

Computable convergence rate bound for ratio consensus algorithms

Balázs Gerencsér

Abstract—The objective of the paper is to establish a computable upper bound for the almost sure convergence rate for a class of ratio consensus algorithms defined via column-stochastic matrices. Our result extends the works of Iutzeler et al. [1] on similar bounds that have been obtained in a more restrictive setup with limited conclusions. The present paper complements the results of Gerencsér and Gerencsér [2], identifying the exact almost sure convergence rate of a wide class of ratio consensus algorithms in terms of a spectral gap, which is, however, not computable in general. The upper bound provided in the paper will be compared to the actual rate of almost sure convergence experimentally on a range of modulated random geographic graphs with random local interactions.

I. INTRODUCTION

Ratio consensus algorithms were initially proposed in a special form by Kempe et al. [3] under the name *push-sum*, with its scope being extended later in [4] under the name *weighted gossip*. The basic setup of these algorithms is a directed graph or network with values associated to each node. The objective is the design of a communication protocol for the computation of the average of the initial input values given at the nodes, using only local, directed, possibly asynchronous communication. Ratio consensus algorithms became the building blocks of further methods requiring distributed computation, such as the analysis of sensors networks [5], the spectral analysis of a network [6] or distributed optimization [7], just to highlight a few.

For the sake of historical context note that ratio consensus is an extension of classic *gossip algorithms* for average consensus, see [8], [9], in which the graph is not directed, and updating the values is restricted to a randomly chosen communicating pair of nodes, replacing their values by the average. Gossip algorithms are linear: updates are defined via (left-)multiplication by a *doubly stochastic* random matrix.

As soon as real-life communication conditions are included in the analysis, additional care is needed as is the case for packet loss [10] where the large size of the network allows controlling the error, or in case of delay [11], where this delay needs to be bounded by spectral properties to achieve average consensus.

The exponential rate of convergence in mean square sense for gossip algorithms, with i.i.d. selection of communicating

pairs, has been determined in [12]. A significant advance, assuming strictly stationary edge selection was presented in [13] establishing almost sure (a.s.) exponential rate of convergence via a spectral gap in the context of Oseledec’s multiplicative ergodic theorem. There is a vast literature for in-depth understanding of such algorithms, for a wider perspective including, e.g., distributed optimization and further references we refer to the survey [14].

Ratio consensus algorithms were designed for possibly asynchronous communication protocols on a directed graph, leading to updates defined via multiplication by a *column stochastic* random matrix, which in itself would fail to reach average consensus. This shortcoming is compensated by running an additional process, allocating weights to each node, with initial weights equal to 1, and considering the quotients value/weight, which are then expected to converge to the required average value for all nodes.

Almost sure convergence of ratio consensus algorithms has been established under a variety of settings see [3] or [4], even for the case of communication protocols with bounded communication delays [15]. However the question on the exact rate of a.s. convergence, raised back in 2010, see [4], was open for a decade.

A partial answer to the question on the a.s. rate of a ratio consensus algorithm was given in Iutzeler et al. [1], providing an upper bound along an unspecified, infinite subset of the timeline. More recently, the paper of Gerencsér and Gerencsér [2] identified the *exact rate of a.s. convergence* as the spectral gap, in the context of Oseledec’s multiplicative ergodic theorem, of the associated random matrix process under very general conditions. However, the spectral gap is known to be *uncomputable* in general [16].

The purpose of this paper is to provide a *computable* upper bound for the rate of a.s. convergence along the full timeline, under technical assumptions that are weaker than those of [1]. This result is obtained by combining arguments of [1], which we simplify and extend with the results of [2]. Apart from a technical tool borrowed from [2] we provide a transparent and self-contained proof.

II. TECHNICAL SETUP AND MAIN RESULT

To describe the technical details in terms of algebraic operations let p be the number of agents, or equivalently, the number of nodes of the communication graph. Let $x_0 \in \mathbf{R}^p$ be a column vector composed of the initial values associated with the nodes in some prefixed order. Our objective is to compute the average $\bar{x} = \sum_{i=1}^p x_0^i / p$. Let $w_0 = \mathbf{1} \in \mathbf{R}^p$ an auxiliary vector, the components of which are called weights. At any time $n \geq 1$ the transmission of an (identical) fraction

B. Gerencsér is with the Alfréd Rényi Institute of Mathematics, Budapest, Hungary and Eötvös Loránd University, Department of Probability and Statistics, Budapest, Hungary, gerencser.balazs@renyi.hu He was supported by NRD (National Research, Development and Innovation Office) grant FK 135711 and KKP 137490, the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and by the ÚNKP-20-5 New National Excellence Program of the Ministry for Innovation and Technology from the source of the NRD Fund.

of values and weights results in updated value and weight vectors as follows:

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

where $(A_n), n \geq 1$ is an i.i.d. sequence of non-negative column-stochastic matrices, implicitly representing all constraints imposed by the network and specifying the local, possibly asynchronous transmissions without packet loss.

