

Tight Bounds on the Convergence Rate of Generalized Ratio Consensus Algorithms

Balázs Gerencsér  and László Gerencsér 

Abstract—The problems discussed in this article are motivated by general ratio consensus algorithms, introduced by Kempe *et al.* in 2003 in a simple form as the push-sum algorithm, later extended by Bénézit *et al.* in 2010 under the name weighted gossip algorithm. We consider a communication protocol described by a strictly stationary, ergodic, sequentially primitive sequence of nonnegative matrices, applied iteratively to a pair of fixed initial vectors, the components of which are called values and weights defined at the nodes of a network. The subject of ratio consensus problems is to study the asymptotic properties of ratios of values and weights at each node, expecting convergence to the same limit for all nodes. The main results of this article provide upper bounds for the rate of the almost sure exponential convergence in terms of the spectral gap associated with the given sequence of random matrices. It will be shown that these upper bounds are sharp. Our results complement previous results of Picci and Taylor in 2013 and Iutzeler *et al.* in 2013.

Index Terms—Asynchronous communication, convergence, communication networks, distributed computing, estimation, iterative algorithms, multiagent systems, random processes, spectral gap.

I. INTRODUCTION

A. Setup and the Ratio Consensus Algorithm

THE problems discussed in this article are motivated by the study of general ratio consensus algorithms, introduced in [1] in a simple form as the push-sum algorithm, and later extended in [2] under the name weighted gossip algorithm for solving a class of distributed computation problems. The

Manuscript received April 29, 2020; revised December 8, 2020; accepted February 27, 2021. Date of publication March 19, 2021; date of current version March 29, 2022. This work was supported in part by the Ministry of Innovation and Technology NRD Office (National Research, Development and Innovation Office) by the projects under Grant PD 121107, Grant KH 126505, and Grant FK 135711, and in part by the NRD Office within the framework of the Autonomous Systems National Laboratory Program, and by the European Commission through the H2020 project EPIC under Grant 739592. Recommended by Associate Editor W. X. Zheng. (Corresponding author: Balázs Gerencsér.)

Balázs Gerencsér is with the Alfréd Rényi Institute of Mathematics, Eötvös Loránd Research Network (ELKH), 1053 Budapest, Hungary, and also with the Department of Probability and Statistics, Eötvös Loránd University, 1117 Budapest, Hungary (e-mail: gerencser.balazs@renyi.hu).

László Gerencsér is with SZTAKI, the Institute for Computer Science and Control, Eötvös Loránd Research Network (ELKH), 1111 Budapest, Hungary (e-mail: gerencser.laszlo@sztaki.hu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TAC.2021.3067629>.

Digital Object Identifier 10.1109/TAC.2021.3067629

algorithm is designed to solve a consensus problem over a network of agents, based on asynchronous communication. The objective of the consensus can be expressed in its simplest way as to achieve the average of certain values given at each node. The original problem formulation and the algorithm has been adapted to model a number of real-life situations such as platooning, sensor networks, or smart grids (see [3] and [4]).

Various relaxations and extensions of the baseline model were proposed in the literature. A nice application of the push-sum algorithm for computing the eigenvectors of a large symmetric matrix, corresponding to the adjacency matrix of an undirected graph, was given in [5]. Another application is distributed convex optimization see [6]. A general class of solvable consensus problems for the distributed function computation was introduced in [7].

The basic setup for this class of methods is a communication network represented by a directed graph $G = (V, E)$, to each node i of which a pair of real numbers x^i and $w^i \geq 0$ is associated, such that not all of the w^i -s are 0. They are often called the values and the weights. The problem is then to compute the ratio $\sum_i x^i / \sum_i w^i$, at all nodes, using only local interactions allowed by $G = (V, E)$ in an asynchronous manner. In the special case, when $w^i = 1$ for all nodes, the problem reduces to the average consensus problem.

A convenient illustration of the aforementioned problem is the following: x^i unit of some chemical is dissolved in a solvent of $w^i \geq 0$ units leading to a solution with concentration x^i/w^i at the node i . The problem equivalent to the aforementioned one is then to compute the concentration of the grand total, defined as $\sum_i x^i / \sum_i w^i$, using only local transfers allowed by $G = (V, E)$ in an asynchronous manner.

Let $|V| = p$ and let $x_0 = x = (x^1, \dots, x^p)^\top$ and $w_0 = w = (w^1, \dots, w^p)^\top$ denote the vectors of initial values and weights, respectively, at time 0, assuming $w \geq 0, w \neq 0$. We update both the values and weights successively as follows. Let x_{n-1} and w_{n-1} denote the p -vector of values and weights, respectively, at time $n - 1$. Select a directed edge $f_n = (i, j) \in E$ randomly, representing the communicating pair at time n . Then, the sender, node i , initiates a transactions by sending a fraction, say $\alpha^{j,i}$ with $0 < \alpha^{j,i} < 1$, of his/her values and weights to the receiver, node j . It is initially assumed that the sequence of edges (f_n) is i.i.d., with the probability of choosing an edge $f = (i, j)$ being denoted by q_{ij} .

In the context of the aforementioned illustration via elementary chemistry, the algorithm is equivalent to mixing a fraction of

the solution at the node i into the current solution at the node j . It is then expected that in the limit we get solutions with identical concentrations at each node.

When setting $\alpha^{ji} = 1/2$ for all edges the aforementioned algorithm is the celebrated push-sum method. The dynamics of the algorithm can be formally described by the equations

$$x_n = A_n x_{n-1} \quad \text{and} \quad w_n = A_n w_{n-1} \quad (1)$$

for $n \geq 1$, where A_n is a $p \times p$ random matrix obtained from the identity matrix by modifying its i th column as follows:

$$A_n^{ii} = 1 - \alpha^{ji} \quad A_n^{ji} = \alpha^{ji} \quad A_n^{ki} = 0 \quad \text{for } k \neq i, j. \quad (2)$$

The aforementioned problem can be modified by allowing packet losses (see [3]). When a packet loss occurs along the edge from i to j , denoted by (j, i) , the content of the node j is not changed. Packet losses are assumed to occur randomly and independently. The functionality of the network at time n is described by a collection of indicators $\rho_n(f)$, $f \in E$: $\rho_n(f) = 1$ if the edge f fails at time n , otherwise $\rho_n(f) = 0$. The probability of failure along edge f is $0 \leq r_f < 1$ at any time so that $P(\rho_n(f) = 1) = r_f$. With these notations, assuming $f_n = (j, i)$, the matrix A_n will have the following structure with a single, possibly nonzero off-diagonal element in the positions (j, i) :

$$\begin{pmatrix} 1 & 0 & \dots & & & & 0 \\ 0 & 1 & \dots & & & & 0 \\ \vdots & \vdots & \ddots & & & & \\ & & & 1 - \alpha^{ji} & & & 0 \\ & & & & \ddots & & \\ & & & (1 - \rho_n(f_n))\alpha^{ji} & & & 1 \\ & & & & & \ddots & \vdots \\ 0 & & & & & \dots & 1 \end{pmatrix}. \quad (3)$$

We note in passing that the coordinates of vectors and the elements of matrices will be indicated by superscripts, while their dependence on the discrete time n will be indicated by subscripts.

B. Generalized Framework

The aforementioned form of the push-sum or weighted gossip algorithm has a natural extension reflecting the possibility of certain schedules in choosing the sequence of interacting pairs of agents, as in the case of geographic gossip, randomized path averaging, or one-way averaging [2], [3], [8].

In addition, we may consider a significantly broader class of matrices, allowing much more complex network dynamics. Technically speaking, we consider a strictly stationary, ergodic sequence of $p \times p$ random matrices with nonnegative entries (A_n) , $n \geq 1$. Let $x, w \in \mathbb{R}^p$ denote a pair of initial vectors, such that $w \geq 0$, $w \neq 0$. Our objective is to study the asymptotic properties of the ratios

$$e_i^\top A_n A_{n-1} \dots A_1 x / e_i^\top A_n A_{n-1} \dots A_1 w, \quad i = 1, \dots, p \quad (4)$$

where e_i is the unit vector with a single 1 in its i th coordinate.

For a start, we provide a brief summary of two classical results on products of strictly stationary, ergodic sequences of random

matrices, and recapitulate and extend a relevant application as Theorem 8. The key results of this article are stated as Theorems 12, 14, 16, and 19, extending previous results on the almost sure (a.s.) exponential convergence in the context of ratio consensus such as given in [9], in particular, providing upper bounds for the a.s. exponential convergence rate in terms of spectral gaps associated with stationary sequences of matrices. It will be shown that these upper bounds are sharp in Theorem 21, thus solving an open problem formulated in the conclusion of one of the fundamental papers [2] under very general conditions.

The proofs are based on the careful analysis of random products $M_n := A_n A_{n-1} \dots A_1$ for a random sequence of non-negative matrices using Oseledec's theorem. The application of results in the theory of products of random matrices in the context of consensus algorithms was previously initiated and elaborated in [10] for the case of linear gossip algorithms with pairwise, bidirectional, symmetric communication. While we rely partially on the same mathematical methodology, the range of communication protocols that we consider is significantly broader, in particular, we consider ratio consensus algorithms.

Our work complements and extends the result of [11] in which an upper bound for the rate (or the exponent) of a.s. exponential convergence of a (sampled) weighted gossip algorithm was derived; see Remark 20. A more recent work to mention is [12], in which an ingenious device was proposed, using auxiliary variables, to solve the average consensus problem with column stochastic matrices via a *linear* asynchronous gossip algorithm, proving exponential mean square stability with an explicit upper bound for the rate. It turns out that our results are applicable to prove a.s. exponential convergence, as well.

Finally, we point out that exponential rate of convergence of ratio consensus algorithms can be established, under appropriate conditions, with purely deterministic methods; however, we are unaware of any tight upper bounds for the rate of convergence in this context. The key idea is to establish exponentially decaying upper bounds for the error of rank-1 approximations of the form $u_n \mathbf{1}^\top$ of products of column stochastic matrices, belonging to a given set, using pure linear algebra. This program is nicely followed through in [4] and the follow-up paper [13], relying on a result of [14]. A specific novelty of these papers is that they develop a ratio consensus algorithm with arbitrary bounded delays. Similar results, with focus on rank-1 approximations, are formulated in [15, Lemma 2 and Prop. 1], although with no proofs. The relevance of rank-1 approximations is also clarified in the stochastic settings (see [16]) pointed out also by the fundamental approximate equations (39) and (40) of this article. It is worth noting that in all the aforesaid papers, the communication matrices are assumed to be column stochastic.

This article is organized as follows. Sections II–VI are devoted to the description of the subject matter and the main results of this article with minimal technical details; starting with two sections presenting a few preliminary technicalities, a section on normalized products, a section with the statements and interpretations of the main results, followed by a brief section on push-sum algorithms. In Sections VII and VIII, we

elaborate on the major mathematical details; in Section VII, we describe the essential fabric of the proofs of the main theorems, while in Section VIII, an interlude on the connection between spectral gap and Birkhoff's contraction coefficient is added. A brief discussion is given in Section IX, and finally, Section X concludes this article. Relevant but minor technical details will be given in the Appendices. Altogether we intend to give a self-contained presentation of the subject matter and of the background material.

II. TECHNICAL PRELIMINARIES

For the formulation of our results, we recall two basic facts on the product of random matrices.

Proposition 1 (Fürstenberg and Kesten's theorem, [17]): Let $(A_n), n \geq 1$ be a strictly stationary, ergodic process of $p \times p$ random matrices over a complete probability space (Ω, \mathcal{F}, P) such that $\mathbb{E} \log^+ \|A_1\| < \infty$. Then, the a.s. limit

$$\lambda_1 = \lim_{n \rightarrow \infty} \frac{1}{n} \log \|A_n A_{n-1} \dots A_1\| < \infty \quad (5)$$

exists and it is equal to

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \log \|A_n A_{n-1} \dots A_1\| \\ &= \inf_n \frac{1}{n} \mathbb{E} \log \|A_n A_{n-1} \dots A_1\|. \end{aligned} \quad (6)$$

Note that we may have $\lambda_1 = -\infty$.

A more refined asymptotic characterization of $A_n A_{n-1} \dots A_1$ is given by Oseledec's theorem. To appreciate the novelty and power of this theorem, we make a brief elementary detour in the field of Lyapunov exponents; see [18]. Let $(A_n), n \geq 1$ be a fixed sequence of $p \times p$ matrices. For any $x \in \mathbb{R}^p$, define the Lyapunov exponent of x with respect to (w.r.t.) (A_n) as

$$\lambda(x) := \limsup_{n \rightarrow \infty} \frac{1}{n} \log |A_n A_{n-1} \dots A_1 x|.$$

Next, for any extended real number $-\infty \leq \mu \leq +\infty$, define the set

$$L_\mu = \{x \in \mathbb{R}^p : \lambda(x) \leq \mu\}. \quad (7)$$

It is easily seen that L_μ is a linear subspace of \mathbb{R}^p and for $\mu < \mu'$, we have $L_\mu \subseteq L_{\mu'}$. It is also readily seen that L_μ is continuous from the right: if $x \in L_{\mu_j}$ for a sequence of μ_j -s such that μ_j tend to μ from above, then we have also $x \in L_\mu$. Since there can be only a finite number of strictly descending subspaces, it follows that there is a finite number of possible values of the Lyapunov exponents, $+\infty \geq \mu_1 > \mu_2 > \dots > \mu_q \geq -\infty$, such that

$$\mathbb{R}^p = L_{\mu_1} \supsetneq L_{\mu_2} \supsetneq \dots \supsetneq L_{\mu_q} \supsetneq \{0\} =: L_{\mu_{q+1}}, \quad (8)$$

where L_μ is a piecewise constant function of μ with points of discontinuity exactly at μ_i . Thus, for $\mu_{r-1} > \mu \geq \mu_r$, we have $L_\mu = L_{\mu_r}$ for $2 \leq r \leq q$, and for $\mu_q > \mu$, we have $L_\mu = \{0\}$. It follows that for $1 \leq r \leq q$

$$x \in L_{\mu_r} \setminus L_{\mu_{r+1}} \text{ implies } \lambda(x) = \mu_r. \quad (9)$$

