Consensus Over Ergodic Stationary Graph Processes

Alireza Tahbaz-Salehi and Ali Jadbabaie

***Abstract*— In this paper, we provide a necessary and sufficient condition for almost sure convergence of consensus algorithms when the underlying graphs of the network are generated by a strictly stationary and ergodic random process. We prove that for such a process, the consensus algorithm converges almost surely, if and only if, the expected graph of the network contains a directed spanning tree. Our results contain the case of independent and identically distributed graph processes as a special case. Moreover, we provide the mean and the variance of the asymptotic random consensus value, as well as a necessary and sufficient condition for its distribution to be degenerate.**

***Index Terms*— Ergodic stationary process, Consensus problem, Random graph, weak ergodicity.**

I. INTRODUCTION

Over the past few years, decentralized iterative schemes such as agreement and consensus problems have attracted a significant amount of interest in various contexts such as motion coordination of autonomous agents [1], parallel computing [2] and sensor networks [3], [4]. In general, the focus of these papers have been on studying the asymptotic behavior of the linear dynamical system $x(k) = W_k x(k-1)$, where $\{W_k\}_{k=1}^{\infty}$ is a sequence of stochastic matrices. More recently, conditions for asymptotic convergence of such systems to a consensus for the case that the weight matrices W_k are random have also been studied. For instance, in [5], the authors prove that all the entries of $x(k)$ converge to a common value almost surely (with probability one), if each edge of $\mathbf{G}(W_k)$, the graph corresponding to matrix W_k, is chosen independently with the same positive probability (what is known as the Erdős-Rényi random graph model), followed by more general models in [6] and [7]. In [8], we showed that these results can be further generalized to the case of i.i.d. weighted and directed random graph sequences, in which the edges of the graph at the same time step might be dependent, while the realizations of the network's graph at two different time steps are independent. We proved that a necessary and sufficient condition for almost sure convergence to consensus is $|\lambda_2(\mathbb{E}W_k)| < 1$, where λ_2 is the eigenvalue with the second largest modulus. The speed of convergence to consensus and some concentration results for the general i.i.d. case is presented in [9].

The common crucial assumption of all the above works is that the weight matrices (and hence, the underlying graphs of the network) are independent and identically distributed. However, in most realistic cases, there exists a strong correlation between the realization of the network's graph over time. For example, the existence of a communication link in a wireless network at a given time instance is highly correlated with the existence of that link at the previous time steps. In this note, we relax this assumption and assume that the weight matrices W_k are generated by an ergodic and stationary process. Building on the results of [10] and by applying Birkhoff's ergodic theorem, we show that the condition $|\lambda_2(\mathbb{E}W_k)| < 1$ that appeared in [8] is still a necessary and sufficient condition for almost sure convergence to consensus. This condition implies that existence of a directed path in the expected graph of the network from some node to all other nodes is both necessary and sufficient for reaching consensus with probability one. The results presented in this note are more general than [5]–[9], which assume that the weight matrices are distributed identically and independently over time. Also contrary to the results in [10], in which the authors consider only unweighted edges one at a time, we consider a general ergodic stationary process of stochastic matrices. Results similar to ours have been stated without proof in [11] and [12] in the context of ergodic theory of Markov chains in random environments. In contrast to those results, our proofs are self-contained and are only based on simple results from the theory of non-negative matrices and the concept of *coefficients of ergodicity*, as introduced by Dobrushin [13].

In addition, we state and prove a necessary and sufficient condition for the distribution of the random consensus value to be degenerate, i.e., the consensus value to be a random variable with a single point as its support. Finally, we compute the mean and the variance of the random consensus value for the i.i.d. case.

This research is supported in parts by the following grants: ARO/MURI W911NF-05-1-0381, DARPA/DSO SToMP, ONR MURI N000140810747, and NSF-ECS-0347285.

A. Tahbaz-Salehi is with Department of Electrical and Systems Engineering and Department of Economics, University of Pennsylvania, Philadelphia, PA 19104 (e-mail: atahbaz@seas.upenn.edu).

A. Jadbabaie is with the General Robotics, Automation, Sensing and Perception (GRASP) Laboratory, Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 (e-mail: jadbabai@seas.upenn.edu).

II. ERGODIC STATIONARY MATRIX PROCESSES

This section is dedicated to some basic definitions and concepts from ergodic theory and the theory of random processes. A thorough treatment of the subject can be found in [14] and [15].