The average at agent i at time n is then estimated by the readout x_n^i/w_n^i . A simple interpretation of the algorithm is obtained by thinking of x_n^i/w_n^i as a concentration of a substance in some solvent, properly re-scaled. Note that we can write $\bar{x} = \mathbf{1}^\top x_0/p = \mathbf{1}^\top x_0/(\mathbf{1}^\top w_0)$. Observe also that

$$\mathbf{1}^\top x_n = \mathbf{1}^\top A_n A_{n-1} \cdots A_1 x_0 = \mathbf{1}^\top x_0$$

since the matrices A_k are column-stochastic. Thus the overall average of x_n is conserved, similarly for w_n , thus $\mathbf{1}^\top w_n = p$ for all n . The rate of a.s. convergence is defined as

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i=1}^p \left| \frac{x_n^i}{w_n^i} - \bar{x} \right|. \quad (2)$$

A significant advance over previous works was the determination of a theoretical and tight upper bound for the rate of a.s. convergence under a variety of reasonable conditions, see Theorems 12-19 of [2], thus settling an open problem raised back in 2010, see [4].

In order to clarify the main result to be stated let us revisit the linear algebraic arguments of [1] in preparation for the analysis of $x_n^i/w_n^i - \bar{x}$. Let $M_n = A_n A_{n-1} \cdots A_1$ denote the total effect of the updates on x_0 or w_0 until time n . Let I denote a $p \times p$ identity matrix and let $J = \mathbf{1}\mathbf{1}^\top/p$. Note that $x_n = M_n x_0$ can be decomposed as

$$\begin{aligned} x_n &= M_n J x_0 + M_n (I - J) x_0 = M_n \mathbf{1} \bar{x} + M_n (I - J) x_0 \\ &= w_n \bar{x} + M_n (I - J) x_0. \end{aligned}$$

Thus, at agent i , the ratio consensus algorithm will yield

$$\frac{x_n^i}{w_n^i} = \bar{x} + \frac{e_i^\top M_n (I - J) x_0}{w_n^i}, \quad (3)$$

where e_i is the unit vector with a single 1 at position i .

It follows that the error of x_n^i/w_n^i is largely controlled by behavior of the matrix $N_n := M_n (I - J)$. We can get a useful alternative expression by noting that A_m being column-stochastic implies

$$(I - J) A_m (I - J) = A_m (I - J). \quad (4)$$

Applying this repeatedly for N_n we get the expression

$$N_n = (A_n (I - J)) \cdot (A_{n-1} (I - J)) \cdots (A_1 (I - J)). \quad (5)$$

Theorem 1: Let Assumption 8 given below be satisfied. Then with $\eta_2 = \log \rho(\mathbb{E}(A_1^{\otimes 2})(I - J)^{\otimes 2})$ we have a.s.

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i=1}^p \left| \frac{x_n^i}{w_n^i} - \bar{x} \right| \leq \frac{\eta_2}{2} < 0.$$

The technical results of the paper also complement previous results on the rate of convergence of linear gossip

algorithms, defined via *doubly-stochastic* matrices, which were available both in mean-squared and a.s. sense, as in [12], [13]. For the current statement see Corollary 13.

To enhance readability the necessary technical results of [2] are restated under the specific conditions of the present paper collected in Assumption 8.

III. SEQUENTIAL PRIMITIVITY

In what follows we present the basic technical assumptions needed for the application of the results of [2]. It is intuitively clear that in order to get convergence to the average at all nodes, we need to ensure all-to-all influence. Technically speaking, we should require that the matrix product $A_n \cdots A_1$ is positive for large enough, possibly random n . This leads to the following definition, in a more general and deterministic context, formulated in [17] as follows:

Definition 2: A set \mathcal{A} of $p \times p$ non-negative matrices is *primitive* if a strictly positive product can be formed by some elements of it, repetitions allowed.