Let the dimension of L_{μ_r} be denoted by i_r , with $1 \leq r \leq q+1$ (with $i_{q+1} = 0$). Then, the codimension of $L_{\mu_{r+1}}$ relative to L_{μ_r} is $i_r - i_{r+1}$, which can be interpreted as the multiplicity of the Lyapunov exponent μ_r . Accordingly, we define the full

spectrum of Lyapunov exponents $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, allowing the values $\pm\infty$, by setting for $1 \leq i \leq p$

$$\lambda_i = \mu_r \text{ if } i_r \geq i > i_{r+1}. \quad (10)$$

If $(A_n) = (A_n(\omega))$ is the realization of a strictly stationary ergodic process, then the above observations can be extended to the following fascinating result, stated first in [19], and restated and proved under weaker condition in [20]:

Proposition 2 (Oseledec's theorem): Assume that (A_n) is a strictly stationary ergodic process of $p \times p$ matrices such that $\mathbb{E} \log \|A_1\|^+ < \infty$. Then, there exists a subset $\Omega' \subset \Omega$ with $P(\Omega') = 1$ such that for all $\omega \in \Omega'$ and for any $x \in \mathbb{R}^p$, the following limit exists:

$$\lambda(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \log |A_n A_{n-1} \dots A_1 x|. \quad (11)$$

Moreover the Lyapunov exponents $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, possibly taking the value $-\infty$, do not depend on $\omega \in \Omega'$. Accordingly, μ_r and i_r for $1 \leq r \leq q$ do not depend on $\omega \in \Omega'$ either. The mapping $\omega \mapsto L_{\mu_r}(\omega)$ is measurable from Ω to the Grassmanian manifold of linear subspaces of dimension i_r . In addition, we have with $M_n = A_n A_{n-1} \dots A_1$

$$M^* = \lim (M_n^T M_n)^{1/2n} \text{ a.s.} \quad (12)$$

From the proof given in [20], it follows that taking a singular value decomposition of M_n as

$$M_n = U_n \Sigma_n V_n \quad (13)$$

where U_n and V_n are orthonormal matrices, and Σ_n is diagonal with entries $\sigma_n^1 \geq \sigma_n^2 \geq \dots \geq \sigma_n^p \geq 0$, we have

$$\lambda_k = \lim_{n \rightarrow \infty} \frac{1}{n} \log \sigma_n^k \text{ a.s. for } k = 1, \dots, p. \quad (14)$$

Therefore, we have, with $o(1)$ denoting a sequence of random variables tending to 0 a.s. as n tends to ∞

$$\Sigma_n = \text{diag}(e^{(\lambda_k + o(1))n}). \quad (15)$$

Surprisingly, the orthonormal matrices V_n will also converge a.s. in a restricted sense. Allowing the possibility of multiplicity of Lyapunov exponents consider a fixed μ_r and define $I_r = \{i : \lambda_i = \mu_r\}$, and let $SV_n^{I_r}$ denote the subspace spanned by the rows of V_n with indices in I_r . Then, we have a.s. $\lim SV_n^{I_r} = SV^{I_r}$ for some random subspace SV^{I_r} . We note in passing that this technical result immediately implies the existence of the a.s. limit in (12).

In particular, if $\lambda_1 > \lambda_2$, then for the first row of V_n , denoted by v_n^1 , we have

$$\lim_{n \rightarrow \infty} v_n^1 = v^1 \quad (16)$$

with probability 1 (w.p.1) for some random v^1 . In fact, it was proved by Ragunathan in [20, Lemma 5] that for any $\varepsilon > 0$

$$v_n^1 - v^1 = O(e^{-(\lambda_1 - \lambda_2 + o(1))n}) \text{ w.p.1.} \quad (17)$$

Writing

$$M_n = u_n^1 v^1 \sigma_n^1 + \sum_{k=2}^p u_n^k v^k \sigma_n^k \quad (18)$$

it follows by straightforward calculations that

$$M_n = u_n^1 v^1 \sigma_n^1 + O(e^{(\lambda_2 + o(1))n}). \quad (19)$$

A rank-1 approximation for the product of a strictly stationary, ergodic sequence of column stochastic matrices has been derived in [16, Th. 3] using different techniques.

A nice corollary of Oseledec's theorem, obtained by a straightforward application of Fubini's theorem, is that for all $x \in \mathbb{R}^p$, except for a set of the Lebesgue-measure zero, we have

$$\lambda_1 = \lim_{n \rightarrow \infty} \frac{1}{n} \log |A_n A_{n-1} \dots A_1 x| \quad \text{w.p.1.} \quad (20)$$

In the special case when $A_n = A$ for all n , arranging the eigenvalues of A , say ν_i , according to their absolute values in a nonincreasing order, we have $\lambda_i = \log |\nu_i|$.

III. SEQUENTIALLY PRIMITIVE NONNEGATIVE MATRIX PROCESSES

In the next section, we present the extension of a result of [21] on the asymptotic behavior of normalized products

$$A_n A_{n-1} \dots A_1 x / \mathbf{1}^T A_n A_{n-1} \dots A_1 x \quad (21)$$

where $\mathbf{1}$ is a p -vector all coordinates of which are 1. For the generalization of [21, Th. 1], the extension of the notion of primitivity for a class of matrices and stochastic processes will be needed. For a nice introduction and motivation on this topic, see [22].

Let $\mathcal{A} = \{A_1, \dots, A_m\}$ be a finite family of $p \times p$ matrices with nonnegative entries. We may then ask if there is a product of these matrices (with repetitions permitted) which is strictly positive. The following definition is essentially given in [22].

Definition 3: A family \mathcal{A} of nonnegative $p \times p$ matrices is called primitive if there is at least one strictly positive product of matrices of this family.

Let $A^0 := \gamma(A)$ denote the $(0,1)$ matrix having a 1 in a position exactly if in that position A has a positive element. Define the set of matrices $\mathcal{A}^0 = \{\gamma(A) : A \in \mathcal{A}\}$. Then, obviously, \mathcal{A} is primitive if and only if \mathcal{A}^0 is primitive. The definition and claim extends to infinite sets of matrices \mathcal{A} .

We will now extend the definition to stationary processes of nonnegative random matrices. A matrix is called allowable, if it has no zero row or zero column. It is called row allowable if it has no zero row.

Definition 4: A strictly stationary process of nonnegative allowable random matrices $(A_n), n \geq 1$, is called (forward) sequentially primitive if $M_\tau = A_\tau A_{\tau-1} \dots A_1$ is strictly positive for some finite stopping time τ with probability 1. For any $n \geq 1$, we define the (forward) index of sequential primitivity as

$$\psi_n = \min\{\psi \geq 1 : A_{n+\psi-1} A_{n+\psi-2} \dots A_n > 0\}. \quad (22)$$

Since by assumption A_n is row allowable, we will have $M_n > 0$ with strict inequality for all $n \geq \psi_1$. It is also clear that a stationary process of nonnegative random matrices $(A_n), n \geq 1$, is (forward) sequentially primitive if and only if the stochastic process $(A_n^0), n \geq 1$, is (forward) sequentially primitive.

The definition extends to two-sided processes. In this case, we may also define the concept of backward sequential primitivity, and the backward index of sequential primitivity as

$$\rho_n = \min\{\rho \geq 1 : A_n A_{n-1} \dots A_{n-\rho+1} > 0\}. \quad (23)$$

Lemma 5: A two-sided strictly stationary sequence (A_n) is forward sequentially primitive if and only if it is backward

sequentially primitive. Moreover, the forward and backward indices of sequential primitivity, ψ_n and ρ_n , have the same distributions.

The point in discussing both forward and backward primitivity will become clear in connection with Theorems 14 and 16 below in which the natural assumption is that $(A_n), n \geq 1$ is forward sequentially primitive, and $\mathbb{E}\psi_1 < \infty$. However, in the proof, we do need to ensure that for a two-sided extension of (A_n) , we have $\mathbb{E}\rho_1 < \infty$.

Consider now the case of an i.i.d. sequence $(A_n), n \geq 1$.

Remark 6: Let $(A_n), n \geq 1$, be an i.i.d. sequence of allowable matrices. Then, it is sequentially primitive if and only if the following set is primitive:

$$\overline{\mathcal{A}}^0 = \{C : P(\gamma(A_1) = C) > 0\}.$$

Obviously, the range of $(\gamma(A_n)), n \geq 1$ is finite. This motivates the assumption in the following lemma.

Lemma 7: Consider an i.i.d. sequence of nonnegative, allowable $p \times p$ matrices $(A_n), -\infty < n < \infty$ having a finite range \mathcal{A} , which is primitive. Then, ψ_n is finite w.p.1, and the tail probabilities of ψ_n decay geometrically, $P(\psi_n > x) < c \exp(-\alpha x)$ with some $c, \alpha > 0$. Analogous results hold for the indices of backward sequential primitivity ρ_n .

The almost trivial proof will be given in Appendix A. The above lemma implies that $\mathbb{E}\psi_n < \infty$, and since ψ_n has the same distribution for all n , the sequence ψ_n is sublinear, i.e., $\psi_n = o(n)$ a.s. Obviously, the same holds for the backward indices of sequential primitivity, i.e., $\rho_n = o(n)$ a.s.

IV. NORMALIZED PRODUCTS OF NONNEGATIVE RANDOM MATRICES

In this section, we describe the extension of a nice result of [21], the proof of which inspired the proofs of the main theorems of this article.

Let $(A_n), n \geq 1$ be a sequence of allowable $p \times p$ matrices. Let $x, w \in \mathbb{R}^p$ be component-wise nonnegative vectors, written as $x, w \geq 0$, the set of which will be denoted by \mathbb{R}_+^p , such that $x, w \neq 0$. Define the sequences

$$x_n := M_n x = A_n A_{n-1} \dots A_1 x \quad (24)$$

$$w_n := M_n w = A_n A_{n-1} \dots A_1 w. \quad (25)$$

Obviously x_n and w_n are nonnegative, and since the A_n -s are allowable and $x, w \neq 0$, we have $x_n, w_n \neq 0$. Therefore, we can define

$$\bar{x}_n = x_n / (\mathbf{1}^\top x_n), \quad \bar{w}_n = w_n / (\mathbf{1}^\top w_n). \quad (26)$$

The following result is a straightforward extension of [21]. In the theorem, $\|\bar{x}_n - \bar{w}_n\|_{\text{TV}} := \frac{1}{2} \sum_{i=1}^p |\bar{x}_n^i - \bar{w}_n^i|$ denotes the total variation distance of the probability vectors \bar{x}_n and \bar{w}_n .

Theorem 8: Assume that $(A_n), n \geq 1$ is a strictly stationary, ergodic process of random $p \times p$ matrices such that $\mathbb{E} \log^+ \|A_1\| < \infty$. In addition, assume that A_n is nonnegative and allowable for all n , and assume that the process (A_n) is sequentially primitive. Then, for all pairs $(x, w) \in \mathbb{R}_+^p \times \mathbb{R}_+^p$, except for a set of Lebesgue measure zero, it holds that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|\bar{x}_n - \bar{w}_n\|_{\text{TV}} = -(\lambda_1 - \lambda_2) \quad \text{w.p.1}$$

where λ_1 and λ_2 are the first and second largest Lyapunov exponents associated with (A_n) . In addition, for any fixed pair $(x, w) \in \mathbb{R}_+^p \times \mathbb{R}_+^p$ with strictly positive components with no exception, it holds that the above limit exists w.p.1, and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \|\bar{x}_n - \bar{w}_n\|_{\text{TV}} \leq -(\lambda_1 - \lambda_2).$$

The proof of Theorem 8 is a straightforward extension of [21, proof of Th. 1] and will be given in Appendix B. We should note, however, that the proof given in [21] contains two nontrivial deficiencies. These will be rectified by the following lemmas, the proofs of which will be given also in Appendix B. The first lemma was implicitly stated in [21], with a minor flaw in the proof.

Lemma 9: Let the sequence of matrices (A_n) be as in Theorem 8. Then, there exists a subset $\Omega' \subset \Omega$ with $P(\Omega') = 1$ such that for all $\omega \in \Omega'$ it holds that any strictly positive vector $x > 0$, $x \in \mathbb{R}^p$ is contained in $x \in L_{\mu_1} \setminus L_{\mu_2}$, see (8)–(10):

$$\lambda_1 = \lim_{n \rightarrow \infty} \frac{1}{n} \log |A_n A_{n-1} \dots A_1 x|. \quad (27)$$

The second result was tacitly used in [21], with no proof. Here, the notion of exterior product of vectors and matrices, denoted by $x \wedge w$ and $A \wedge B$, resp., is used. Here, $x \wedge w$ can be identified with the antisymmetric matrix $xw^\top - wx^\top$, and $(A \wedge B)(x \wedge w) = Ax \wedge Bw$; see [23].

Lemma 10: Let (A_n) , $n \geq 1$ be a strictly stationary, ergodic process of $p \times p$ random matrices such that $\mathbb{E} \log^+ \|A_n\| < \infty$. Consider the exterior product space $\mathbb{R}^p \wedge \mathbb{R}^p$ and the matrices $A_n \wedge A_n$ acting on it. Then, for all pairs $(x, w) \in \mathbb{R}^p \times \mathbb{R}^p$, except for a set of Lebesgue measure zero, the a.s. limit

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |((A_n A_{n-1} \dots A_1) \wedge (A_n A_{n-1} \dots A_1))(x \wedge w)|$$

exists and is equal to $\lambda_1 + \lambda_2$.

Motivated by Theorem 8, we consider the possibility of an extension of the results concerning the push-sum or weighted gossip algorithms under significantly more general conditions.

V. GENERALIZED RATIO CONSENSUS

In this section, we will formalize our main results on the convergence rate of a generalized ratio consensus algorithm. The common setup for our results will be based on Theorem 8. However, this will have to be complemented by a variety of additional conditions imposed on (A_n) .

For the formulation of our technical results, we will need to impose further conditions on the positive elements of A_n , controlling the possibility of moving a random fraction (or share) of values and weights during a transaction. Let us introduce the following notations for the minimal and maximal positive elements of A_n :

$$\alpha_n := \min_{ij} \{A_n^{ij} : A_n^{ij} > 0\}, \quad \beta_n := \max_{ij} A_n^{ij}. \quad (28)$$

Since β_n is equivalent to $\|A_n\|$, it follows immediately that $\mathbb{E} \log^+ \beta_n < \infty$. A direct consequence of this is that for any $\varepsilon > 0$, we have a.s. $\beta_n = O(e^{\varepsilon n})$, i.e., β_n is subexponential (see below). A twin pair of the condition $\mathbb{E} \log^+ \beta_n < \infty$ is the following.