Let $(\Omega_0, \mathcal{B})$ be a measurable space, where $\Omega_0 = \{$set of stochastic matrices of order n with strictly positive diagonal entries$\}$ and $\mathcal{B}$ is the Borel σ-algebra on Ω_0. Consider a probability measure $\mathbb{P}$ defined on the sequence space $(\Omega, \mathcal{F})$ defined as

$$\Omega = \{(\omega_1, \omega_2, \dots) : \omega_k \in \Omega_0\}$$
$$\mathcal{F} = \mathcal{B} \times \mathcal{B} \times \dots,$$

so that $(\Omega, \mathcal{F}, \mathbb{P})$ forms a probability space. If we let $\varphi : \Omega \to \Omega$ be the shift operator defined as $\varphi(\omega_1, \omega_2, \dots) = (\omega_2, \omega_3, \dots)$ and let $W(\omega) = \omega_1$, then we can define a sequence of stochastic matrices as $W_k(\omega) = W(\varphi^k \omega)$. For notational simplicity, we denote $W_k(\omega)$ by W_k through the rest of the note.

Definition 1: The sequence of random stochastic matrices $W_1, W_2, \dots$ is *stationary* if the families $\{W_{k_1}, W_{k_2}, \dots, W_{k_r}\}$ and $\{W_{k_1+h}, W_{k_2+h}, \dots, W_{k_r+h}\}$ have the same joint distribution for all $k_1, k_2, \dots, k_r$ and all $h > 0$.

The above definition states that the process $\{W_k : k \geq 1\}$ is stationary if all of its finite dimensional distributions are invariant under time shifts. Equivalently, one can define stationarity as the case that the shift operator is a *measure-preserving* transformation, i.e., $\mathbb{P}(\varphi B) = \mathbb{P}(B)$ for all sets $B \in \mathcal{F}$. Clearly, any i.i.d. sequence of random matrices is stationary.

Definition 2: Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and suppose that the shift operator $\varphi : \Omega \to \Omega$ is measure-preserving. φ is said to be *ergodic* if every invariant set $B \in \mathcal{F}$ is trivial.

In other words, the transformation φ is ergodic if for every $B \in \mathcal{F}$ satisfying $\mathbb{P}(B \Delta \varphi^{-1} B) = 0$, we have $\mathbb{P}(B) \in \{0, 1\}$, where Δ denotes the symmetric difference between the two sets. Given the above definitions, we say the random matrix process $\{W_k : k \geq 1\}$ is ergodic stationary, if the shift operator defined over $(\Omega, \mathcal{F}, \mathbb{P})$ is measure-preserving and ergodic. For example, a time-invariant Markov chain with its unique stationary distribution as the initial distribution forms a stationary and ergodic process. Clearly, any i.i.d. sequence of matrices is also ergodic stationary. In this note, we use the following lemma for proving our results.

Lemma 1: Suppose $W_1, W_2, \dots$ is an ergodic stationary process of stochastic $n \times n$ matrices. If the event $\{W_k \in A\}$ has positive probability $p > 0$, then such events occur infinitely often almost surely, that is, $\mathbb{P}(W_k \in A \text{ for infinitely many } k) = 1$.

Proof: Since the process $\{W_k : k \geq 1\}$ is ergodic stationary, so is the process $\{\mathbb{I}_{\{W_k \in A\}} : k \geq 1\}$, where $\mathbb{I}$ is the indicator function. Therefore, by Birkhoff's ergodic theorem [14], [15],

$$\frac{1}{T}\sum_{k=1}^{T} \mathbb{I}_{\{W_k \in A\}} \to \mathbb{P}\{W_1 \in A\} = p \quad \textit{almost surely},$$

which implies

$$\mathbb{P}\left(\sum_{k=1}^{\infty} \mathbb{I}_{\{W_k \in A\}} = \infty\right) = 1.$$

This means that the events $\{W_k \in A\}$ occur infinitely often almost surely. ∎

III. Consensus Over Random Graphs

Consider the discrete-time autonomous dynamical system

$$x(k) = W_k(\omega)x(k-1), \tag{1}$$

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 $W_k(\omega)$ is a random stochastic matrix with strictly positive diagonals, defined as before. Since the matrices W_k are random, the state vector $x(k)$ is also a random vector for all $k \geq 1$.