Recall that a non-negative matrix is called *allowable* if all rows and all columns contain at least one strictly positive element, see [18]. Now, if the matrices $A_n \in \mathcal{A}$ are chosen according to some random process, we get a natural extension of the notion of primitivity:

Definition 3: A strictly stationary process $(A_n), n \geq 1$ of $p \times p$ non-negative allowable matrices is *sequentially primitive* if $A_\tau A_{\tau-1} \cdots A_1$ is strictly positive for a finite stopping time τ .

Sequential primitivity is easily established for an i.i.d. sequence of matrices by the lemma below:

Lemma 4: Consider a set of $p \times p$ non-negative matrices \mathcal{A} , such that all $A \in \mathcal{A}$ is allowable. Assume that \mathcal{A} is primitive. Let μ be a fully supported distribution on \mathcal{A} , i.e., $\text{supp } \mu = \mathcal{A}$. Consider the i.i.d. sequence $(A_n), n \geq 1$, distributed according to μ . Then (A_n) is sequentially primitive.

The proof will be given in the Appendix. For the sake of historical perspective we note that the following simple alternative sufficient condition for sequential primitivity was given in [4]:

Proposition 5: Let $(A_n), n \geq 1$ be an i.i.d. sequence of $p \times p$ matrices. Assume that A_1 has a strictly positive diagonal almost surely, and $\mathbb{E}A_1$ is irreducible. Then (A_n) is sequentially primitive.

We note in passing that in [1] the conditions of [4] are assumed to be satisfied for \mathcal{A} , and in addition $|\mathcal{A}| < \infty$ is assumed. It is easily seen that the condition that $\mathbb{E}A_1$ is irreducible is in fact necessary for sequential primitivity:

Lemma 6: Let $(A_n), n \geq 1$ be a strictly stationary sequence of non-negative matrices. Assume that (A_n) is sequentially primitive. Then $\mathbb{E}A_1$ is irreducible.

The proof will be given in the Appendix. Note, however, that irreducibility in expectation is not sufficient by itself. For example consider an i.i.d. sequence distributed uniformly on $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. On the other hand, the condition that A_1 has a strictly positive diagonal almost surely is not necessary for sequentially primitivity. For example consider the i.i.d.

sequence of matrices with distributed uniformly on $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, which is readily seen to be sequentially primitive.

Lemma 7: Let $(A_n), n \geq 1$ be a strictly stationary sequence of non-negative matrices. Assume that (A_n) is sequentially primitive. Then for any $k \in \mathbb{Z}^+$, $(A_n^{\otimes k}), n \geq 1$ is also sequentially primitive.

Proof: Observe that for any n

$$A_n^{\otimes k} A_{n-1}^{\otimes k} \cdots A_1^{\otimes k} = (A_n A_{n-1} \cdots A_1)^{\otimes k},$$

hence the left hand side is strictly positive exactly if $A_n A_{n-1} \cdots A_1$ is strictly positive, proving the claim. \blacksquare

The examples above indicate that sequential primitivity may be a fundamental concept, and this is indeed fully justified in [2], see in particular Theorem 19, restated as Proposition 9 below. The conditions of the latter, to be used throughout the paper, can be reformulated as follows:

Assumption 8: Let \mathcal{A} be a set of $p \times p$ matrices, and let $(A_n), n \geq 1$ be a \mathcal{A} -valued stochastic process, satisfying the following conditions:

- \mathcal{A} is a Borel set of non-negative, allowable, column-stochastic matrices.
- \mathcal{A} is primitive.
- $(A_n), n \geq 1$ is an i.i.d. sequence of matrices in \mathcal{A} .
- The distribution of A_1 is fully supported on \mathcal{A} .
- Setting $\alpha_n := \min_{i,j} \{A_n^{i,j} : A_n^{i,j} \neq 0\}$, we have $\mathbb{E} \log^- \alpha_1 > -\infty$.

It is readily seen that the above assumptions on $(A_n), n \geq 1$ are significantly weaker than those in [1], requiring the conditions of [4] to be satisfied for a finite \mathcal{A} .

IV. TIGHT BOUNDS FOR A.S. CONVERGENCE

In this section we highlight the relevant conditions and results of [2], specialized to the context of the present paper. First of all we note that the condition $\mathbb{E} \log^+ \|A_1\| < \infty$ required by the Fürstenberg–Kesten theorem and also by Oseledec’s theorem, restated as Proposition 1 and 2 in [2], and used throughout that paper, is automatically satisfied for column-stochastic matrices. Following these fundamental results, let λ_1 and λ_2 be the first and second largest Lyapunov exponents associated with (A_n) .