Condition 11: Let (A_n) , $n \geq 1$ be a strictly stationary, ergodic process of random, $p \times p$ nonnegative matrices. We assume that $\mathbb{E} \log^- \alpha_n > -\infty$, where α_n is the minimal positive element of A_n defined previously.

A direct consequence of this condition is that $\mathbb{E} \log^+ \frac{1}{\alpha_n} < \infty$, implying that $\frac{1}{\alpha_n}$ is subexponential. The above condition is obviously satisfied if (A_n) takes its values from a finite set, say \mathcal{A} , w.p.1, which is the case with the push-sum algorithm allowing packet loss.

Theorem 12: Assume that the conditions of Theorem 8 are satisfied, in addition, the sequence (A_n) is i.i.d., and $\lambda_1 - \lambda_2 > 0$. Furthermore, assume that the minimal positive elements of A_n satisfy Condition 11. Let e_k denote the k th unit vector for any $k = 1, \dots, p$. Take an arbitrary vector of initial values $x \in \mathbb{R}^p$, and a nonnegative vector of initial weights $w \in \mathbb{R}_+^p$ such that $w \neq 0$. Then, ratio consensus takes place and an explicit upper bound for the rate of convergence can be given as follows: for all $i = 1, \dots, p$, we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left| \frac{e_i^\top M_n x}{e_i^\top M_n w} - \frac{v^1 x}{v^1 w} \right| \leq -(\lambda_1 - \lambda_2) \quad \text{w.p.1.} \quad (29)$$

By Theorem 12 for all agents i , the values x_n^i/w_n^i will converge to the same limit $\pi^T x$ a.s., where π is the random vector defined by $\pi = v^1/v^1 w$, with at least the given rate. The limit is random, in contrast to the case of classic push-sum or weighted gossip algorithms without packet loss. On the other hand, there is ample empirical evidence that *decreasing the probability of packet loss* leads to higher concentration of the distribution of $\pi^T x$, around \bar{x} ; see [9].

An extension of the aforementioned scenario is obtained if the communicating pairs of agents are chosen according to some time-homogeneous random pattern, which may be different from an i.i.d. choice, see geographic gossip, randomized path averaging, or one-way averaging, [2], [3], [8]. Thus, we come to consider the case when (A_n) is a general, *strictly stationary ergodic* sequence. As for the additional conditions to be imposed, we consider two levels of complexity.

Condition 13: Let (A_n) , $n \geq 1$ be a strictly stationary, ergodic process of random, $p \times p$ nonnegative matrices. We say that (A_n) is bounded from below and from above, if there exist $\alpha, \beta > 0$ such that, with the notations of (28), we have a.s.

$$\alpha_n \geq \alpha > 0, \quad \beta_n \leq \beta. \quad (30)$$

Again, the above condition is obviously satisfied if the range of (A_n) , denoted above by \mathcal{A} , is finite.

Theorem 14: Assume that the conditions of Theorem 8 are satisfied, $\lambda_1 - \lambda_2 > 0$, and for the forward index of sequential primitivity ψ_n , we have $\mathbb{E} \psi_n < \infty$. Furthermore, assume that the positive elements of A_n are bounded from below and from above in the sense of Condition 13. Then, for any vector of initial values $x \in \mathbb{R}^p$, and any nonnegative vector of initial weights $w \in \mathbb{R}_+^p$ such that $w \neq 0$ ratio consensus takes place, in fact, (29) holds.

A further extension of this result is obtained if the elements of A_n are *not bounded* from above and from below, thus allowing for the possibility of moving a negligible fraction of values

and weights. In this case, we need an extra technical condition ensuring some kind of mixing of the process (A_n) .

Condition 15: A two-sided strictly stationary process (ξ_n) satisfies a q th order M -mixing condition, with $q \geq 1$, if $\mathbb{E}|\xi_n|^q < \infty$, and for any positive integer N , we have, with some constant $C > 0$

$$\mathbb{E} \left| \sum_{n=1}^N (\xi_n - \mathbb{E}\xi_n) \right|^q \leq CN^{q/2}. \quad (31)$$

Theorem 16: Assume that the conditions of Theorem 8 are satisfied, $\lambda_1 - \lambda_2 > 0$, and for the index of forward sequential primitivity ψ_n , we have $\mathbb{E}\psi_n < \infty$. Furthermore, assume that $a_n = \log \alpha_n$ and $b_n = \log \beta_n$ satisfy a q th order M -mixing condition, given in Condition 15, with some $q > 4$. Then for any vector of initial values $x \in \mathbb{R}^p$, and any nonnegative vector of initial weights $w \in \mathbb{R}_+^p$ such that $w \neq 0$ ratio consensus takes place, in fact, (29) holds.

It may be of interest to consider an estimate of the average at any time n by taking a weighted average of the respective values of x_n^i and w_n^i . In this case, Theorems 12, 14, and 16 easily generalize to the following.

Corollary 17: Let $q \in \mathbb{R}_+^p, q \neq 0$ be a nonnegative weight vector. Assume that any of the sets of conditions of Theorems 12, 14, or 16 is satisfied. Then, for any vector of initial values $x \in \mathbb{R}^p$, and any nonnegative vector of initial weights $w \in \mathbb{R}_+^p$ such that $w \neq 0$, we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left| \frac{q^\top M_n x}{q^\top M_n w} - \frac{v^1 x}{v^1 w} \right| \leq -(\lambda_1 - \lambda_2) \quad \text{a.s.} \quad (32)$$

Proof of Corollary 17: The claim is obtained by a direct and standard convexity argument (see [2]): for any pair of vectors $a, b \in \mathbb{R}^p$ such that $b > 0$, we have

$$\min_i \frac{a^i}{b^i} \leq \frac{q^\top a}{q^\top b} \leq \max_i \frac{a^i}{b^i}. \quad (33)$$

Indeed, this follows from

$$\frac{q^\top a}{q^\top b} = \frac{\sum_i q^i a^i}{\sum_i q^i b^i} = \sum_i \left(\frac{a^i}{b^i} \right) \frac{q_i b^i}{\sum_j q^j b^j}. \quad (34)$$

Setting $a^i = e_i^\top M_n x$ and $b^i = e_i^\top M_n w$, we get

$$\min_i \frac{e_i^\top M_n x}{e_i^\top M_n w} \leq \frac{q^\top M_n x}{q^\top M_n w} \leq \max_i \frac{e_i^\top M_n x}{e_i^\top M_n w} \quad (35)$$

from which the claim follows by Theorems 12, 14, and 16. ■

Let the left-hand side (LHS) and the right-hand side (RHS) of (35) be denoted by y_n and z_n , respectively. The following elementary lemma, which will be used later on, has been established in [9] for the case of the push-sum algorithm with packet loss.

Lemma 18: The values y_n and z_n are monotone nondecreasing and nonincreasing, respectively. In particular, it follows that for any time n , we have

$$\min_i \frac{e_i^\top M_n x}{e_i^\top M_n w} \leq \frac{v^1 x}{v^1 w} \leq \max_i \frac{e_i^\top M_n x}{e_i^\top M_n w} \quad \text{a.s.} \quad (36)$$

Proof of Lemma 18: Indeed, for any index j , write

$$h_{n+1,j} := \frac{e_j^\top M_{n+1} x}{e_j^\top M_{n+1} w} = \frac{e_j^\top A_{n+1} M_n x}{e_j^\top A_{n+1} M_n w} = \frac{q_j^\top M_n x}{q_j^\top M_n w} \quad (37)$$

with $q_j^\top = e_j^\top A_{n+1}$. Since A_{n+1} is nonnegative and allowable, we have $q_j \geq 0, q_j \neq 0$. Thus, we get by (35), the inequality $y_n \leq h_{n+1,j} \leq z_n$ for all j from which the first claim follows. The second claim follows trivially from the established monotonicity, and the fact that, according to Theorem 12, we have a.s.

$$\lim_{n \rightarrow \infty} \min_i \frac{e_i^\top M_n x}{e_i^\top M_n w} = \frac{v^1 x}{v^1 w} = \lim_{n \rightarrow \infty} \max_i \frac{e_i^\top M_n x}{e_i^\top M_n w}. \quad \blacksquare$$

In the special case, when A_n is column stochastic for all n , as in the case of the push-sum or weighted gossip algorithm with no packet loss, M_n will be column-stochastic for all n . It follows that $\|M_n\|$ is bounded from above and bounded away from 0, hence, it readily follows that for the top-Lyapunov exponent we have $\lambda_1 = 0$, and we obtain the following result.

Theorem 19: Assume that any of the sets of conditions of Theorems 12, 14, or 16 is satisfied, and in addition, A_n is column stochastic for all n . Then, for any vector of initial values $x \in \mathbb{R}^p$, and any nonnegative vector of initial weights $w \in \mathbb{R}_+^p$ such that $w \neq 0$, we have for all $i = 1, \dots, p$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left| \frac{e_i^\top M_n x}{e_i^\top M_n w} - \frac{\mathbf{1}^\top x}{\mathbf{1}^\top w} \right| \leq \lambda_2 < 0 \quad \text{a.s.}$$

Choosing $w = \mathbf{1}$, Theorem 19 implies that ratio consensus will take place in the classic sense: for all agents k , the values x_n^k/w_n^k will converge to the same nonrandom limit $\bar{x} = \sum_{i=1}^p x_0^i/p$, with at least the given rate.

Remark 20: It may come as a pleasing surprise that the a.s. rate of convergence for weighted gossip algorithms provided by Theorem 19 is identical with the a.s. rate of convergence of a class of *linear gossip* algorithms, described in [10], defined via a strictly stationary ergodic edge process. By [10, Th. 5.2], with A_n denoting the associated *doubly stochastic* matrices, we have for any $x \in \mathbb{R}^p$ and any i

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log |e_i^\top A_n \cdots A_1 x - \frac{\mathbf{1}^\top x}{p}| \leq \lambda_2 \quad \text{a.s.} \quad (38)$$

We note that an extension of this result can be easily derived from the proof of Theorem 19: assuming the additional condition that A_n is *doubly stochastic* for all n , inequality (38) holds. Unfortunately the problem of deciding if $\lambda_2 < 0$ is generally not only NP hard, but undecidable [24], [25].

An upper bound for the rate of a.s. exponential convergence of an appropriately sampled process $x_{\tau_n}^i/w_{\tau_n}^i$, generated by the weighted gossip algorithms, was derived in [11] assuming, among others, that (A_n) is i.i.d. and column stochastic. These upper bounds for the rate, obtained via the analysis of the mean squared error of $A_n \cdots A_1 \cdot (I - \mathbf{1}\mathbf{1}^\top/p)$, are given by $\kappa = -\frac{1}{2} \log \rho(R)$, with $\rho(\cdot)$ denoting the spectral radius, and

$$R = \mathbb{E}[A_1 \otimes A_1] \cdot ((I - \mathbf{1}\mathbf{1}^\top/p) \otimes (I - \mathbf{1}\mathbf{1}^\top/p)).$$

We should note that the same *computable* upper bound for the rate of a.s. exponential convergence of the complete process x_n^i/w_n^i can be readily derived by combining the arguments of [11] with Lemma 27 of this article.

The upper bounds for the rates in the preceding theorems seem to have been unknown prior to this article. As for the exact rate, the best we can claim is the following theorem.

Theorem 21: Assume that any of the sets of conditions of Theorems 12, 14, or 16 is satisfied. Then, for all pairs of non-negative vectors $(x, w) \in \mathbb{R}_+^p \times \mathbb{R}_+^p$, such that $x, w \neq 0$, except perhaps for a set of Lebesgue-measure zero, it holds that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \max_i \left| \frac{e_i^\top M_n x}{e_i^\top M_n w} - \frac{v^1 x}{v^1 w} \right| = -(\lambda_1 - \lambda_2) \quad \text{a.s.}$$

VI. SPECIFICATION FOR PUSH-SUM WITH PACKET LOSS

In this section, we summarize the implications of the previously stated results for the classic push-sum or weighted gossip algorithm, allowing packet loss as described in the Introduction, which is in line with the setting of [9].

Theorem 22: Let $(A_n), n \geq 1$ be the associated i.i.d. sequence of matrices defined under (3). Assume that the directed communication graph (G, E) is strongly connected. Then, for any initial values $x \in \mathbb{R}^p$, and a nonnegative vector of initial weights $w \in \mathbb{R}_+^p$ such that $w \neq 0$ ratio consensus takes place, and for all $i = 1, \dots, p$, an explicit upper bound for the a.s. rate of convergence can be given as follows:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left| \frac{e_i^\top M_n x}{e_i^\top M_n \mathbf{1}} - \frac{v^1 x}{v^1 \mathbf{1}} \right| \leq -(\lambda_1 - \lambda_2).$$

In the case of no packet loss, we have $\lambda_1 = 0$ and $v^1 = \mathbf{1}^\top$.

Proof of Theorem 22: For the first step of the proof, we verify the only nontrivial condition of Theorem 8 requiring that (A_n) is sequentially primitive. Since (A_n) is an i.i.d. sequence, we can resort to Lemma 6. Consider, therefore, the (finite) range of the random matrices A_n given by (3), denoted by \mathcal{A}_{PS} . The first of the following two lemmas restates a well-known result in the consensus literature (see [2]), while the second one claims the validity of the key condition $\lambda_1 > \lambda_2$. The proofs will be given in Appendix C.

Lemma 23: Assume that the directed communication graph $G = (V, E)$ is strongly connected. Then, the set \mathcal{A}_{PS} is primitive.

Lemma 24: Let $(A_n), n \geq 1$ be an i.i.d. sequence of matrices corresponding to the push-sum algorithm allowing packet loss, defined in (3), satisfying the condition described in the Introduction. Then, we have for the spectral gap $\lambda_1 - \lambda_2 > 0$.

To complete the proof of Theorem 22, we apply Theorem 12, the conditions of which are partially assumed, and partially ensured by the above lemmas. This confirms the general case with possible packet loss. In the case of no packet loss, the claim $\lambda_1 = 0$ and $v^1 = \mathbf{1}$ is implied by Theorem 19. ■

Remark 25: Note that the argument used in [21] to estimate $\lambda_1 - \lambda_2$ from below cannot be used in our case. Namely, [21] refers to a result of [26]

$$\lambda_1 - \lambda_2 \geq -\mathbb{E} \log \tau(A_1)$$

where $\tau(A_1)$ is the Birkhoff contraction coefficients of A_1 (see in the following). However, in our case, we have $\tau(A_n) = 1$ a.s., hence, the lower bound is simply 0.