In a general *randomized consensus problem*, we are interested in the asymptotic behavior of the state vector $x(k)$ as k goes to infinity.

Definition 3: Dynamical system (1) reaches consensus almost surely, if for any initial state value $x(0)$,

$$|x_i(k) - x_j(k)| \to 0 \quad \textit{almost surely}$$

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

Note that reaching almost sure consensus is stronger than reaching consensus *in probability*. In the former, not only the probability of the event $\{|x_i(k) - x_j(k)| > \epsilon\}$ goes to zero for an arbitrary $\epsilon > 0$ as $k \to \infty$, but also such events occur only finitely many times [14].

In this paper, we are interested in necessary and sufficient conditions for almost sure convergence of dynamical system (1) to consensus, when the weight matrices W_k are generated by an ergodic stationary process as defined in section II. Since any sequence of independent and identically distributed matrices is also ergodic stationary, any sufficient condition for almost sure consensus in the latter case serves also as a sufficient condition for the i.i.d. case, investigated in [5]–[8].

A. Random Graph Interpretation

For a general weight matrix $W \in \mathbb{R}^n$, one can define its corresponding graph $\mathbf{G}(W)$ as a weighted directed graph on n vertices, with an edge (i, j) from vertex i to vertex j with weight W_{ji} if and only if $W_{ji} \neq 0$. As a consequence of this definition, the update equation $x(k) = Wx(k-1)$ 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 $\{x_j(k-1) : W_{ij} \neq 0\} = \{x_j(k-1) : (j, i) \text{ is an edge of } \mathbf{G}(W)\}$. Therefore, one can interpret (1) as a local iterative update over an ergodic stationary graph process. The assumption that the sequence $\{W_k : k \geq 1\}$ is a general stationary and ergodic process implies that the edges of the graphs $\mathbf{G}(W_k)$ are not necessarily independent from one another over time. In other words, the existence of an edge in the network at some time step k_1 may provide some information (in the probabilistic sense) about the existence and weights of other edges of the graph at some other time k_2.

Given a general weight matrix W, if (i, j) is an edge of $\mathbf{G}(W)$, 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*. Later in the note, we use the following lemma, the proof of which can be found in [16].

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 α_r is initial, if and only if the spectral radius of $\alpha_r[W]$ equals to one, where $\alpha_r[W]$ is the submatrix of W corresponding to the vertices in the class α_r.

B. Weak Ergodicity

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

$$x(k) = W_k \dots W_2 W_1 x(0). \tag{2}$$

As it is evident from (2), one needs to investigate the behavior of infinite products of stochastic matrices in order to check for 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 4: The sequence $\{W_k\}_{k=1}^{\infty} = W_1, W_2, \cdots$, of $n \times n$ stochastic matrices is weakly ergodic, if for all $i, j, s = 1, \cdots, n$ and all integer $p \geq 0$

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

as $k \to \infty$, 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[1] if the rows of the product matrix converge to each other, as the number of terms in the product grows. Note that the definition of weak ergodicity does not require $\lim_{k\to\infty} U^{(k,p)}$ to exist. All it requires is the convergence of the pairwise differences of all the rows to zero. This is not a matter of concern as one can show that whenever a sequence of matrices is weakly ergodic, then the limit of the infinite left products exists [17], [18].

Theorem 1: Suppose the matrix sequence $\{W_k\}_{k=0}^{\infty}$ is weakly ergodic and denote the left products by $U^{(k,p)} = W_{k+p} \cdots W_{p+1}$. Then, for all $i, s = 1, \cdots, n$ and all integers $p \geq 0$ there exist vectors $d^{(p)}$ not depending on i such that

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

as $k \to \infty$. This is called *strong ergodicity*.

Proof: 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

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

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

Therefore, weak ergodicity implies the existence of a non-negative vector d satisfying $U^{(k,0)} \to \mathbf{1}d^T$, where $\mathbf{1}$ denotes a vector with all entries equal to one. As a consequence, if one proves that the sequence $\{W_k : k \geq 1\}$ is weakly ergodic almost surely, it guarantees the convergence of (1) to consensus, with probability one. We use

[1] To be more precise, we have stated the definition of weak ergodicity *in the backward direction*.

this fact as the basis of our proofs for convergence to consensus. It is important to 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 (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.