The condition of Theorem 8 of [2], serving as a benchmark for subsequent discussion, requiring that A_n is non-negative and allowable for all n , and that the process (A_n) is sequentially primitive, is implied by Assumption 8. Condition 11 of [2], imposing a kind of lower bound on the strictly positive elements of A_n , is required by Assumption 8 in identical form.

Finally, the condition $\lambda_1 - \lambda_2 > 0$, required in the first main result of [2], stated as Theorem 12, is in fact implied by Assumption 8, see Theorem 36 of [2].

Now we are in a position to restate Theorem 19 of [2] in a specialized form, the reference result for identifying the convergence rate of ratio consensus, with $w = \mathbf{1}$ as follows, with $e_i, i = 1, \dots, p$ denoting the i -th unit vector:

Proposition 9: Let Assumption 8 be satisfied. Then for an arbitrary vector of initial values $x \in \mathbb{R}^p$ and initial weights

$w = \mathbf{1}$, we have for all $i = 1, \dots, p$

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

In the current development will need a critical auxiliary technical result on the evolution of the weight vector $w_n = A_n A_{n-1} \cdots A_1 \mathbf{1}$.

Lemma 10: Let Assumption 8 be satisfied. Then $1/\min_i w_n^i$ is sub-exponential:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{\min_i w_n^i} \leq 0.$$

The proof will be given in the Appendix.

V. RESTRICTED CONTRACTION OF $A_n A_{n-1} \cdots A_1$

We will estimate higher order moments of N_n by considering higher order tensor products $N_n^{\otimes k}$ with some $k \in \mathbb{Z}^+$. To set the notations, let $B_n = A_n(I - J)$ and $\|\cdot\|_F$ denote the Frobenius norm of a square matrix. Note that for any square matrix S , the sum of squares of the elements of S , expressing $\|S\|_F^2$, is in fact a sum of selected elements of $S \otimes S$, and thus we can write, with an appropriate linear functional L ,

$$\|S\|_F^2 = L(S \otimes S). \quad (6)$$

Finally, let $\rho(\cdot)$ denote the spectral radius a square matrix.

Lemma 11: Under Assumption 8 we have with $\eta_{2k} = \log \rho(\mathbb{E}(B_1^{\otimes 2k}))$

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \|N_n^{\otimes k}\|_F^2 \leq \eta_{2k} < 0. \quad (7)$$

The lemma above was given in [1] for the case $k = 1$ with a proof, relying on another paper of the authors. Lemma 11 is thus a generalization for all integers k , together with a direct, simple proof. This generalization is also relevant in estimating higher order moments of the error obtained in the course of linear gossip algorithms.

Proof: Taking the $2k$ -th tensor power of (5), followed by taking expectation, recalling that (A_m) is i.i.d., we get

$$\mathbb{E}(N_n^{\otimes 2k}) = \mathbb{E}(B_n^{\otimes 2k}) \cdots \mathbb{E}(B_1^{\otimes 2k}) = (\mathbb{E}(B_1^{\otimes 2k}))^n.$$

From here using (6) we get

$$\mathbb{E} \|N_n^{\otimes k}\|_F^2 = \mathbb{E}(L N_n^{\otimes 2k}) = L \mathbb{E}(N_n^{\otimes 2k}) = L(\mathbb{E}(B_1^{\otimes 2k}))^n.$$

L is a fixed linear functional, thus

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \|N_n^{\otimes k}\|_F^2 &\leq \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log (\|L\| \cdot \|\mathbb{E}(B_1^{\otimes 2k})^n\|) = \eta_{2k}, \end{aligned}$$

using a standard expression of the spectral radius. This confirms the first inequality in (7).

For the second part of the inequality, note that the expectation of the column-stochastic $A_1^{\otimes 2k}$ is itself column-stochastic and the primitivity assumptions provide irreducibility by Lemmas 4 and 6. Therefore the Perron-Frobenius theorem ensures a single maximal eigenvalue with left eigenvector $\mathbf{1}^{\otimes 2k \top}$. Therefore multiplying $\mathbb{E}(A_1^{\otimes 2k})$ by the projection $P_1 = (I^{\otimes 2k} - J^{\otimes 2k})$, which maps $\mathbb{R}^{\otimes 2k}$ into

the orthogonal complement of $\mathbf{1}^{\otimes 2k\top}$ and acts as identity there will result in the stable matrix $\tilde{B} = \mathbb{E}(A_1^{\otimes 2k})P_1$. By the same observation, $P_1(I-J)^{\otimes 2k} = (I-J)^{\otimes 2k}$. Consequently we may express the log spectral radius of interest as