By this, we end the description of the key points of our work and switch to slightly heavier mathematical details. First, we describe the critical steps of the proofs of our main theorems, with some technical details relegated to Appendix D, and then, a mathematical interlude on the spectral gap is added.

VII. PROOFS OF THEOREMS 12, 14, 16, 19, AND 21

For the proof of Theorem 12, a natural starting point would be Theorem 8. However, we will see that nothing is gained compared to a direct proof. On the other hand, the situation is completely different in the case of Theorem 21, the proof of which will rely essentially on Theorem 8.

For the description of the proofs, we need the following definition. A stochastic process $\xi_n, n \geq 1$ is called subexponential, if for any $\varepsilon > 0$, we have for all n , with finitely many exceptions, a.s. $|\xi_n| \leq e^{\varepsilon n}$. We will use the notation $\xi_n = e^{o(1)n}$. Equivalently, $\xi_n, n \geq 1$ is subexponential if $\limsup_{n \rightarrow \infty} \frac{1}{n} \log |\xi_n| \leq 0$.

In view of (19), assuming $\lambda_1 > \lambda_2$, the matrix product M_n is asymptotically equivalent to the sequence of rank-1 matrices $u_n^1 v^1 \sigma_n^1$, a.s. A weak, *a priori* estimate of a measure of collinearity of the rows of M_n is formalized in Condition 26, under which the proofs of Theorems 12–21 will be completed. The validity of Condition 26 itself will be verified by Lemma 44 in Appendix D.

Condition 26: Letting $M_n = A_n A_{n-1} \cdots A_1$, as before, we assume that for any pair of row indices i, j , and any column index k , it holds that M_n^{ik}/M_n^{jk} is subexponential.

Lemma 27: Under the conditions of Theorem 8, the additional assumption that $\lambda_1 > \lambda_2$, and Condition 26, it holds that $1/u_n^{i1}$ is subexponential a.s. for all i .

Proof of Lemma 27: Recall that according to (19), we have a.s. $M_n = u_n^1 v^1 \sigma_n^1 + O(e^{(\lambda_2 + o(1))n})$. Take an arbitrary pair of row indices j, i , and compare the rows M_n^j and M_n^i . Choosing a column index k such that $v^{1k} > 0$, we consider

$$\frac{M_n^{jk}}{M_n^{ik}} = \frac{u_n^{j1} v^{1k} \sigma_n^1 + O(e^{(\lambda_2 + o(1))n})}{u_n^{i1} v^{1k} \sigma_n^1 + O(e^{(\lambda_2 + o(1))n})}. \quad (39)$$

Taking into account $v^{1k} > 0$, we would have for any j, i

$$\frac{M_n^{jk}}{M_n^{ik}} = \frac{u_n^{j1} + O(e^{(-\lambda_1 + \lambda_2 + o(1))n})}{u_n^{i1} + O(e^{(-\lambda_1 + \lambda_2 + o(1))n})}. \quad (40)$$

From this, it follows that $1/u_n^{i1}$ is subexponential as stated. Indeed, assume that this is not the case, then for some small $\varepsilon > 0$, we have $1/u_n^{i1} \geq e^{\varepsilon n}$ for an infinite subsequence, say $n = n_r$, consequently $u_n^{i1} \leq e^{-\varepsilon n}$ for $n = n_r$. Select j so that for some infinite subsequence of (n_r) , which we identify with (n_r) , we have $u_{n_r}^{j1} \geq 1/\sqrt{p}$. The indirect assumption and the choice of j would then imply $M_{n_r}^{jk}/M_{n_r}^{ik} \geq C e^{\varepsilon n}$ with some $C > 0$ infinitely many times a.s., which is a contradiction to Condition 26. ■

Lemma 28: Under the conditions of Theorem 8, with the additional assumption that $\lambda_1 > \lambda_2$, and Condition 26, it holds that $v^{1i} > 0$ for all $i = 1, \dots, p$.

Proof of Lemma 28: Consider the matrix process $\bar{A}_n = A_n^\top$. First, we show that the Lyapunov exponents for the processes (\bar{A}_n) and (A_n) are identical, $\bar{\lambda}_k = \lambda_k$ for all $k = 1, \dots, p$. Define for any pair of integers $n > m$, the products $M_{n,m} = A_n A_{n-1} \cdots A_m$ and $\bar{M}_{n,m} = \bar{A}_n \bar{A}_{n-1} \cdots \bar{A}_m$. Then, we have

$$\begin{aligned} M_{n,m}^\top &= (A_n A_{n-1} \cdots A_m)^\top = A_m^\top \cdots A_{n-1}^\top A_n^\top \\ &= \bar{A}_m \cdots \bar{A}_{n-1} \bar{A}_n = \bar{M}_{m,-n}. \end{aligned}$$

Let a singular value decomposition (SVD) of $M_{n,m}$ be

$$M_{n,m} = U_{n,m} \Sigma_{n,m} V_{n,m}.$$

Then, an SVD for $\bar{M}_{-m,-n}$ is obtained as follows:

$$\bar{M}_{-m,-n} = V_{n,m}^\top \Sigma_{n,m} U_{n,m}^\top =: \bar{U}_{-m,-n} \bar{\Sigma}_{-m,-n} \bar{V}_{-m,-n}$$

with the notations

$$\bar{U}_{-m,-n} = V_{n,m}^\top \quad (41)$$

$$\bar{\Sigma}_{-m,-n} = \Sigma_{n,m} \quad (42)$$

$$\bar{V}_{-m,-n} = U_{n,m}^\top. \quad (43)$$

To prove $\bar{\lambda}_1 = \lambda_1$, note that (41)–(43) implies

$$\begin{aligned} \bar{\lambda}_1 &= \lim_{-m \rightarrow \infty} \frac{1}{n-m+1} \log \bar{\sigma}_{-m,-n}^1 \\ &= \lim_{-m \rightarrow \infty} \frac{1}{n-m+1} \log \sigma_{n,m}^1 \end{aligned}$$

w.p.1, and hence, also in distribution. But $\sigma_{n,m}^1$ and $\sigma_{n-m+1,1}^1$ have the same distribution, and for the latter, we have

$$\lambda_1 = \lim_{-m \rightarrow \infty} \frac{1}{n-m+1} \log \sigma_{n-m+1,1}^1$$

w.p.1, and hence, also in distribution. Thus, the distribution of $\bar{\lambda}_1$ and λ_1 agree implying $\bar{\lambda}_1 = \lambda_1$.

Applying the same argument to the k th exterior product sequences formed by $A_n \wedge \cdots \wedge A_n$ and $\bar{A}_n \wedge \cdots \wedge \bar{A}_n$, we conclude that $\bar{\lambda}_1 + \cdots + \bar{\lambda}_k = \lambda_1 + \cdots + \lambda_k$ for all k implying the claim.

Next, consider the matrices $V_{n,m}$ with m fixed and n tending to ∞ . The first rows of $V_{n,m}$ denoted by $v_{n,m}^1$ converge a.s. to a limit, say v_m^1 , with exponential rate by [20, Lemma 5], the error being $O(e^{(-\lambda_1 + \lambda_2 + o(1))(n-m)})$. This implies, that the first columns of $\bar{U}_{-m,-n}$, denoted by $\bar{u}_{-m,-n}^1$ also converge to a limit $\bar{u}_{-m}^1 = v_m^{1\top}$ a.s. with the same exponential rate when n tends to ∞ .

Take $m = 1$ and assume in contrary to the statement of the lemma that $v^{1i} = v_1^{1i} = 0$ for some i . Then, $\bar{u}_{-1}^{1i} = 0$, and thus, $\bar{u}_{-1,-n}^{1i}$ is exponentially small a.s. when n tends to ∞ : writing $\xi_n := \bar{u}_{-1,-n}^{1i}$, we have for any $0 < \mu < \lambda_1 - \lambda_2$, with some $C(\omega) > 0$, the inequality $\xi_n \leq C(\omega)e^{-\mu n}$. This implies for the distribution of ξ_n that for any $\mu' < \mu < \lambda_1 - \lambda_2$

$$\begin{aligned} P(\xi_n \leq e^{-\mu'n}) &\geq P(C(\omega)e^{-\mu n} \leq e^{-\mu'n}) \\ &= P(C(\omega) \leq e^{(\mu-\mu')n}) \rightarrow 1, \quad \text{as } n \rightarrow \infty. \quad (44) \end{aligned}$$

On the other hand, shifting the time indices in $\bar{u}_{-1,-n}^{1i}$ by $n+1$, we get the random variables $\xi'_n := \bar{u}_{n,1}^{1i}$ having the same distribution as ξ_n . Applying Lemma 27 to the process (\bar{A}_n) , where the conditions are easily verified, we get that $1/\xi'_n$ is subexponential. Thus, for any $\varepsilon > 0$, we have $1/\xi'_n \leq C'(\omega)e^{\varepsilon n}$, with some $C'(\omega) > 0$. Following the argument given previously, we get for the distribution of $1/\xi'_n$ that for any $\varepsilon' > \varepsilon > 0$, it holds that $P(1/\xi'_n \leq e^{\varepsilon'n}) \rightarrow 1$ as $n \rightarrow \infty$, implying $P(e^{-\varepsilon'n} \leq \xi'_n) \rightarrow 1$, which in turn yields

$$P(\xi'_n < e^{-\varepsilon'n}) \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (45)$$

Choosing $0 < \varepsilon < \varepsilon' < \mu'$ and recalling that ξ'_n and ξ_n have the same distribution, we get a contradiction with (44), and thus, the proof is complete. ■

Proofs of Theorems 12, 14, and 16: Assuming the validity of Condition 26, to be established separately under each set of conditions of Theorems 12, 14, and 16, the proof of the quoted three theorems are identical.

Recall that we have by (19), $M_n = u_n^1 v^1 \sigma_n^1 + O(e^{(\lambda_2 + o(1))n})$, hence

$$\frac{e_i^\top M_n x}{e_i^\top M_n w} = \frac{e_i^\top u_n^1 v^1 x \sigma_n^1 + O(e^{(\lambda_2 + o(1))n})}{e_i^\top u_n^1 v^1 w \sigma_n^1 + O(e^{(\lambda_2 + o(1))n})}. \quad (46)$$

Divide both the numerator and the denominator by σ_n^1 to get

$$\frac{e_i^\top M_n x}{e_i^\top M_n w} = \frac{e_i^\top u_n^1 \cdot v^1 x + O(e^{(-\lambda_1 + \lambda_2 + o(1))n})}{e_i^\top u_n^1 \cdot v^1 w + O(e^{(-\lambda_1 + \lambda_2 + o(1))n})}. \quad (47)$$

Note that $v^{1\cdot} > 0$ by Lemma 28, and thus, $w \geq 0$, $w \neq 0$ imply $v^{1\cdot} w > 0$. Divide both the numerator and the denominator by $v^{1\cdot} w$ and also by $e_i^\top u_n^1$. The proof is then completed by noting that $1/e_i^\top u_n^1 = 1/u_n^{1i}$ is subexponential for all i , as stated in Lemma 27. ■

Proof of Theorem 19: First note that M_n is column-stochastic for all n , hence, $\|M_n\|$ is bounded from above and bounded away from zero. It follows that $\lambda_1 = 0$. To complete the proof, it is sufficient to show that $v^{1\cdot}$ is proportional to $\mathbf{1}^\top$, (implying that $v^{1\cdot} = \mathbf{1}^\top / \sqrt{p}$). Writing

$$\mathbf{1}^\top = \mathbf{1}^\top M_n = \mathbf{1}^\top u_n^1 v^1 \sigma_n^1 + O(e^{(\lambda_2 + o(1))n}) \quad \text{a.s.} \quad (48)$$

and noting that $\mathbf{1}^\top u_n^1$ and $\sigma_n^1 = \|M_n\|$ are bounded and bounded away from 0, after dividing by these, we get

$$c_n \mathbf{1}^\top = v^{1\cdot} + O(e^{(\lambda_2 + o(1))n}) \quad \text{a.s.} \quad (49)$$

with some possibly random scalar c_n . Letting $n \rightarrow \infty$, and taking into account $\lambda_2 < 0$, the RHS will converge to $v^{1\cdot}$, and thus, the LHS will also converge, implying that c_n converges to some c , yielding $c \mathbf{1}^\top = v^{1\cdot}$, as claimed. ■

Proof of Theorem 21: Note that the a.s. inequality

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \max_i \log \left| \frac{e_i^\top M_n x}{e_i^\top M_n w} - \frac{v^1 x}{v^1 w} \right| \leq -(\lambda_1 - \lambda_2) \quad (50)$$

follows directly from Theorem 12. For the proof that the inequality is actually an equality we will rely on Theorem 8. First note that, in addition to $w > 0$, we may assume $x > 0$, since the set of pairs $(x, w) \in \mathbb{R}^p \times \mathbb{R}^p$, having a 0 component in x has a zero Lebesgue measure. Now, note that for any pairs or probability vectors (\bar{x}, \bar{w}) , we have

$$\frac{1}{2} \max_i |\bar{x}^i - \bar{w}^i| \leq \|\bar{x} - \bar{w}\|_{\text{TV}} \leq \frac{p}{2} \max_i |\bar{x}^i - \bar{w}^i|.$$

Therefore, Theorem 8 can be restated as follows: for all pairs $(x, w) \in \mathbb{R}_+^p \times \mathbb{R}_+^p$, $x, w \neq 0$, except for a set of Lebesgue-measure zero, it holds a.s. that

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{1}{n} \log \max_i |\bar{x}_n^i - \bar{w}_n^i| \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log \max_i \left| \frac{x_n^i}{\mathbf{1}^\top x_n} - \frac{w_n^i}{\mathbf{1}^\top w_n} \right| = -(\lambda_1 - \lambda_2). \quad (51) \end{aligned}$$

We may relate this equality to a ratio consensus problem by rewriting the middle term as

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log \max_i \left| \frac{x_n^i}{w_n^i} - \frac{\mathbf{1}^\top x_n}{\mathbf{1}^\top w_n} \right| \cdot \frac{w_n^i}{\mathbf{1}^\top x_n} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \max_i \left(\log \left| \frac{x_n^i}{w_n^i} - \frac{\mathbf{1}^\top x_n}{\mathbf{1}^\top w_n} \right| + \log \frac{w_n^i}{\mathbf{1}^\top x_n} \right). \end{aligned} \quad (52)$$

Now, if a^i and b^i are real numbers, then $\max_i (a^i + b^i) \leq \max_i a^i + \max_i b^i$. Apply this inequality to the RHS of (52) and take into account (51) to get that $-(\lambda_1 - \lambda_2)$ is bounded from above by

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \left(\max_i \log \left| \frac{x_n^i}{w_n^i} - \frac{\mathbf{1}^\top x_n}{\mathbf{1}^\top w_n} \right| + \max_i \log \frac{w_n^i}{\mathbf{1}^\top x_n} \right).$$

Furthermore, if α_n and $\beta_n, n \geq 1$, are real numbers and $\gamma_n = \alpha_n + \beta_n$, then $\liminf_n \gamma_n \leq \liminf_n \alpha_n + \limsup_n \beta_n$.