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 $0 \le \tau(\cdot) \le 1$. A coefficient of ergodicity is said to be *proper* if

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

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

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

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

as $k \to \infty$ for a proper coefficient of ergodicity τ. We can now state the following theorem, proved in [18]:

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

$$\tau(W_m \ldots W_2 W_1) \le \prod_{k=1}^{m} \tau(W_k). \tag{3}$$

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

$$\sum_{r=1}^{\infty} \left(1 - \tau(W_{k_{r+1}} \ldots W_{k_r+1})\right) = \infty. \tag{4}$$

Proof: Suppose that there exists a strictly increasing sequence of positive integers k_r such that (4) holds. Then, the inequality $\log x \le x - 1$ implies that

$$\sum_{r=1}^{\infty} \log\left(\tau(W_{k_{r+1}} \ldots W_{k_r+1})\right) = -\infty,$$

and as a result, $\prod_{r=1}^{\infty} \tau(W_{k_{r+1}} \ldots W_{k_r+1}) = 0$. Because we assumed that τ is proper, (3) guarantees weak ergodicity of the sequence. ■

In this note, we use a coefficient of ergodicity which is defined as

$$\kappa(W) = \frac{1}{2} \max_{i,j} \sum_{s=1}^{n} |W_{is} - W_{js}|.$$

Note that $\kappa(\cdot)$ is a proper coefficient that satisfies the submultiplicative property (3).

IV. Almost Sure Convergence of the Consensus Algorithm

In this section, we prove a necessary and sufficient condition for linear dynamical system (1) to converge to consensus almost surely, when the process that generates the weight matrices $\{W_k : k \ge 1\}$ is ergodic and stationary. Our results contain the results of [5]–[8] as special cases, which simply assume an i.i.d. matrix process.

Theorem 3: Let $\{W_k : k \ge 1\} = W_1, W_2, \ldots$ denote a sequence of stochastic matrices with positive diagonals generated by an ergodic stationary process. This random sequence is weakly ergodic almost surely, if and only if $|\lambda_2(\mathbb{E}W_k)| < 1$, where λ_2 is the eigenvalue with the second largest modulus.[2]

[2]Note that $\mathbb{E}W_k$ is time-invariant because of the stationarity assumption.

Proof: First, we prove the necessity. Suppose $|\lambda_2(\mathbb{E}W_k)| = 1$. Since all weight matrices 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 [16], $|\lambda_2(\mathbb{E}W_k)| < 1$, which contradicts our assumption. Therefore, $|\lambda_2(\mathbb{E}W_k)| = 1$ implies reducibility of $\mathbb{E}W_k$. As a result, without loss of generality, one can label the vertices such that $\mathbb{E}W_k$ gets the following block triangular form

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

where each Q_{ii} is an irreducible matrix and represents the vertices in the communication class α_i. Since $\lambda_1(\mathbb{E}W_k) = |\lambda_2(\mathbb{E}W_k)| = 1$, submatrices corresponding to at least two of the classes have unit spectral radii (because of irreducibility and aperiodicity of Q_{ii}'s, the multiplicity of the unit-modulus eigenvalue of each one cannot be more than one). 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. Since all the weight matrices are non-negative, W_k has the same type (zero block pattern) as does $\mathbb{E}W_k$ for all time k with probability one. Therefore, $U^{(k,0)} = W_k \cdots W_2 W_1$ has two orthogonal rows almost surely for any k,[3] which means that there are initial conditions for which random discrete-time dynamical system (1) reaches consensus with probability zero. Since weak ergodicity of $\{W_k\}$ is a subset of convergence of (1) to consensus, the random sequence of weight matrices is weakly ergodic almost never.