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left\| \left(\mathbb{E}(A_1^{\otimes 2k})P_1(I-J)^{\otimes 2k} \right)^n \right\| \quad (8)$$

Note that (4) can be extended to the tensor power, also inserting P_1 using the invariance observed above, i.e.,

$$(I-J)^{\otimes 2k} \tilde{B} (I-J)^{\otimes 2k} = \tilde{B} (I-J)^{\otimes 2k}.$$

Repeatedly applying this to the product inside the expression of (8) we arrive at

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left\| \tilde{B}^n (I-J)^{\otimes 2k} \right\| \leq \log \rho(\tilde{B}) < 0.$$

Corollary 12: Under Assumption 8,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \left\| N_n^{\otimes k} \right\|_F^2 \leq \eta_{2k} \quad \text{a.s.} \quad (9)$$

Proof: Given the moment bound of Theorem 11, by a standard combination of the Chernoff-inequality and the Borel-Cantelli lemma, for any fixed $l \in \mathbb{Z}^+$ the event $\frac{1}{n} \log \left\| N_n^{\otimes k} \right\|_F^2 > \eta_{2k} + \frac{1}{l}$ occurs finitely many times a.s. which then combined for all $l \in \mathbb{Z}^+$ confirms the claim. ■

Proof: [Proof of Theorem 1] We perform a slight rearrangement so that we can introduce the $2k$ -th power of a single term. For any positive p -tuple of u_i we may write $\log \sum_{i=1}^p u_i \leq \log(p \max_i u_i) = \frac{1}{2k} \log(p \max_i u_i)^{2k}$. For our target expression this translates to

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i=1}^p \left| \frac{x_n^i}{w_n^i} - \bar{x} \right| \leq \limsup_{n \rightarrow \infty} \frac{1}{2kn} \log \max_i \left| \frac{e_i^\top N_n x_0}{w_n^i} \right|^{2k}. \quad (10)$$

To get a hand on this quantity, recall that the denominator is sub-exponential by Lemma 10, thus it does not alter the rate. For the numerator, there holds for some $c_{2k} > 0$

$$\begin{aligned} |e_i^\top N_n x_0|^{2k} &\leq c_{2k} \|e_i^\top\|^{2k} \|N_n\|_F^{2k} \|x_0\|^{2k} \\ &= c_{2k} \|e_i^\top\|^{2k} \|N_n^{\otimes k}\|_F^2 \|x_0\|^{2k}, \end{aligned} \quad (11)$$

using that for any S , both $\|S^{\otimes k}\|_F^2, \|S\|_F^{2k}$ express the sum of all k -fold products of the squared elements of S and are thus equal. Plugging this back to (10) and using the result of Corollary 12 we get the upper bound of $\eta_{2k}/2k$ on the rate. Set $k = 1$ to confirm the claim. ■

Combining Lemma 11 with (11) above we get a $2k$ -moment convergence rate bound for *linear* consensus:

Corollary 13: Under Assumption 8 further requiring A_n to be doubly stochastic there holds

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \|M_n x_0 - \bar{x} \mathbf{1}\|^{2k} \leq \eta_{2k}.$$

VI. OPTIMIZING THE TENSOR EXPONENT k

As we have seen in the proof of Theorem 1, the main term in (11) quantifying the error becomes $\|N_n\|_F^{2k}$ once k is chosen, which is then further bounded with the tools obtained before. Directly examining $\|N_n\|_F^{2k}$ we would get the so-called $s = 2k$ -th mean Lyapunov exponent that could be defined for any $s > 0$, see [19], as

$$\lambda^s = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \|B_n \cdots B_1\|^s. \quad (12)$$

It is easy to see that the limit on the right hand side does exist, and the function λ^s is convex in s , and λ^s/s is monotone non-decreasing. We now show that the same holds as a discrete series for the computable bound $\eta_{2k}/(2k)$, implying $k = 1$ is optimal, in line with the choice in Theorem 1.

Towards showing the (mid-point) convexity of η_{2k} we present an inequality of general interest, a Cauchy-Schwartz type comparison for tensor products. Among various versions available in the literature, the current proof is significantly shorter than the one found in [20]. The main ideas are built on [21], referring to [22], but now interpreted in a simple way that is sufficient for our finite dimensional setting without needing to delve into the operator space theoretical context.