Also note that $w_n^i \leq \mathbf{1}^\top w_n$ implies $\max_i \log(w_n^i / \mathbf{1}^\top w_n) \leq \log(\mathbf{1}^\top w_n / \mathbf{1}^\top x_n)$. Thus, we get

$$\begin{aligned} -(\lambda_1 - \lambda_2) &\leq \liminf_{n \rightarrow \infty} \frac{1}{n} \max_i \log \left| \frac{x_n^i}{w_n^i} - \frac{\mathbf{1}^\top x_n}{\mathbf{1}^\top w_n} \right| \\ &\quad + \limsup_{n \rightarrow \infty} \frac{1}{n} \log \frac{\mathbf{1}^\top w_n}{\mathbf{1}^\top x_n}. \end{aligned} \quad (53)$$

Now, by Corollary 17, $\mathbf{1}^\top w_n / \mathbf{1}^\top x_n$ has a finite, nonzero limit w.p.1, hence

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \frac{\mathbf{1}^\top w_n}{\mathbf{1}^\top x_n} = \lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{\mathbf{1}^\top w_n}{\mathbf{1}^\top x_n} = 0.$$

Hence, we conclude that

$$-(\lambda_1 - \lambda_2) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \max_i \log \left| \frac{x_n^i}{w_n^i} - \frac{\mathbf{1}^\top x_n}{\mathbf{1}^\top w_n} \right| \quad (54)$$

and combining this with (50), we can write equality and lim in place of lim inf on the RHS as

$$-(\lambda_1 - \lambda_2) = \lim_{n \rightarrow \infty} \frac{1}{n} \max_i \log \left| \frac{x_n^i}{w_n^i} - \frac{\mathbf{1}^\top x_n}{\mathbf{1}^\top w_n} \right|. \quad (55)$$

Now, in view of Corollary 17, we have

$$\min_i \frac{x_n^i}{w_n^i} \leq \frac{\mathbf{1}^\top x_n}{\mathbf{1}^\top w_n} \leq \max_i \frac{x_n^i}{w_n^i}. \quad (56)$$

On the other hand, the trivial inequalities

$$\begin{aligned} \frac{1}{2} \left| \max_i \frac{x_n^i}{w_n^i} - \min_i \frac{x_n^i}{w_n^i} \right| &\leq \max_i \left| \frac{x_n^i}{w_n^i} - \frac{\mathbf{1}^\top x_n}{\mathbf{1}^\top w_n} \right| \\ &\leq \left| \max_i \frac{x_n^i}{w_n^i} - \min_i \frac{x_n^i}{w_n^i} \right| \end{aligned} \quad (57)$$

combined with (55) yield

$$-(\lambda_1 - \lambda_2) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \left| \max_i \frac{x_n^i}{w_n^i} - \min_i \frac{x_n^i}{w_n^i} \right| \quad \text{a.s.} \quad (58)$$

except for a set of initial (x, w) -s of Lebesgue measure zero. Considering (57) and replacing $\mathbf{1}^\top x_n / \mathbf{1}^\top w_n$ by an arbitrary sequence of intermediate values v_n such that

$$\min_i \frac{x_n^i}{w_n^i} \leq v_n \leq \max_i \frac{x_n^i}{w_n^i}$$

we get by the same logic

$$-(\lambda_1 - \lambda_2) = \lim_{n \rightarrow \infty} \frac{1}{n} \max_i \log \left| \frac{x_n^i}{w_n^i} - v_n \right|. \quad (59)$$

Taking $v_n = v^1 x / v^1 w$ for all n , in view of Lemma 18, we get the claim. \blacksquare

Remark 29: In the special case, when M_n is column stochastic, we have $\mathbf{1}^\top x_n = \mathbf{1}^\top M_n x = \mathbf{1}^\top x$, and similarly, $\mathbf{1}^\top w_n = \mathbf{1}^\top w$ for all n . Furthermore, by Theorem 19, we have $v^1 = \mathbf{1}^\top$. Thus, in this special case, (55) immediately implies the claim without any further deliberations.

VIII. REPRESENTATION OF THE SPECTRAL GAP $\lambda_1 - \lambda_2$

As we have seen, the spectral gap $\lambda_1 - \lambda_2$ plays a key role in characterizing the stability of normalized products and the convergence rate of the ratio consensus method. In this section, we present a set of simple results providing computable lower bounds and alternative representations for the spectral gap under the conditions of Theorems 8, 12, 14, or 16.

A lower bound for the spectral gap was established in [26, Prop. 5], under the condition that A_1 is strictly positive with positive probability. In fact, this result is a simple corollary of Theorem 8 relying on its less restrictive conditions. For the formal statement, we introduce the following definitions and notations.

Definition 30: Let $x, y \in \mathbb{R}_+^p$ be strictly positive vectors, $x, y > 0$. Then, their Hilbert distance is defined as

$$h(x, y) := \log \max_{k,l} \left(\frac{x_k}{y_k} / \frac{x_l}{y_l} \right). \quad (60)$$

The Hilbert distance satisfies the properties of a metric within the set of strictly positive vectors in \mathbb{R}^p , except that $h(x, y) = 0$ if and only if $y = cx$ with some $c > 0$. The operator norm of a nonnegative allowable matrix A corresponding to the Hilbert distance is called the Birkhoff contraction coefficient of A . More exactly, we set the following.

Definition 31: The Birkhoff contraction coefficient of a nonnegative allowable matrix A is defined as

$$\tau(A) := \sup \left\{ \frac{h(Ax, Ay)}{h(x, y)} \mid x, y \in \mathbb{R}_+^p, h(x, y) \neq 0 \right\}.$$

Note that $x, y > 0$ and the assumption that A is allowable imply that $Ax, Ay > 0$, and thus, $h(Ax, Ay)$ is well-defined. It is easy to see that $\tau(A)$ is submultiplicative, i.e., $\tau(AB) \leq \tau(A) \cdot \tau(B)$, and obviously, $\tau(A) \leq 1$.

A beautiful theorem due to Birkhoff yields an explicit expression of $\tau(A)$ in terms of the elements of A , which we present for allowable matrices. Define an intermediary quantity $\varphi(A)$ as follows. Let $\varphi(A) = 0$ if A has any 0 element. Otherwise, we set

$$\varphi(A) := \log \max_{i,j,k,l} \left(\frac{A^{ik}}{A^{jk}} \right) / \left(\frac{A^{il}}{A^{jl}} \right) = \max_{i,j} h(A^i, A^j). \quad (61)$$

By Birkhoff's theorem (see [27, Th. 3.12] or [28])

$$\tau(A) = \tanh \left(\frac{\varphi(A)}{4} \right) = \frac{e^{\varphi(A)/4} - e^{-\varphi(A)/4}}{e^{\varphi(A)/4} + e^{-\varphi(A)/4}}. \quad (62)$$

Theorem 32: Let $(A_n), n \geq 1$ be a strictly stationary, ergodic stochastic process of $p \times p$ matrices satisfying the conditions of Theorem 8. Then

$$\lambda_1 - \lambda_2 \geq -\mathbb{E} \log \tau(A_1).$$

Proof of Theorem 32: Since A_m is allowable for all m and x, w are strictly positive, the Hilbert distances of $x_n = A_n A_{n-1} \cdots A_1 x$ and $w_n = A_n A_{n-1} \cdots A_1 w$ are well defined, and we have

$$\begin{aligned} h(x_n, w_n) &= h(A_n A_{n-1} \cdots A_1 x, A_n A_{n-1} \cdots A_1 w) \\ &\leq \prod_{k=1}^n \tau(A_k) \cdot h(x, w). \end{aligned} \quad (63)$$

Therefore, we get

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log h(x_n, w_n) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log \tau(A_k) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log \tau(A_k) = \mathbb{E} \log \tau(A_1) \quad \text{a.s.} \end{aligned} \quad (64)$$

where the last two equalities follow from the ergodic theorem. Note that we can handle also the case when $\mathbb{E} \log \tau(A_1) = -\infty$ since $\log \tau(A_1)$ is bounded from above by 0. Now, the LHS can be bounded from below via the total variation $\|\bar{x}_n - \bar{w}_n\|_{\text{TV}}$ using the following elementary lemma.

Lemma 33: Let ξ and η be two strictly positive probability vectors in \mathbb{R}^p . Then, for their total variation distance, we have

$$\|\xi - \eta\|_{\text{TV}} \leq \frac{1}{2} (e^{h(\xi, \eta)} - 1).$$

Proof of Lemma 33: Let us write briefly $h = h(\xi, \eta)$. First note that for any k, l , we have

$$\frac{\xi^k}{\eta^k} / \frac{\xi^l}{\eta^l} \leq e^h.$$

Define $R = \max_k \frac{\xi^k}{\eta^k}$, $r = \min_l \frac{\xi^l}{\eta^l}$. Since ξ, η are probability vectors, we have $R \geq 1 \geq r$, and thus, from the above inequality, we get $e^{-h} \leq r \leq R \leq e^h$. Taking a k such that $\xi^k \geq \eta^k$, we have

$$|\xi^k - \eta^k| = \xi^k - \eta^k = \left(\frac{\xi^k}{\eta^k} - 1 \right) \eta^k \leq (e^h - 1) \eta^k.$$

On the other hand, for $\xi^k \leq \eta^k$, we get

$$\begin{aligned} |\xi^k - \eta^k| &= \eta^k - \xi^k = \left(1 - \frac{\xi^k}{\eta^k} \right) \eta^k \\ &\leq (1 - e^{-h}) \eta^k \leq (e^h - 1) \eta^k. \end{aligned}$$

Summation over k gives the claim. \blacksquare

To complete the proof of Theorem 32, we note that due to the above lemma, we can bound $h = h(\xi, \eta)$ from below for small h , say for $0 \leq h \leq 1/2$, we get $\|\xi - \eta\|_{\text{TV}} \leq h$. Taking into account that the Hilbert distance is invariant w.r.t. scaling its arguments, we have $h(x_n, w_n) = h(\bar{x}_n, \bar{w}_n)$, and this is exponentially small by Theorem 8, thus we can use $\|\xi - \eta\|_{\text{TV}} \leq h$ in (64) to get

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \|\bar{x}_n - \bar{w}_n\|_{\text{TV}} &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log h(\bar{x}_n, \bar{w}_n) \\ &= \limsup_{n \rightarrow \infty} \frac{1}{n} \log h(x_n, w_n) \leq \mathbb{E} \log \tau(A_1) \quad \text{a.s.} \end{aligned} \quad (65)$$

But we know by Theorem 8 that for almost all pairs (x, w) , $x > 0, w > 0$, the left side is equal to $-(\lambda_1 - \lambda_2)$ a.s., even with lim

instead of lim sup. From here after rearrangement, we get the claim. \blacksquare

Note that the result proven is directly not applicable for the analysis of the push-sum algorithm allowing packet loss, since all off-diagonal elements of A_1 , except at most one, is 0, and hence, $\tau(A_1) \equiv 1$ for all ω . A set of alternative lower bounds can be obtained by segmenting the product $A_n \cdots A_1$ into the product of blocks of fixed length, say $m \geq 1$. Let $A_n(\omega) = A_1(T^n \omega)$, where T is a measure-preserving ergodic transformation of Ω . Theorem 32 has the following extension.

Theorem 34: Let $(A_n), n \geq 1$ be a strictly stationary, ergodic stochastic process of $p \times p$ matrices satisfying the conditions of Theorem 8. Then, for all integers $m \geq 1$, we have

$$\lambda_1 - \lambda_2 \geq -\frac{1}{m} \mathbb{E} \log \tau(M_m). \quad (66)$$

Proof of Theorem 34: Let $m \geq 1$, and define $B_n = A_{nm} \cdot A_{(n-1)m+1} \cdots A_{(n-1)m+1}$. Obviously, $B_{n+1}(\omega) = B_n(T^m \omega)$, thus (B_n) is a strictly stationary process. Now, in analogy with (63), we have

$$\begin{aligned} h(x_{nm}, w_{nm}) &= h(B_n B_{n-1} \cdots B_1 x, B_n B_{n-1} \cdots B_1 w) \\ &\leq \prod_{k=1}^n \tau(B_k) \cdot h(x, w). \end{aligned} \quad (67)$$

Therefore, we get

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{nm} \log h(x_{nm}, w_{nm}) &\leq \limsup_{n \rightarrow \infty} \frac{1}{nm} \sum_{k=1}^n \log \tau(B_k) \\ &= \frac{1}{m} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log \tau(B_k) \quad \text{w.p.1} \end{aligned} \quad (68)$$

where the last equality follows from the ergodic theorem. Here, the LHS is bounded from below by $-(\lambda_1 - \lambda_2)$ w.p.1 as seen previously. Applying the ergodic theorem once again, the RHS converges to $\frac{1}{m} \mathbb{E} [\log \tau(B_1) | \mathcal{F}_{T^m}]$, where \mathcal{F}_{T^m} denotes the σ -algebra of invariant sets w.r.t. T^m . Thus, we get the a.s. upper bound for $-(\lambda_1 - \lambda_2)$ as

$$\frac{1}{m} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log \tau(B_k) = \frac{1}{m} \mathbb{E} [\log \tau(B_1) | \mathcal{F}_{T^m}].$$

Taking expectation of both sides, we get the claim. \blacksquare

Now, it is easy to see that the sequence $\mathbb{E} \log \tau(M_m)$ is subadditive (for any ergodic T), therefore, $\mathbb{E} \log \tau(M_m)/m$ has a limit (the value of which may be $-\infty$). Additionally

$$\lim_{m \rightarrow \infty} \frac{1}{m} \mathbb{E} \log \tau(M_m) = \inf_m \frac{1}{m} \mathbb{E} \log \tau(M_m).$$

Thus, we get the following corollary.