Our proof of the reverse implication is based on Lemma 1. When $|\lambda_2(\mathbb{E}W_k)| < 1$, Lemma 2 implies that $\mathbf{G}(\mathbb{E}W_k)$ has exactly one initial class. In other words, there exists i such that for all $j \neq i$ there exists a sequence $i = j(1), j(2), \ldots, j(s_j) = j$ for which $(\mathbb{E}W_k)_{j(q+1),j(q)} > 0$. Equivalently, there exists a path of length $s_j - 1$ from some node i to any other node j in the expected graph of the network. As a result, there exists $\epsilon > 0$ such that

$$\mathbb{P}\left((W_k)_{j(q+1),j(q)} > \epsilon\right) > 0 \text{ for all } q = 1, 2, \ldots, s_j - 1.$$

for all vertices j. Hence, Lemma 1 implies

$$\mathbb{P}\left((W_k)_{j(q+1),j(q)} > \epsilon \text{ infinitely often}\right) = 1 \quad 0 < q < s_j$$

for all $j \neq i$. Therefore, the countable intersection of these events also occurs with probability one. As a result, there exists a deterministic time T for which

$$\mathbb{P}\left(\delta(W_T \ldots W_2 W_1) > \zeta\right) > 0$$

for some $\zeta > 0$, where $\delta(W) = \max_j(\min_i W_{ij})$. In other words, there exists a deterministic time T, for which all the entries of at least one column of the matrix product $W_T \ldots W_2 W_1$ is bounded away from zero with positive probability. Now, once again the ergodicity and stationarity of the sequence $\{W_k : k \ge 1\}$ implies that such an event occurs infinitely often almost surely, i.e.,

$$\mathbb{P}\left(\delta(W_{(r+1)T} \ldots W_{rT+1}) > \zeta \text{ for infinitely many } r\right) = 1.$$

Therefore, by defining $k_r = rT$, we have

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

[3]This is because of the fact that this event is the intersection of countably many unit-measure events.

Since $\delta(W) \leq 1 - \kappa(W)$ for any stochastic matrix W,

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

which is exactly (4), the sufficient condition for weak ergodicity. Therefore, the sequence is weakly ergodic almost surely. ∎

Theorem 3 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 linear dynamical system (1).

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

Proof: Based on Theorem 3, $|\lambda_2(\mathbb{E}W_k)| < 1$ guarantees weak ergodicity with probability one, and as a result, the event of asymptotic consensus occurs almost surely, since it is a superset of the weak ergodicity event. The reverse implication was shown in the proof of Theorem 3. ∎

Therefore, $|\lambda_2(\mathbb{E}W_k)| < 1$ provides a necessary and sufficient condition for almost sure asymptotic consensus in (1) when the weight matrices in the sequence (and hence their corresponding graphs) are generated by an ergodic stationary process. This result is a generalization of our results in [8], which provides a similar criterion for the i.i.d. case.

Remark 1: The ergodicity of the graph process can be interpreted as the property that the time averages over the process are equal to the ensemble averages almost surely. In other words, when the expected graph of the network contains a directed spanning tree, that is, when $|\lambda_2(\mathbb{E}W_k)| < 1$, then there exists a time sequence $\{k_r : r \geq 1\}$ such that collection of graphs $\{\mathbf{G}(W_{k_r+1}), \cdots, \mathbf{G}(W_{k_{r+1}})\}$ are *jointly connected* almost surely, and therefore, asymptotic consensus is guaranteed with probability one [1].

Remark 2: Theorem 3 states that depending on the second largest eigenvalue modulus of the expected weight matrix, weak ergodicity occurs with either probability 1 or 0, i.e., weak ergodicity is a trivial event. This was to be expected, as the event $B = \{W_1, W_2, \ldots \text{ is weakly ergodic}\}$ satisfies $B = \varphi B$ and therefore, is invariant, i.e., $\mathbb{P}(B \Delta \varphi B) = 0$. Hence, as a consequence of ergodicity of φ, such an event is trivial.

In order to illustrate the results presented in this section we provide a simple example.

Example 1: Consider a graph on n vertices with its potential undirected edges numbered 1 through $n(n-1)/2$. We assume that at time k, the edge e exists in the graph with weight $1/n$ if and only if the e-th entry of the random vector z_k is non-negative, where z_k is generated by an autoregressive process of order one. More precisely, for any $i \neq j$ and the edge $e = (i, j)$,

$$(W_k)_{ij} = \frac{1}{n}\mathbb{I}_{\{z_{k_e} \geq 0\}} \quad (5)$$

$$z_k = \gamma z_{k-1} + (1-\gamma)\epsilon_k \mathbf{1} \quad (6)$$

where $\gamma \in [0, 1)$ is a scalar, $z_0 \sim \mathcal{N}(0, \frac{1-\gamma}{1+\gamma}\mathbf{1}\mathbf{1}^T)$, and $\{\epsilon_k\}$ are a sequence of i.i.d. unit normal random variables independent from z_0. The diagonal elements of W_k are defined such that the matrix is stochastic. Note that (6) defines z_k as simply a convex combination of its value at the previous time step and an independent noise term ϵ_k. In other words, not only the existence of an edge at time k is correlated with the existence of other edges at the same time step (due to the common noise term), it is also correlated with the realization of the random vector z at all other times (as long as $\gamma \neq 0$).

