, if  $|\lambda_2(\mathbb{E}W_k)|$  is subunit, then the linear dynamical system (1) converges with probability 1 and  $x(k) \xrightarrow{\text{a.s.}} c\mathbf{1}$ , where  $c$  is a scalar random variable depending on the initial state value  $x(0)$  and the random sequence of weight matrices. The following theorem states that the distribution of  $c$  is concentrated at one point if all weight matrices have a common left eigenvector corresponding to their unit eigenvalue.

*Theorem 5:* For  $y \in \mathbb{R}^n$ , set  $S(y) = \{W \in S_n | y^T W = y^T\}$ . For a given initial state value vector  $x(0)$ , if  $|\lambda_2(\mathbb{E}W_k)| < 1$  and  $\mu(S_n - S(y)) = 0$  hold, then

$$\lim_{k \rightarrow \infty} x(k) = (y^T x(0)) \mathbf{1} \quad \text{a.s.}$$

*Proof:* A variant of this theorem is proved in [2]. A similar proof can be used here as well. ■

One special case of the above theorem is when the weight matrices are doubly stochastic almost surely. In such a case, the weight matrices have vector  $\mathbf{1}$  as their common left eigenvector at all times, and therefore, all the entries of the state vector converge to  $(1/n)(\mathbf{1}^T x(0))\mathbf{1}$ , the average of the initial state values, with probability 1. This special case is addressed in [20].

## VI. CONCLUSION

In this note, we showed how the problem of reaching consensus can be reduced to the problem of weak ergodicity of a sequence of matrices. In particular, for the case of i.i.d. weight matrices, we showed that ergodicity is a trivial event. Moreover, we showed that the discrete-time linear dynamical system  $x(k) = W_k x(k-1)$  reaches state consensus almost surely if and only if  $\mathbb{E}W_k$  has exactly one eigenvalue with unit modulus. Finally, we showed how our theorem simply recovers the other known results in the field of consensus over random graphs.

## REFERENCES

- [1] M. H. DeGroot, "Reaching a consensus," *J. Am. Statistical Assoc.*, vol. 69, no. 345, pp. 118–121, Mar. 1974.
- [2] S. Chatterjee and E. Seneta, "Towards consensus: Some convergence theorems on repeated averaging," *J. Appl. Probab.*, vol. 14, no. 1, pp. 89–97, Mar. 1977.
- [3] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," *Ph.D.*, dissertation, Massachusetts Inst. Technol., Cambridge, MA, 1984.
- [4] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.
- [5] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom. Control*, vol. 48, no. 6, pp. 988–1001, Jun. 2003.
- [6] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," in *Proc. Joint 44th IEEE Conf. Decision Control Eur. Control Conf.*, Seville, Spain, Dec. 2005, pp. 2996–3000.
- [7] S. Boyd, A. Gosh, B. Prabhakar, and D. Shah, "Gossip algorithms: Design, analysis and applications," in *Proc. IEEE INFOCOM 2005*, vol. 3, Miami, FL, Mar. 2005, pp. 1653–1664.
- [8] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. 4th Int. Conf. Inf. Process. Sensor Netw.*, Apr., Apr. 2005, pp. 63–70.
- [9] 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.
- [10] J. Cortes, S. Martinez, and F. Bullo, "Robust rendezvous for mobile autonomous agents via proximity graphs in arbitrary dimensions," *IEEE Trans. Autom. Control*, vol. 51, no. 8, pp. 1289–1298, Aug. 2006.
- [11] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Trans. Autom. Control*, vol. 50, no. 11, pp. 1867–1872, Nov. 2005.
- [12] C. W. Wu, "Synchronization and convergence of linear dynamics in random directed networks," *IEEE Trans. Autom. Control*, vol. 51, no. 7, pp. 1207–1210, Jul. 2006.
- [13] R. Cogburn, "The ergodic theory of Markov chains in random environments," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 66, no. 1, pp. 109–128, 1984.
- [14] R. Cogburn, "On products of random stochastic matrices," in *Proc. Am. Math. Soc. Summer Res. Conf. Random Matrices Appl.*, vol. 50, Providence, RI, 1986, pp. 199–213.
- [15] D. Griffeath, "A maximal coupling for Markov chains," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 31, no. 2, pp. 95–106, 1975.
- [16] R. L. Dobrushin, "Central limit theorem for non-stationary Markov chains, I, II," *Theory Probab. Appl.*, vol. 1, pp. 65–80, 329–383, 1956, (English Translation).
- [17] R. Durrett, *Probability: Theory and Examples*, 3rd ed. Belmont, CA: Duxbury Press, 2005.
- [18] E. Seneta, *Non-Negative Matrices and Markov Chains*, 2nd ed. New York: Springer-Verlag, 1981.
- [19] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. New York: Academic, 1979.
- [20] S. Boyd, A. Gosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *Special Issue IEEE Trans. Inf. Theory IEEE/ACM Trans. Netw.*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.