Corollary 35: Let $(A_n), n \geq 1$ be a strictly stationary, ergodic stochastic process of $p \times p$ matrices satisfying the conditions of Theorem 8. Then, $\lambda_1 - \lambda_2$ is bounded from below by

$$\lim_{m \rightarrow \infty} -\frac{1}{m} \mathbb{E} \log \tau(M_m) = \sup_m -\frac{1}{m} \mathbb{E} \log \tau(M_m). \quad (69)$$

A nice application of Corollary 35, providing a lower bound for the spectral gap, is the following.

Theorem 36: Let $(A_n), n \geq 1$ be a strictly stationary, ergodic stochastic process of $p \times p$ matrices satisfying the conditions of Theorem 8. Then, we have $\lambda_1 - \lambda_2 > 0$.

Proof: Since (A_n) is sequentially primitive, there exists a finite m such that $P(M_m > 0) > 0$. But then $P(\tau(M_m) < 1) > 0$, and hence, $-\mathbb{E} \log \tau(M_m) > 0$. The claim now follows from the second part of Corollary 35. ■

A natural question that arises at this point is if we can drop the expectation from (69). We show that in fact this can be done using Kingmans's subadditive ergodic theorem; see [29]–[32].

Theorem 37: Let $(A_n), n \geq 1$ be a strictly stationary, ergodic stochastic process of $p \times p$ matrices satisfying the conditions of Theorem 8. Then, we have

$$\lim_{m \rightarrow \infty} \frac{1}{m} \log \tau(M_m) = \lim_{m \rightarrow \infty} \frac{1}{m} \mathbb{E} \log \tau(M_m) \quad \text{w.p.1.}$$

Proof of Theorem 37: The double index series $M_{m,k} = A_m A_{m-1} \cdots A_k$ is obviously strictly stationary, $M_{m+1,k+1}(\omega) = M_{m,k}(T\omega)$, where T is ergodic. It follows that the double index series $\log \tau(M_{m,k})$ is also strictly stationary. Moreover, it is obviously subadditive, and $\mathbb{E} \log^+ \tau(M_{1,1}) = 0$ since $\tau(M_{1,1}) \leq 1$. Thus, by the subadditive ergodic theorem, we have

$$\lim_{m \rightarrow \infty} \frac{1}{m} \log \tau(M_{m,1}) = \lim_{m \rightarrow \infty} \frac{1}{m} \mathbb{E} \log \tau(M_{m,1}) \quad \text{w.p.1}$$

which proves our claim. ■

Combining this theorem with Corollary 35, we get the following extension.

Corollary 38: Let $(A_n), n \geq 1$ be a strictly stationary, ergodic stochastic process of $p \times p$ matrices satisfying the conditions of Theorem 8. Then, we have the following lower bound for the spectral gap:

$$\lambda_1 - \lambda_2 \geq \lim_{m \rightarrow \infty} -\frac{1}{m} \log \tau(M_m) \quad \text{w.p.1.} \quad (70)$$

The aforementioned results can be interpreted as providing various lower bounds for $\log \tau(M_m)$. We will now develop an a.s. asymptotic upper bound for $\log \tau(M_m)$ using techniques developed in the previous sections. Taking into account (62), the Birkhoff contraction coefficient $\tau(M_m)$, for its small values and for $M_m > 0$, is equivalent to $\varphi(M_m)$. On the other hand, $\varphi(M_m)$ is a measure of collinearity of the rows of M_m ; see (61). Thus, an asymptotic upper bound for $\tau(M_m)$ provides a bound on the speed with which M_m converges to a rank-1 matrix.

Theorem 39: Assume that any of the sets of conditions of Theorems 12, 14, or 16 is satisfied. Then, we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \tau(M_n) \leq -(\lambda_1 - \lambda_2) \quad \text{w.p.1.} \quad (71)$$

Proof of Theorem 39: The conditions of the theorem are identical to those of Lemma 44, implying that for any pair of row indices i, j and any column index k , the quotient M_n^{ik}/M_n^{jk} is subexponential, and thus, Condition 26 is satisfied. It follows that the conditions of Lemma 27 are also satisfied, implying that $1/u_n^{i1}$ is subexponential a.s. for all i .

Now, consider the equality (40), developed in the course of the proof of Lemma 27. Recall that $|u_n^{j1}| \leq 1$ and $1/u_n^{i1}$ is subexponential for all i and j . Hence, dividing both the numerator and the denominator of (40) by u_n^{i1} , we get, independently of the

column index k

$$\frac{M_n^{jk}}{M_n^{ik}} = \frac{u_n^{j1}}{u_n^{i1}} + O(e^{(-\lambda_1 + \lambda_2 + o(1))n}) \quad \text{a.s.} \quad (72)$$

By assumption for sufficiently large (random) n , the matrix M_n is strictly positive, hence, we can write, see (61)

$$\varphi(M_n) = \max_{i,j,k,l} \log \left(\frac{M_n^{jl}}{M_n^{il}} \right) / \left(\frac{M_n^{jk}}{M_n^{ik}} \right). \quad (73)$$

Taking into account (72), and once again noting that $|u_n^{j1}| \leq 1$ and $1/u_n^{i1}$ is subexponential for all i and j , we get a.s.

$$\varphi(M_n) = O(\log(1 + e^{(-\lambda_1 + \lambda_2 + o(1))n})) = O(e^{(-\lambda_1 + \lambda_2 + o(1))n}).$$

Taking into account Birkhoff's quoted theorem, stating that $\tau(M_n) = \tanh(\varphi(M_n)/4)$, we immediately get

$$\tau(M_n) = O(e^{(-\lambda_1 + \lambda_2 + o(1))n}) \quad (74)$$

from which the theorem immediately follows. ■

From the theorem obtained, we get via a trivial rearrangement an a.s. upper bound for the spectral gap in terms of the Birkhoff contraction coefficient as

$$\lambda_1 - \lambda_2 \leq -\limsup_{m \rightarrow \infty} \frac{1}{m} \log \tau(M_m) \quad \text{w.p.1.} \quad (75)$$

We have seen that on the RHS lim sup can be replaced with lim. Combining the above upper bound for the gap with the lower bound obtained in Corollary 38, we get the following result.

Theorem 40: Assume that any of the sets of conditions of Theorems 12, 14, or 16 is satisfied. Then, we have

$$\lambda_1 - \lambda_2 = \lim_{m \rightarrow \infty} -\frac{1}{m} \log \tau(M_m) \quad \text{w.p.1.} \quad (76)$$

IX. DISCUSSION

We should point out that the characterization of the a.s. rate of convergence via the spectral gap $\lambda_1 - \lambda_2$ may provide a solid ground for further investigations of direct practical interest, such as explicit estimates on the relation of spectral gap with respect to the number of nodes, the failure probabilities or the strength of connectivity, see [11] on related empirical results to this effect. Let us mention two simple facts that may be relevant in such investigations.

First, we note that $\lambda_1(\mathcal{A})$ is monotone non-decreasing in \mathcal{A} . More precisely, letting $\mathcal{A} = (A_n)$ and $\mathcal{A}' = (A'_n)$, and assuming $A_n \leq A'_n$ entry-wise for all n w.p.1 implies $\lambda_1(\mathcal{A}) \leq \lambda_1(\mathcal{A}')$. Indeed, $A'_n A'_{n-1} \cdots A'_1$ is entry-wise not less than $A_n A_{n-1} \cdots A_1$, hence letting $\|B\| = \sum_{i,j} b_{ij}$, we have $\|A_n A_{n-1} \cdots A_1\| \leq \|A'_n A'_{n-1} \cdots A'_1\|$, implying the stated inequality. From the above observation we immediately get the following simple result:

Lemma 41: Let (A_n) and (A'_n) be two strictly stationary, ergodic processes of matrices associated with the push-sum method on the same underlying network but with packet loss probabilities $r_{ij} \leq r'_{ij}$ for all i, j . Then $\lambda_1(\mathcal{A}) \geq \lambda_1(\mathcal{A}')$.

Unfortunately, the effect of increasing the packet loss probabilities on λ_2 is yet unknown. If we had $\lambda_2(\mathcal{A}) \leq \lambda_2(\mathcal{A}')$ then we could conclude that increasing the packet loss probabilities would decrease, or at least not increase the gap. A nice observation here is that although we do not know if $\lambda_2(\mathcal{A}) \leq \lambda_2(\mathcal{A}')$ we do know that $\sum_{i=2}^p \lambda_i(\mathcal{A}) \leq \sum_{i=2}^p \lambda_i(\mathcal{A}')$. The last inequality

follows from a simple relationship for the sum of the Lyapunov exponents given in the lemma below.

Lemma 42: Let (A_n) be a sequence of $p \times p$ matrices satisfying the conditions of Proposition 1. Then we have

$$\lambda_1 + \cdots + \lambda_p = \mathbb{E} \log(|\det A_1|).$$

In the case of the push-sum algorithm allowing packet loss we get $\lambda_1 + \cdots + \lambda_p = -\log 2$.

The magic of the lemma is in the fact that the l.h.s. depends only on the marginal distribution of A_1 .

Proof of Lemma 42: For the p -factor exterior product, we have

$$A_n \wedge \cdots \wedge A_n = \det A_n.$$

Therefore

$$\prod_{k=1}^n (A_k \wedge \cdots \wedge A_k) = \prod_{k=1}^n \det A_k.$$

On the other hand, using the singular-value decomposition $A_n \cdots A_1 = U_n \Sigma_n V_n$, we can write

$$\begin{aligned} & \prod_{k=1}^n (A_k \wedge \cdots \wedge A_k) \\ &= \prod_{k=1}^n (U_k \wedge \cdots \wedge U_k) \cdot \prod_{k=1}^n (\Sigma_k \wedge \cdots \wedge \Sigma_k) \\ & \cdot \prod_{k=1}^n (V_k \wedge \cdots \wedge V_k). \end{aligned}$$

Therefore

$$\prod_{k=1}^n \det A_k = \pm \prod_{k=1}^n \det \Sigma_k = \pm \prod_{k=1}^n \sigma_k^1 \cdots \sigma_k^p.$$

Taking absolute value and logarithm, dividing by n , and then taking the limit, we get

$$\mathbb{E} \log(|\det A_1|) = \lambda_1 + \cdots + \lambda_p.$$

In the case of the push-sum algorithm allowing packet loss, we have $|\det A_n| = 1/2$ for all n and all ω , thus, we get the claim. ■

Setting $p = 2$, the combination of the aforementioned two observations give that in the case of the push-sum algorithm increasing the probabilities of packet loss will decrease the spectral gap

$$\lambda_1(\mathcal{A}) - \lambda_2(\mathcal{A}) \geq \lambda_1(\mathcal{A}') - \lambda_2(\mathcal{A}') \quad (77)$$

for any strictly stationary, ergodic 2×2 matrix-valued processes (A_n) and (A'_n) of the form (3), no matter what the dependence structure is.

X. CONCLUSION

The problems discussed in the article are motivated by the ratio consensus problems and algorithms, such as the push-sum or weighted gossip algorithms. We have considered fairly general, strictly stationary communication protocols, covering as special cases broadcast algorithms, geographic gossip, randomized path averaging, or one-way averaging. We have given sharp upper bounds for the rate of the a.s. exponential convergence in terms of the spectral gap of the associated matrix sequence under various technical conditions. We have presented a variety of connections between the spectral gap and the Birkhoff contraction coefficient of the product of the associated matrices. Among others, our results provide a solution to an open problem raised in [2].

APPENDIX A SEQUENTIAL PRIMITIVITY

Proof of Lemma 5: The lemma is a direct consequence of the observation below, in which the indices of forward sequential primitivity will be replaced by a sequence of generic nonnegative waiting times Δ_n for all n , while the analogue of the indices of backward sequential primitivity, denoted by Δ'_n , will be defined in terms of the sequence (Δ_n) . Thus, let (Δ_n) be a two-sided strictly stationary, nonnegative process. Define for all n

$$m_n = \max_{m \leq n} \{m : m + \Delta_m \leq n\} \quad \text{and} \quad \Delta'_n = n - m_n. \quad (78)$$

Then, we claim that the probability distributions of Δ_n and Δ'_n are the same for all n . In particular, $\mathbb{E} \Delta_n = \mathbb{E} \Delta'_n$. Indeed

$$\begin{aligned} P(\Delta'_n > x) &= P(n - m_n > x) = P(m_n < n - x) \\ &= P(n - x + \Delta_{n-x} > n) = P(\Delta_{n-x} > x). \end{aligned} \quad (79)$$

for any $x \geq 0$. Since (Δ_n) is strictly stationary, we have $P(\Delta_{n-x} > x) = P(\Delta_n > x)$, and thus, the proof is complete. ■

Proof of Lemma 7: Let the elements of \mathcal{A} be denoted by B_1, B_2, \dots, B_r so that $P(A_1 = B^i) > 0$ for all i . The i.i.d. sequence (A_n) can be identified with an i.i.d. sequence of indices i_1, i_2, \dots , with $1 \leq i_k \leq r$. Since \mathcal{A} is primitive, there exists a word $w = (j_s, j_{s-1}, \dots, j_1)$ such that $B_{j_s} B_{j_{s-1}} \cdots B_{j_1} > 0$. Segment the full sequence of indices into an i.i.d. sequence of s -tuples v_m . Let $\tau := \min\{m : v_m = w\}$. Since $p := P(v_m = w) > 0$ implies $P(\tau > x) = (1 - p)^x$ and $\psi_1 \leq m\tau$, the claim follows. ■

APPENDIX B NORMALIZED PRODUCTS

In this section, we present the proof of Theorem 8, starting with the proofs of the auxiliary results, Lemmas 9 and 10.

Proof of Lemma 9: Consider

$$M_n^\top M_n = V_n^\top \text{diag}((\sigma_n^i)^2) V_n.$$

For $n \geq \tau$, this is a symmetric positive semidefinite matrix with strictly positive elements. Its eigenvalues are $(\sigma_n^i)^2$ with corresponding eigenvectors $(v_n^i)^\top$. By the Perron–Frobenius theorem, $M_n^\top M_n$ has a unique eigenvalue with maximal modulus, which is positive as is the corresponding eigenvector. It follows that $(\sigma_n^1)^2$ is a single eigenvalue, and $v_n^1 > 0$ elementwise.