Equations (5) and (6) together define a weight matrix process $\{W_k : k \geq 1\}$. Note that the assumption $\gamma \neq 1$ and the given distribution for z_0 guarantee that the process is ergodic stationary. Therefore, we can apply Theorem 3. It is easy to verify that $\mathbb{P}(z_{k_e} \geq 0) = 1/2$, and as a consequence, $\mathbb{E}W_k = \frac{1}{2}I + \frac{1}{2n}\mathbf{1}\mathbf{1}^T$, which is irreducible. Thus, the given sequence is weakly ergodic almost surely.

V. Asymptotic Distribution of the Consensus Value

As stated in the previous section, the ergodic stationary matrix process $\{W_k : k \geq 1\}$ of stochastic matrices is weakly ergodic almost surely, if and only if, $|\lambda_2(\mathbb{E}W_k)| < 1$. In other words, if the expected weight matrix has a unique unit-modulus eigenvalue, then the left products of the stochastic matrices converge to rank one matrices with probability one. Also, Theorem 1 implies that for left products, weak and strong ergodicity are equivalent. Therefore, whenever $|\lambda_2(\mathbb{E}W_k)| < 1$, then there exists a random non-negative vector d satisfying $\mathbf{1}^T d = 1$ such that $U^{(k,0)} \to \mathbf{1}d^T$ almost surely, as $k \to \infty$. As a consequence, the asymptotic consensus value of linear dynamical system (1) is the random variable $x^* = d^T x(0)$.

A natural question to ask is whether one can determine the distribution of this random consensus value. Answering such a question is far from trivial, even for the case that the weight matrices are i.i.d. In this section, we investigate a special case, for which one can compute the distribution analytically. More specifically, we provide a necessary and sufficient condition for the random consensus value to be degenerate, i.e., a condition under which the consensus algorithm in (1) converges to a deterministic constant almost surely. We also compute the mean and the variance of the random consensus value x^* for the case of i.i.d. weight matrices.

Theorem 5: Let $\{W_k : k \geq 1\} = W_1, W_2, \ldots$ denote a sequence of stochastic matrices with positive diagonals generated by an ergodic stationary process with $|\lambda_2(\mathbb{E}W_k)| < 1$. Also consider the deterministic vector y satisfying $\mathbf{1}^T y = 1$. Then, the left product $U^{(k,0)} = W_k \ldots W_1$ converges to $\mathbf{1}y^T$ almost surely, if and only if y is a left eigenvector of W_k corresponding to the unit eigenvalue, with probability one.

A special case of interest is when all the matrices that can appear with positive probability are doubly-stochastic. In this special case, $y = (1/n)\mathbf{1}$ is a common left eigenvector of all the matrices in the sequence. Theorem 5 states that this is a necessary and sufficient condition for the limiting consensus value to be equal to the average of the initial values $x(0)$ almost surely. In such a case, we say the linear dynamical system reaches an *average consensus* with probability one.

Proof: The sufficiency proof is trivial and quite well-known [4], [17]. Since $|\lambda_2(\mathbb{E}W_k)| < 1$, Theorem 3 guarantees that the product (2) converges to a rank one matrix with probability one, i.e., $W_k \ldots W_2 W_1 \to \mathbf{1}d^T$ almost surely, for some random vector d. In the case that all the weight matrices share the same left eigenvector y corresponding to the unit eigenvalue with probability one,[4] then any product $U^{(k,0)}$ has also the same left eigenvector, and so does its limit as $k \to \infty$. Therefore, $W_k \ldots W_2 W_1 \to \mathbf{1}y^T$ almost surely, or in other words, $\mathbb{P}(d = y) = 1$.