Lemma 14: Let us consider random matrices X and Y . Then there is a constant $C > 0$ depending on the dimensions such that

$$\|\mathbb{E}(X \otimes Y)\| \leq C \sqrt{\|\mathbb{E}(X \otimes X)\|} \cdot \sqrt{\|\mathbb{E}(Y \otimes Y)\|}.$$

For square matrices we further have

$$\rho(\mathbb{E}(X \otimes Y)) \leq \sqrt{\rho(\mathbb{E}(X \otimes X))} \cdot \sqrt{\rho(\mathbb{E}(Y \otimes Y))}.$$

Proof: We prove the claim with $C = 1$ for a special choice of norms, then the claim follows by the equivalence of norms. The construction is indirect at first, we handle an element Z expressed as a mixture of tensor products by some auxiliary measure μ on some auxiliary set S , with proper choice of X, Y , i.e., assume $Z = \int_S X \otimes Y d\mu$. Define the norm of such a mixture as

$$\left\| \int_S X \otimes Y d\mu \right\|_* = \sup_{\substack{\|a_x\| = \|b_x\| = \\ \|a_y\| = \|b_y\| = 1}} \left\{ \int_S a_x^\top X b_x b_y^\top Y^\top a_y d\mu \right\},$$

where a_x, b_x, a_y, b_y are vectors of appropriate dimensions. This is a general scheme for all tensor product spaces encountered.

First we need to check this is a norm indeed. It is defined: all big matrices can be expressed as mixtures of tensor products. It is well defined: observe that for any a_x, b_x, a_y, b_y the integral depends only on the value, not the representation, then taking supremum keeps this property. Linearity is straightforward from the definition. For the triangle inequality we use the freedom of representation, w.l.o.g we may express an addition $Z = Z_1 + Z_2$ as merging disjoint representations, i.e., take $S = S_1 \cup S_2$ with $S_1 \cap S_2 = \emptyset$ with X, Y, μ merging the action of X_i, Y_i, μ_i . Then by construction

$$\int_S X \otimes Y d\mu = \int_{S_1} X_1 \otimes Y_1 d\mu_1 + \int_{S_2} X_2 \otimes Y_2 d\mu_2.$$

Similarly, towards computing the norm we have

$$\int_S a_x^\top X b_x b_y^\top Y^\top a_y d\mu = \int_{S_1} a_x^\top X_1 b_x b_y^\top Y_1^\top a_y d\mu_1 + \int_{S_2} a_x^\top X_2 b_x b_y^\top Y_2^\top a_y d\mu_2.$$

Taking supremum in a_x, b_x, a_y, b_y for the l.h.s. the norm appears there. These are not necessary the optimal parameter vectors for the r.h.s. expressions, immediately confirming the triangle inequality.

Let us turn to our main claim. Notice that the underlying probability space appears naturally for expressing the mixtures to handle. For fixed a_x, b_x, a_y, b_y , the claim is a standard ‘‘Cauchy-Schwartz’’ between the two random scalars $a_x^\top X b_x$ and $b_y^\top Y^\top a_y$, i.e.,

$$\mathbb{E}(a_x^\top X b_x b_y^\top Y^\top a_y) \leq \sqrt{\mathbb{E}(a_x^\top X b_x b_x^\top X^\top a_x)} \cdot \sqrt{\mathbb{E}(a_y^\top Y b_y b_y^\top Y^\top a_y)}.$$

Optimizing in a_x, b_x, a_y, b_y for the l.h.s. we get the inequality needed to conclude the first part on norms. Considering the spectral radius, take n independent copies of X, Y and apply the claim for their products, noting $\mathbb{E}((X_1 \cdots X_n) \otimes (Y_1 \cdots Y_n)) = \mathbb{E}(X \otimes Y)^n$ arriving at

$$\|\mathbb{E}(X \otimes Y)^n\|_* \leq \sqrt{\|\mathbb{E}(X \otimes X)^n\|_*} \cdot \sqrt{\|\mathbb{E}(Y \otimes Y)^n\|_*}.$$

Taking n th root and letting $n \rightarrow \infty$ the spectral radii appear as required. ■

Now we can conveniently apply the above in our context:

Lemma 15: In the setting of Assumption 8, η_{2k} is (mid-point) convex in k . Also, $\eta_{2k}/(2k)$ is non-decreasing.