Expand x in the orthonormal system defined by the rows of V_n : $x^\top = \sum_i \alpha_n^i v_n^i$. Here, $\alpha_n^i := v_n^i x$. Then

$$x^\top M_n^\top M_n x = \sum_i (\sigma_n^i)^2 (\alpha_n^i)^2.$$

Now, $v_n^1 > 0$ and $|v_n^1| = 1$, together with $x > 0$ imply that $\alpha_n^1 > \alpha_1 > 0$ with some α_1 . Thus, $x^\top M_n^\top M_n x > (\sigma_n^1)^2 \alpha_1^2$, from which we get $\liminf_n \frac{1}{n} \log |x^\top M_n^\top M_n x| \geq 2\lambda_1$, implying $\liminf_n \frac{1}{n} \log |M_n x| \geq \lambda_1$, and thus, the claim of the lemma follows. ■

Proof of Lemma 10: Write $V'_1 = \mathbb{R}^p \wedge \mathbb{R}^p$. According to Oseledec's theorem, there is a proper random subspace of V'_1

of fixed dimension, say V_2' , such that for $z \in V_1' \setminus V_2'$

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \log |((A_n A_{n-1} \cdots A_1) \wedge (A_n A_{n-1} \cdots A_1))z| \\ & = \lambda_1 + \lambda_2 \quad \text{a.s.} \end{aligned}$$

Consider the tensor product space $\mathbb{R}^p \otimes \mathbb{R}^p$ and its canonical linear mapping to $V_1' = \mathbb{R}^p \wedge \mathbb{R}^p$, denoted by S , defined by

$$\sum_{i,j} x_{ij} e_i \otimes e_j \longrightarrow \sum_{i,j} x_{ij} e_i \wedge e_j = \sum_{i < j} (x_{ij} - x_{ji}) e_i \wedge e_j.$$

Equivalently, interpreting $\mathbb{R}^p \otimes \mathbb{R}^p$ as the linear space of matrices of size $p \times p$, and identifying $\mathbb{R}^p \wedge \mathbb{R}^p$ as the linear space of antisymmetric matrices, the linear transformation S takes the form $S(X) = X - X^\top$.

It is readily seen that $V_2'' = S^{-1}V_2'$ is a proper subspace of the tensor product space $\mathbb{R}^p \otimes \mathbb{R}^p$. Indeed, any $X \in \mathbb{R}^p \otimes \mathbb{R}^p$ can be written as $X = X_a + X_s$, as a sum of its antisymmetric and symmetric part, and we have $S(X) = 2X_a$. Therefore, the linear subspace $V_2'' = S^{-1}V_2'$ consists of matrices for which $X_a \in V_2'$, and thus, it is indeed a proper subspace.

Let E denote the random set of exceptional pairs $(x, w)(\omega)$ defined as

$$E_{xw}(\omega) = \{(x, w) : x \otimes w \in V_2''(\omega)\}. \quad (80)$$

We claim that $E_{xw}(\omega) \in \mathbb{R}^p \times \mathbb{R}^p$ has zero Lebesgue-measure for all almost all ω . Assuming the contrary, there is a set $E_x(\omega) \in \mathbb{R}^p$ of positive Lebesgue measure such that for each $x \in E_x(\omega)$, the set

$$E_{w|x}(\omega) = \{w : (x, w) \in E_{xw}(\omega)\}.$$

has positive Lebesgue measure in \mathbb{R}^p . Taking any $x \in E_x(\omega)$, the elements of $E_{w|x}(\omega)$ span the full \mathbb{R}^p , therefore, (x, w) , $w \in E_{w|x}(\omega)$ span the linear space $x \otimes \mathbb{R}^p$. Letting x vary through the positive set $E_x(\omega)$, we get that the elements of $x \otimes \mathbb{R}^p$ span the whole $\mathbb{R}^p \times \mathbb{R}^p$. This is in contradiction with the assumption that for any $(x, w) \in E_{xw}(\omega)$, the tensor product $x \otimes w$ lies in the proper subspace V_2'' .

We conclude by Fubini's theorem that the exceptional set in $\mathbb{R}^p \times \mathbb{R}^p \times \Omega$

$$E_{xw\omega} = \{(x, w, \omega) : (x, w) \in V_2''(\omega)\}, \quad (81)$$

has $\lambda \times \lambda \times P$ -measure zero. Applying Fubini's theorem once again in the opposite direction, we get the claim. \blacksquare

Proof of Theorem 8: Note that since \bar{x}_n and \bar{w}_n belong to the simplex of probability vectors, we have

$$\|\bar{x}_n - \bar{w}_n\|_{\text{TV}} \sim |\sin(\bar{x}_n, \bar{w}_n)| = |\sin(x_n, w_n)| = \frac{|x_n \wedge w_n|}{|x_n| \cdot |w_n|},$$

where $a_n \sim b_n$ means that a_n/b_n and b_n/a_n are bounded by a deterministic constant. After taking logarithm, we get that $\log \|\bar{x}_n - \bar{w}_n\|_{\text{TV}}$ can be written as

$$\log |x_n \wedge w_n| - \log |x_n| - \log |w_n| + O(1), \quad (82)$$

where $O(1)$ is bounded by a deterministic constant.

To deal with the second and third terms of (82), we use Lemma 9, from which we get for any strictly positive initial vectors $x, w > 0$ almost surely

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |x_n| = \lambda_1 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log |w_n| = \lambda_1. \quad (83)$$

To deal with the first term of (82), we use Lemma 10 implying that for all initial pairs $(x, w) \in \mathbb{R}_+^p \times \mathbb{R}_+^p$, except for a set of Lebesgue measure zero, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log |(x_n \wedge w_n)| = \lambda_1 + \lambda_2 \quad \text{a.s.} \quad (84)$$

Moreover, Oseledec's theorem implies that for all initial pairs $(x, w) \in \mathbb{R}^p \times \mathbb{R}^p$ the LHS of (84) exists, and it is majorized by the RHS w.p.1. Combining these facts with (82), we immediately get Theorem 8. \blacksquare

APPENDIX C PUSH-SUM ALGORITHM

Proof of Lemma 23: The basic idea is what is called flooding. A convenient reference is [2, Lemma 4.2], the conditions of which can be readily verified, implying that there exist an N such that $p := P(A_N \cdots A_1 > 0) > 0$, where the strict inequality is meant entry wise. It follows that for any $m \geq 1$, we have $P(A_{mN} \cdots A_1 \not\geq 0) \leq (1-p)^m$, and the claim follows by a Borel–Cantelli argument. \blacksquare

Proof of Lemma 24: The following proof relies on a combination of [9] and Theorem 8. Note that our conditions are identical with those of [9], except that in [9], $\alpha^{ji} = 1/2$ for all $(j, i) \in E$ and $w = \mathbf{1}$ were assumed. It is easily seen that the analysis of [9, Th. 3] carries over for general $w \geq 0, w \neq 0$ and $\alpha^{ji} \in (0, 1)$. In particular, setting $s_n = \mathbf{1}^\top x_n$ and $t_n = \mathbf{1}^\top w_n$, we get by a straightforward extension of [9, Th. 3] : for any vector of initial values $x \in \mathbb{R}^p$, and a nonnegative vector of initial weights $w \in \mathbb{R}_+^p$ such that $w \neq 0$, we have for all $i = 1, \dots, p$ a.s.

$$\lim_{n \rightarrow \infty} \frac{x_{ni}}{w_{ni}} = \lim_{n \rightarrow \infty} \frac{s_n}{t_n} \cdot \frac{\bar{x}_{ni}}{\bar{w}_{ni}} = x^* \quad (85)$$

for some random x^* . In fact, the convergence is at least exponential with a deterministic rate: for all $i = 1, \dots, p$ a.s.

$$\frac{s_n}{t_n} \cdot \frac{\bar{x}_{ni}}{\bar{w}_{ni}} = x^* + O(e^{-\alpha n}). \quad (86)$$

It follows by a simple convexity argument (see the proof of Corollary 17) that we also have $s_n/t_n \rightarrow x^*$ a.s. exponentially fast with the same rate

$$\frac{s_n}{t_n} = x^* + O(e^{-\alpha n}) \quad \text{a.s.} \quad (87)$$

In addition, x^* is a convex combination of the initial ratios x_k/w_k . It follows that choosing $x, w > 0$, we will have $x^* > 0$.

Hence, dividing (86) by (87), we get for all $i = 1, \dots, p$ a.s.

$$\frac{\bar{x}_{ni}}{\bar{w}_{ni}} = 1 + O(e^{-\alpha n}). \quad (88)$$

From this, the exponential decay of the total variation distance of \bar{x}_{ni} and \bar{w}_{ni} immediately follows: multiplying both sides of (88) by $0 < \bar{w}_{ni} \leq \max_i w_i$, followed by summation over i gives the a.s. asymptotics

$$|\bar{x}_{ni} - \bar{w}_{ni}| = O(e^{-\alpha n}) \quad \text{and} \quad \|\bar{x}_n - \bar{w}_n\|_{\text{TV}} = O(e^{-\alpha n}),$$

and hence, for all strictly positive pairs (x, w) , we get

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \|\bar{x}_n - \bar{w}_n\|_{\text{TV}} < 0 \quad \text{a.s.} \quad (89)$$

But the LHS is equal to $-(\lambda_1 - \lambda_2)$ a.s. for Lebesgue-almost all $(x, w) \in \mathbb{R}_+^p \times \mathbb{R}_+^p, x, w \neq 0$ by Theorem 8 with \limsup

replaced by \lim . Thus, $-(\lambda_1 - \lambda_2) < 0$ follows as stated in the lemma. \blacksquare

APPENDIX D M_n^{ik}/M_n^{jk} IS SUBEXPONENTIAL

We first provide an elementary *a priori* estimate of M_n^{ik}/M_n^{jk} using the following lemma, a variant of which has been stated by Bellman (see [17] and [33]).

Lemma 43: Let M, B , and X be $p \times p$ matrices such that $M = BX$. Assume that B is strictly positive, and X is a non-negative, allowable matrix. Then, M is strictly positive, and for any fixed pair of row indices (i, j) and any column index k , we have

$$\min_r \frac{B^{ir}}{B^{jr}} \leq \frac{M^{ik}}{M^{jk}} \leq \max_r \frac{B^{ir}}{B^{jr}}.$$

Proof of Lemma 43: The (i, k) and the (j, k) element of M can be expressed as

$$M^{ik} = \sum_r B^{ir} X^{rk} \quad \text{and} \quad M^{jk} = \sum_r B^{jr} X^{rk}.$$

It is easily seen that the ratio M^{ik}/M^{jk} , i.e.,

$$\frac{M^{ik}}{M^{jk}} = \frac{\sum_r B^{ir} X^{rk}}{\sum_r B^{jr} X^{rk}}$$

can be written as a convex combination of B^{ir}/B^{jr} with weights

$$\mu_r = B^{jr} X^{rk} / \sum_s B^{js} X^{sk}$$

which implies the claim. \blacksquare

Lemma 44: Under any set of conditions given in Theorem 12, 14, and 16, it holds that for any pair of row indices i and j and any column index k , the quotient M_n^{ik}/M_n^{jk} is subexponential.

Proof of Lemma 44: In order to apply Lemma 43, let us first extend the sequence (A_n) for $n \leq 0$, with eventual extension of the underlying probability space so that we get a two-sided strictly stationary, ergodic sequence, or even i.i.d. sequence in the case of Theorem 12. Recall the definition of the index of backward sequential primitivity

$$\rho_n = \min\{\rho \geq 0 : A_n A_{n-1} \dots A_{n-\rho+1} > 0\}.$$

Note that under any set of conditions given in Theorems 12, 14, and 16, we can claim that $\mathbb{E}\rho_n < \infty$. Indeed, under the conditions of Theorem 12, $\mathbb{E}\rho_n < \infty$ follows from Lemma 7. On the other hand, $\mathbb{E}\rho_n < \infty$ follows from the condition $\mathbb{E}\psi_n < \infty$, that was *a priori* assumed to hold in the case of Theorems 14 and 16, due to Lemma 5. Consider now the sets

$$\Omega_n^G = \{\omega : \rho_n \leq n\} \quad \text{and} \quad \Omega_n^{Gc} = \Omega \setminus \Omega_n^G.$$

Note that $\mathbb{E}\rho_n < \infty$ implies that

$$\sum_{n=1}^{\infty} P(\Omega_n^{Gc}) = \sum_{n=1}^{\infty} (1 - P(\Omega_n^G)) = \sum_{n=1}^{\infty} P(\rho_n > n) < \infty$$

and thus, Ω_n^{Gc} occurs finitely many times w.p.1 by the Borel-Cantelli lemma. Equivalently, the set

$$\Omega^{Gc} := \limsup_{n \rightarrow \infty} \Omega_n^{Gc} = \bigcap_{m \geq 1} \bigcup_{n \geq m} \Omega_n^{Gc} \quad (90)$$

has measure 0, and consequently, its complement

$$\Omega^G := \liminf_{n \rightarrow \infty} \Omega_n^G = \bigcup_{m \geq 1} \bigcap_{n \geq m} \Omega_n^G \quad (91)$$

on which $\rho_n \leq n$ for $n \geq m(\omega)$ for some random $m(\omega)$, has probability 1. On the set Ω_n^G consider the following decomposition of M_n by separating a strictly positive factor B_n on the left

$$M_n = A_n A_{n-1} \dots A_{n-\rho_n+1} \tilde{M}_n = B_n \tilde{M}_n. \quad (92)$$

Let $\beta'_n = \sum_{k,l} A_n^{kl}$. Obviously, β'_n is equivalent to $\beta_n = \max_{k,l} A_n^{kl}$, and also to $\|A_n\|$, i.e., $\beta'_n \sim \beta_n \sim \|A_n\|$. Then, a simple crude estimator of $\min_r B_n^{ir}/B_n^{jr}$ can be obtained on the set Ω_n^G , with α_n defined under (28), as follows:

$$\frac{\prod_{m=n-\rho_n+1}^n \alpha_m}{\prod_{m=n-\rho_n+1}^n \beta'_m} \leq \frac{B_n^{ir}}{B_n^{jr}} \leq \frac{\prod_{m=n-\rho_n+1}^n \beta'_m}{\prod_{m=n-\rho_n+1}^n \alpha_m}. \quad (93)$$

Obviously, the lower bound is the reciprocal of the upper bound. We will estimate the latter as previously mentioned. From the inequality (93), we get on Ω_n^G

$$\log^+ \frac{B_n^{ir}}{B_n^{jr}} \leq \sum_{m=n-\rho_n+1}^n \log^+ \beta'_m - \sum_{m=n-\rho_n+1}^n \log^- \alpha_m =: \pi_n.$$

Note that the middle term, and thus, π_n , is actually well-defined on all Ω (since m can take on negative values), and obviously, their distributions are independent of n .