To prove the reverse implication assume $|\lambda_2(\mathbb{E}W_k)| < 1$. Also, suppose that there exists a non-random stochastic vector y such that $U^{(k,0)} = W_k \ldots W_1 \to \mathbf{1}y^T$ almost surely. Since the sequence $\{W_k : k \geq 1\}$ is stationary, $U^{(k,1)} = W_k \ldots W_2$ should also converge to $\mathbf{1}y^T$ almost surely. Combining the above, we have,

$$U^{(k,0)} = U^{(k,1)} W_1 \to \mathbf{1}y^T W_1 \quad \textit{almost surely}.$$

As a consequence, $\mathbb{P}(y^T W_1 = y^T) = 1$, which means that all the weight matrices have the same common left eigenvector y corresponding to the unit eigenvalue, with probability one. ∎

[4]Note that since $|\lambda_2(\mathbb{E}W_k)| < 1$, there is only one such vector y.

Remark 3: Stationarity of the matrix process plays a crucial role in proving the necessity part of the above theorem. In fact, if the weight matrix process is not stationary, having a common left eigenvector corresponding to the unit eigenvalue is not necessary anymore. As an example, consider the following two stochastic matrices:

$$W_1 = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} \quad , \quad W_2 = \begin{bmatrix} \frac{1}{5} & \frac{4}{5} \\ \\ \frac{3}{5} & \frac{2}{5} \end{bmatrix}.$$

It is easy to verify that neither matrix is doubly stochastic. However, the product $W_2 W_1$ is a doubly stochastic matrix. Therefore, if matrices W_k that appear in the sequence are doubly stochastic for $k \geq 3$, the linear dynamical system converges to the average consensus, even though W_1 and W_2 are not doubly stochastic.

A. Computing First and Second Moments

As stated before, computing the distribution of the consensus value in terms of the distribution of the weight matrices is not an easy problem in general. However, one can compute the moments of the random rank one matrix that the left products $U^{(k,0)}$ converge to. In what follows, we compute the mean and the variance of the consensus value for the case of i.i.d. weight matrices.

Assuming $|\lambda_2(\mathbb{E}W_k)| < 1$, we know that $W_k \dots W_2 W_1 \to \mathbf{1} d^T$ almost surely, for some random stochastic vector d. By taking expectation and applying the dominated convergence theorem [14],

$$\mathbb{E}(W_k \dots W_2 W_1) \to \mathbb{E}(\mathbf{1} d^T),$$

which implies $[\mathbb{E}W_1]^k \to \mathbf{1}(\mathbb{E}d^T)$, due to independence. Therefore, by the Perron-Frobenius theorem, $\mathbb{E}d$ is simply equal to the normalized left eigenvector $\mathbb{E}W_k$ corresponding to the unit eigenvalue. [5]

In order to compute the variance, first note that

$$\frac{1}{n}(W_k \dots W_1)^T (W_k \dots W_1) \to dd^T \quad a.s.,$$

which can be rewritten as

$$\begin{aligned} &\tfrac{1}{n} vec[(W_k \dots W_1)^T (W_k \dots W_1)] = \\ &\quad = \tfrac{1}{n}(W_1^T \otimes W_1^T)(W_2^T \otimes W_2^T) \dots (W_k^T \otimes W_k^T) vec(I_n) \\ &\quad \longrightarrow vec(dd^T) \quad \textit{almost surely}, \end{aligned}$$

where vec is the vectorization operator, $\otimes$ denotes the Kronecker product and I_n is the identity matrix of size n. By applying the dominated convergence theorem once again, and using the assumption that the weight matrices are independent, we get

$$\frac{1}{n}\left[\mathbb{E}(W_1^T \otimes W_1^T)\right]^k vec(I_n) \longrightarrow \mathbb{E}[vec(dd^T)] = \mathbb{E}(d \otimes d).$$

Hence, by the Perron-Frobenius theorem,

$$\mathbb{E}(d \otimes d) = \frac{1}{n} v_1\left(\mathbb{E}\left[W_k \otimes W_k\right]\right)\left(\mathbf{1}_{2n}^T vec(I_n)\right) = v_1\left(\mathbb{E}\left[W_k \otimes W_k\right]\right),$$

where $v_1(\cdot)$ denotes the normalized left eigenvector corresponding to the unit eigenvalue. Therefore, the covariance matrix of the random vector d satisfies

$$vec(cov(d)) = v_1\left(\mathbb{E}\left[W_k \otimes W_k\right]\right) - v_1\left(\mathbb{E}W_k\right) \otimes v_1\left(\mathbb{E}W_k\right).$$