Proof: By applying Lemma 14 with $X = B^{\otimes k-1}$ and $Y = B^{\otimes k+1}$ for $k \geq 1$ we directly get

$$\exp(\eta_{2k}) \leq \exp(\eta_{2k-2}/2) \exp(\eta_{2k+2}/2),$$

showing the convexity of η_{2k} . To complete the sequence, extend to $\eta_0 = 0$. Lemma 14 still provides convexity at $k = 1$, using the identity matrix when necessary, this easily implies that $\eta_{2k}/(2k)$ must be non-decreasing for $k \geq 1$. ■

VII. NUMERICAL RESULTS

The main question is the sharpness of the upper bound on the exponential convergence rate obtained. We do not launch the processes from a single initial x_0 , but rather from $(I - J)$:

$$\frac{1}{n} \log \left(\frac{1}{\sqrt{p}} \|\text{diag}(w_n)^{-1} M_n (I - J)\|_F \right), \quad (13)$$

bounding the empirical rate of convergence of the process for the range of any starting vector spanned by the columns of $I - J$: those with 0 average. We take $n = 100000$, which is generous in view of the size of the network and the communication pattern to follow. The Julia computing platform is used to carry out the simulations [23], [24].

For the underlying network, we consider a model based on Random Geometric Graphs (RGG) [25] with a simple perturbation where a dependence on the positions of the

agents can be introduced. We interpolate between a grid and uniform random placement. To be more precise, let $c \in [0, 1]$ be an interpolation parameter together with a reference number of nodes, p_0 . For each node i , two preliminary positions are assigned: z_g^i a unique point on the $\sqrt{p_0} \times \sqrt{p_0}$ square grid fitted into $[0, 1]^2$ and z_r^i , a uniform random position in $[0, 1]^2$. The final position is then declared as $z^i = cz_r^i + (1 - c)z_g^i$.

We still need to define the graph on the points obtained on the unit square. We still want to stay with the concept of connecting those that are close.

As the structure of positions are changing, the clear connectivity thresholds for RGGs [26] does not apply anymore. Instead to get a graph with balanced density, we optimize the threshold for the distance of two nodes getting connected so that the largest connected component contains $\approx 90\%$ of the nodes. Then this giant component is kept for further work also determining the final dimension p . In Figure 1 we see two examples for $c = 0.1$ and $c = 0.8$ for $p_0 = 64$ initial points, which will be the default size parameter for our simulations. This leads to a typical dimension of $p \approx 58$.

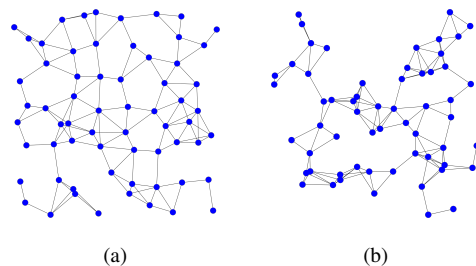


Fig. 1: Example random graphs interpolating between square grid and uniform random node placement, for (a) $c = 0.1$ and (b) $c = 0.8$ coefficient for the random component.

The reference dynamics is asynchronous directed gossip: every step a uniformly chosen node communicates towards a single uniformly chosen neighbor, sending $\frac{1}{2}$ fraction of its value and weight. Figure 2 presents the empirical rate according to (13) together with $\eta_2/2$ for 500 simulations for various c , we see the two move together despite the wild randomness of the graph instances. We also see the difference of the two, showing that it is reliable estimate even point-wise. The numerical stability is demonstrated by a single instance out of the 500 when there is a positive difference of $\approx 2 \cdot 10^{-6}$.

We compare the reference gossip with two modified strategies. First, we have two-way randomized gossip: the transmitting node selects two receivers and uniformly randomly splits the total fraction of $\frac{1}{2}$ to be sent between them. Note that the conditions of Theorem 1, Assumption 8 still holds, here \mathcal{A} is a union of segments in the space of non-negative matrices. Second, the reference gossip is modified to send only a fraction of $\frac{1}{4}$ to a single recipient, but we allow twice as many steps to take place, we will name this slowed gossip for convenience. Figure 3 presents the comparison of rates of

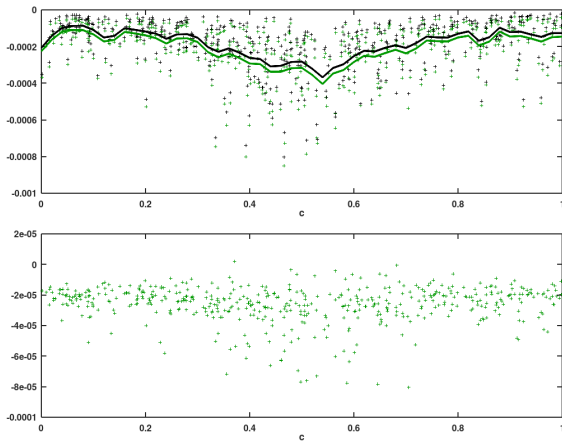


Fig. 2: Empirical (green) rates versus bounding $\eta_2/2$ (black) and their moving averages, below the difference of the two.