Claim: Under any set of conditions given in Theorems 12, 14, and 16, it holds that $\mathbb{E}\pi_n < \infty$.

For the proof in the case of Theorem 12, note that ρ_n is a stopping time for the backward process with finite expectation. In addition, $\mathbb{E} \log^+ \beta'_n < \infty$. Moreover $\mathbb{E} \log^- \alpha_n > -\infty$, by Condition 11. Since $\log^+ \beta'_n$ and $\log^- \alpha_n$ form i.i.d. sequences, we get by Wald's theorem

$$\begin{aligned} & \mathbb{E} \left(\sum_{m=n-\rho_n+1}^n \log^+ \beta'_m - \sum_{m=n-\rho_n+1}^n \log^- \alpha_m \right) \\ &= \mathbb{E} \rho_n \cdot \mathbb{E} \log^+ \beta'_1 - \mathbb{E} \rho_n \cdot \mathbb{E} \log^- \alpha_1 < \infty. \end{aligned} \quad (94)$$

The proof for the case of Theorem 14, in which the positive elements of A_n are assumed to be bounded from below by a positive bound α and from above by β , is trivial. We have

$$\begin{aligned} & \mathbb{E} \left(\sum_{m=n-\rho_n+1}^n \log^+ \beta'_m - \sum_{m=n-\rho_n+1}^n \log^- \alpha_m \right) \\ & \leq \mathbb{E} \rho_n \cdot \log^+(p^2 \beta) - \mathbb{E} \rho_n \cdot \log^- \alpha < \infty. \end{aligned} \quad (95)$$

Finally, consider the less trivial case of Theorem 16, in which the positive elements of A_n may spread all over \mathbb{R}_+ . Setting $\lambda := \mathbb{E} \log^+ \beta'_n$, and noting that $(\log^+ \beta'_n)$ is ergodic, the random variable defined by

$$C_n(\omega, \varepsilon) = \max_{k \geq 0} \left(\sum_{m=n-k}^n (\log^+ \beta'_m - \lambda - \varepsilon) \right)^+ \quad (96)$$

is finite w.p.1 for any $\varepsilon > 0$. Obviously, by rearranging, we have

$$\sum_{m=n-\rho_n+1}^n \log^+ \beta'_m \leq C_n(\omega, \varepsilon) + (\lambda + \varepsilon) \rho_n. \quad (97)$$

We can proceed with the estimation of $\sum_{m=n-\rho_n+1}^n \log^- \alpha_m$ analogously. Under the conditions of Theorem 16, we have $\mathbb{E}\rho_n = \mathbb{E}\rho_1 < \infty$. Obviously, $(C_n(\omega, \varepsilon))$ is a strictly stationary sequence; therefore, to complete the proof of the Claim, it is sufficient to prove that $\mathbb{E}C_n(\omega, \varepsilon) < \infty$. For any fixed

n , this follows directly from the following lemma by setting $\xi_k = \log^+ \beta'_{n-k} - \lambda - \varepsilon$.

Lemma 45: Let $(\xi_k), k \geq 1$ be a strictly stationary, ergodic process such that $\mathbb{E}\xi_k =: -c > 0$. Define

$$\eta = \max_{m \geq 1} \left(\sum_{k=1}^m \xi_k \right)^+. \quad (98)$$

Assume that (ξ_k) is M -mixing of order q with some $q > 4$. Then, $\mathbb{E}\eta < \infty$.

Proof of Lemma 45: For any $x > 0$, we have

$$\begin{aligned} P(\eta \geq x) &\leq \sum_{m=1}^{\infty} P\left(\sum_{k=1}^m \xi_k \geq x\right) \\ &= \sum_{m=1}^{\infty} P\left(\sum_{k=1}^m (\xi_k + c) \geq x + mc\right). \end{aligned} \quad (99)$$

The m th term on the RHS can be bounded from above by using Markov's inequality for the q th absolute moment and the condition that (ξ_k) is M -mixing of the order q as follows:

$$\begin{aligned} \frac{C_q m^{q/2}}{(x + mc)^q} &= \frac{C_q m^{q/2}}{c^q (x/c + m)^q} \leq \frac{C_q (x/c + m)^{q/2}}{c^q (x/c + m)^q} \\ &= \frac{C_q}{c^q (x/c + m)^{q/2}} \end{aligned} \quad (100)$$

with some $q > 4$. Thus, the sum over m on the RHS of (99) can be majorized, by noting that the RHS of (100) are monotone decreasing, as follows:

$$\begin{aligned} \sum_{m=1}^{\infty} \frac{C_q}{c^q (x/c + m)^{q/2}} &\leq \int_0^{\infty} \frac{C_q}{c^q (x/c + t)^{q/2}} dt \\ &= \int_{x/c}^{\infty} \frac{C_q}{c^q t^{q/2}} dt = \frac{-C_q}{c^q (-q/2 + 1)} \left(\frac{x}{c}\right)^{-q/2+1}. \end{aligned} \quad (101)$$

Summing through the positive integers $x = n$, and recalling that $q > 4$, we conclude that

$$\sum_{n=1}^{\infty} P(\eta \geq n) \leq \sum_{n=1}^{\infty} \frac{C_q}{c^q (q/2 - 1)} \left(\frac{n}{c}\right)^{-q/2+1} < \infty \quad (102)$$

hence, $\mathbb{E}\eta < \infty$, as stated in the lemma. ■

Having proved $\mathbb{E}\pi_n < \infty$, we conclude that π_n is sublinear on Ω , implying that B_n^{ir}/B_n^{jr} is subexponential a.s. on Ω^G for any pair (i, j) and any r . This, in combination with Lemma 43 yields the proof of Lemma 44. ■

ACKNOWLEDGMENT

B. Gerencs r would like to thank A. Ozdaglar for encouraging him to study the push-sum algorithm with packet loss and J. M. Hendrickx for an ongoing inspiring collaboration in consensus problems in general. The authors would also like to thank G. Picci for helpful preliminary discussions on the subject and their anonymous reviewers for numerous useful comments and for calling their attention to additional relevant papers.

REFERENCES

[1] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. IEEE 44th Annu. Symp. Found. Comput. Sci.*, 2003, pp. 482–491.

[2] F. Benezit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *Proc. IEEE Int. Symp. Inf. Theory*, 2010, pp. 1753–1757.

[3] A. G. Dimakis, A. D. Sarwate, and M. J. Wainwright, "Geographic gossip: Efficient aggregation for sensor networks," in *Proc. 5th Int. Conf. Inf. Process. Sensor Netw.*, 2006, pp. 69–76.

[4] C. N. Hadjicostis and T. Charalambous, "Asynchronous coordination of distributed energy resources for the provisioning of ancillary services," in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput.*, 2011, pp. 1500–1507.

[5] D. Kempe and F. McSherry, "A decentralized algorithm for spectral analysis," *J. Comput. Syst. Sci.*, vol. 74, no. 1, pp. 70–83, 2008.

[6] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, "Push-sum distributed dual averaging for convex optimization," in *Proc. IEEE 51st Annu. Conf. Decis. Control*, 2012, pp. 5453–5458.

[7] J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, "Distributed anonymous discrete function computation," *IEEE Trans. Autom. Control*, vol. 56, no. 10, pp. 2276–2289, Oct. 2011.

[8] F. Benezit, A. G. Dimakis, P. Thiran, and M. Vetterli, "Order-optimal consensus through randomized path averaging," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5150–5167, Oct. 2010.

[9] B. Gerencs r and J. M. Hendrickx, "Push sum with transmission failures," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1019–1033, Mar. 2019.

[10] G. Picci and T. J. Taylor, "Almost sure exponential convergence to consensus of random gossip algorithms," *Int. J. Robust Nonlinear Control*, vol. 23, no. 9, pp. 1033–1045, 2013.

[11] F. Iutzeler, P. Ciblat, and W. Hachem, "Analysis of sum-weight-like algorithms for averaging in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2802–2814, Jun. 2013.

[12] D. Silvestre, J. P. Hespanha, and C. Silvestre, "Broadcast and gossip stochastic average consensus algorithms in directed topologies," *IEEE Control Netw. Syst.*, vol. 6, no. 2, pp. 474–486, Jun. 2019.

[13] C. N. Hadjicostis and T. Charalambous, "Average consensus in the presence of delays in directed graph topologies," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 763–768, Mar. 2014.

[14] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," *Proc. Amer. Math. Soc.*, vol. 14, no. 5, pp. 733–737, 1963.

[15] J. Liu and A. S. Morse, "Asynchronous distributed averaging using double linear iterations," in *Proc. IEEE Amer. Control Conf.*, 2012, pp. 6620–6625.

[16] A. Tahbaz-Salehi and A. Jadbabaie, "Consensus over ergodic stationary graph processes," *IEEE Trans. Autom. Control*, vol. 55, no. 1, pp. 225–230, Jan. 2010.

[17] H. Furstenberg and H. Kesten, "Products of random matrices," *Ann. Math. Statist.*, vol. 31, no. 2, pp. 457–469, 1960.

[18] R. Jungers, *The Joint Spectral Radius: Theory and Applications*, vol. 385. Berlin, Germany: Springer, vol. 385, 2009.

[19] V. I. Oseledec, "A multiplicative ergodic theorem. Lyapunov characteristic number for dynamical systems," *Trans. Moscow Math. Soc.*, vol. 19, pp. 197–231, 1968.

[20] M. S. Raghunathan, "A proof of Oseledec's multiplicative ergodic theorem," *Israel J. Math.*, vol. 32, no. 4, pp. 356–362, 1979.

[21] R. Atar and O. Zeitouni, "Lyapunov exponents for finite state nonlinear filtering," *SIAM J. Control Optim.*, vol. 35, no. 1, pp. 36–55, 1997.

[22] V. Y. Protasov and A. Voynov, "Sets of nonnegative matrices without positive products," *Linear Algebra Appl.*, vol. 437, no. 3, pp. 749–765, 2012.

[23] H. Flanders, *Differential Forms With Applications to the Physical Sciences*. Dover Books on Mathematics, 1989.

[24] J. N. Tsitsiklis and V. D. Blondel, "Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate," *Math. Control, Signals Syst.*, vol. 10, no. 1, pp. 31–40, 1997.

[25] V. D. Blondel and J. N. Tsitsiklis, "A survey of computational complexity results in systems and control," *Automatica*, vol. 36, no. 9, pp. 1249–1274, 2000.

[26] Y. Peres, "Domains of analytic continuation for the top Lyapunov exponent," *Annales de l'Institut Henri Poincar e, Probabilit es et Statistiques*, vol. 28, no. 1, pp. 131–148, 1992.

[27] E. Seneta, *Non-Negative Matrices and Markov Chains*, 2nd ed. Berlin, Germany: Springer, 2006.

[28] R. Cavazos-Cadena, "An alternative derivation of Birkhoff's formula for the contraction coefficient of a positive matrix," *Linear Algebra Appl.*, vol. 375, pp. 291–297, 2003.

- [29] J. M. Hammersley and D. J. Welsh, "First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory," in *Bernoulli 1713, Bayes 1763, Laplace 1813*. Berlin, Germany: Springer, 1965, pp. 61–110.
- [30] J. F. Kingman, "The ergodic theory of subadditive stochastic processes," *J. Roy. Statist. Soc., Ser. B*, vol. 30, pp. 499–510, 1968.
- [31] J. F. Kingman, "Subadditive processes," in *Ecole d'Été de Probabilités de Saint-Flour V-1975*. Berlin, Germany: Springer, 1976, pp. 167–223.
- [32] J. M. Steele, *Probability Theory and Combinatorial Optimization*, vol. 69. Philadelphia, PA, USA: SIAM, 1997.
- [33] R. Bellman *et al.* "Limit theorems for non-commutative operations. I," *Duke Math. J.*, vol. 21, no. 3, pp. 491–500, 1954.



Balázs Gerencsér received the M.Sc. and Ph.D. degrees in mathematics from the Eötvös Loránd University, Budapest, Hungary, in 2007 and 2013, respectively.

He was a Visiting Student Researcher with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, with the support of a Fulbright Student Scholarship, in 2009–2010. During his Ph.D. studies, he was a Research Assistant with the

Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences, from 2011 to 2013. After graduation, he held a postdoctoral position with the Université catholique de Louvain, Belgium, from 2013 to 2015. Since 2015, he has been a Postdoctoral Fellow with the Alfréd Rényi Institute of Mathematics, then affiliated to the Hungarian Academy of Sciences, currently affiliated to the Eötvös Loránd Research Network, Budapest, Hungary, tenured in 2020, and also an Assistant Professor with the Department of Probability Theory and Statistics, Eötvös Loránd University. His research interests include Markov chains and mixing times, consensus problems and their applications, Markov chains in random environments, synchronizing automata, and mathematical genetics.

Dr. Gerencsér was the recipient of the Gyula Farkas Commemorative Prize of the János Bolyai Mathematical Society awarded to young researchers in some field of applied mathematics in 2014.



László Gerencsér received the M.Sc. and Doctorate degree in mathematics from the Eötvös Loránd University, Budapest, Hungary, in 1969 and 1970, respectively.

Since 1970, he has been with the Institute for Computer Science and Control, then affiliated to the Hungarian Academy of Sciences, currently affiliated to the Eötvös Loránd Research Network, Budapest, Hungary, interrupted with long-term academic visits to Moscow, Oxford, London, and Göteborg. In addition, he was a

Visiting Professor with the Department of Electrical Engineering, McGill University, Montreal, QC, Canada, from 1998 to 1991. His research interests include system identification, stochastic adaptive control, randomization methods in optimization, hidden Markov-models, financial mathematics, change detection, continuous-time linear stochastic systems, adaptive input design, linear stochastic Lévy-systems, and estimation of point processes.

Dr. Gerencsér was the recipient of a Széchenyi Professorship by Eötvös Loránd University, from 1997 to 2001 and the Hart Prize for Independent Research, awarded by The Johns Hopkins University, 2002. He served as an Associate Editor for the *SIAM Journal of Control and Optimization*, 1998–2004, and for the *IEEE TRANSACTIONS ON AUTOMATIC CONTROL*, 2001–2004. He was also the recipient of the title of Doctor of the Hungarian Academy of Sciences in 2000.