By combining all the above, we can compute the conditional first and second moments of the random consensus value $x^* = d^T x(0)$ in terms of the moments of the weight matrices. The mean and variance conditional on the initial condition $x(0)$ are respectively given by

$$\begin{aligned} \mathbb{E}\ x^* &= x(0)^T v_1(\mathbb{E}W_k) \\ \text{var}(x^*) &= \left[x(0) \otimes x(0)\right]^T v_1(\mathbb{E}\left[W_k \otimes W_k\right]) - \left[x(0)^T v_1(\mathbb{E}W_k)\right]^2 \end{aligned}$$

It is easy to verify that the variance is equal to zero, whenever all the weight matrices share the same left eigenvector corresponding to the unit eigenvalue, and therefore, generating a consensus value with a degenerate distribution, as shown in Theorem 5.

[5]The assumption $|\lambda_2(\mathbb{E}W_k)| < 1$ guarantees that such an eigenvector is unique.

VI. Conclusions

In this note, we proved a necessary and sufficient condition for almost sure consensus for a general weighted and directed stationary and ergodic random graph process. We showed that the 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. Our results contain the cases of i.i.d. and (ergodic and stationary) Markovian graph processes as special cases. We also showed that, given the assumptions of ergodicity and stationarity, in order to have an asymptotic consensus value with a degenerated distribution, it is both necessary and sufficient for all the weight matrices to share a common left eigenvector corresponding to the unit eigenvalue. Finally, we provided expressions for the mean and the variance of the consensus value.

Acknowledgment

The authors would like to thank Asu Ozdaglar and Frank Schofheide for helpful comments and discussions.

References

[1] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.

[2] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 1984.

[3] S. Boyd, A. Gosh, B. Prabhakar, and D. Shah, "Gossip algorithms: Design, analysis and applications," in *Proceedings of IEEE INFOCOM 2005*, vol. 3, Miami, Mar. 2005, pp. 1653–1664.

[4] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proceedings of the 4th International Conference on Information Processing in Sensor Networks*, Apr. 2005, pp. 63–70.

[5] Y. Hatano and M. Mesbahi, "Agreement over random networks," *IEEE Transactions on Automatic Control*, vol. 50, no. 11, pp. 1867–1872, 2005.

[6] C. W. Wu, "Synchronization and convergence of linear dynamics in random directed networks," *IEEE Transactions on Automatic Control*, vol. 51, no. 7, pp. 1207–1210, July 2006.

[7] M. Porfiri and D. J. Stilwell, "Consensus seeking over random weighted directed graphs," *IEEE Transactions on Automatic Control*, vol. 52, no. 9, pp. 1767–1773, sep 2007.

[8] A. Tahbaz-Salehi and A. Jadbabaie, "A necessary and sufficient condition for consensus over random networks," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 791–795, Apr 2008.

[9] F. Fagnani and S. Zampieri, "Randomized consensus algorithms over large scale networks," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 634–649, May 2008.

[10] G. Picci and T. Taylor, "Almost sure convergence of random gossip algorithms," in *Proceedings of the 46th IEEE Conference on Decision and Control*, New Orleans, LA, dec 2007, pp. 282–287.

[11] 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.

[12] ——, "On products of random stochastic matrices," in *Random matrices and their applications, Proceedings of American Mathematical Society Summer Research Conference*, vol. 50, Providence, 1986, pp. 199–213.

[13] R. L. Dobrushin, "Central limit theorem for non-stationary Markov chains, I, II," *Theory of Probability and its Applications*, vol. 1, pp. 65–80, 329–383, 1956, (English Translation).

[14] R. Durrett, *Probability: Theory and Examples*, 3rd ed. Belmont, CA: Duxbury Press, 2005.

[15] I. P. Cornfeld, S. V. Fomin, and Y. G. Sinai, *Ergodic theory*, ser. Grundlehren der mathematischen Wissenschaften. New York, NY: Springer, 1982, vol. 245.

[16] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. New York: Academic Press, 1979.

[17] S. Chatterjee and E. Seneta, "Towards consensus: Some convergence theorems on repeated averaging," *Journal of Applied Probability*, vol. 14, no. 1, pp. 89–97, Mar. 1977.

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