the modified strategies. For the empirical rates we observe no consistent ordering of the three strategy, however, the difference is an order of magnitude smaller than the variance caused by the graph variability, see the range in Figure 2. $\eta_2/2$ provides two-way gossip a stronger bound than for the reference process, even more for the slowed gossip.

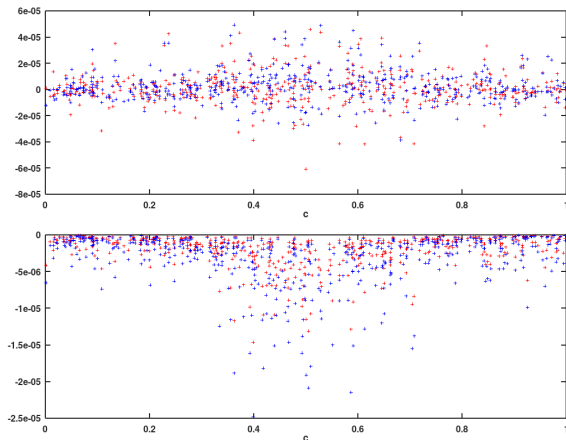


Fig. 3: Empirical rates (above) and $\eta_2/2$ (below) compared to that of the reference gossip on structured RGGs, for two-way randomized gossip (red) and slowed gossip (blue).

Another natural question to ask is the dependence of the rate on the connection structure. We consider the following Erdős-Rényi process inspired model to study this phenomenon: starting with a cycle on $p = 50$ nodes we add random edges uniformly one by one, up to 500 (an average extra degree of 20), and at each step, we evaluate both the empirical rates and $\eta_2/2$ of the three process variants. Figure 4 shows the aggregated picture. It is apparent that initially in the sparse region an extra edge is much more game-changing than later on. Also, in the late phase there is a clear ordering of efficiency of slowed gossip being the best, followed by two-way gossip, then the reference. Note however, this ordering is not fully present in the earlier

sparse, less interconnected phases.

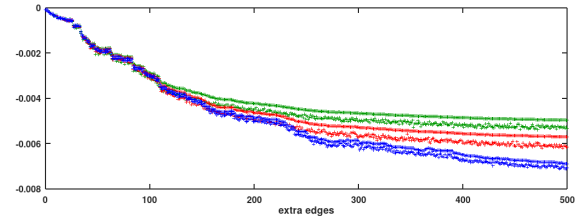


Fig. 4: Empirical rates and $\eta_2/2$ for reference gossip (green) two-way randomized gossip (red) and slowed gossip (blue) on the cycle with random edges.

Let us zoom in once again for this connection structure to compare the three process variants. We compare separately the empirical rates and the computed $\eta_2/2$ for both process variants, normalized against the reference gossip in Figure 5. We observe that in terms of the empirical rate, initially there is a non-trivial race among the three strategies, then between the two-way and the slowed gossip, before the final order is settled for dense graphs. Interestingly, for the bounding $\eta_2/2$ the ordering is robust.

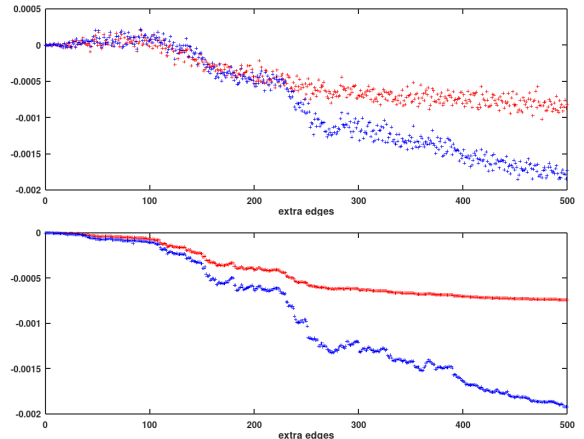


Fig. 5: Empirical rates (above) and $\eta_2/2$ (below) compared to that of the reference gossip on the cycle with random edges, for two-way randomized gossip (red) and slowed gossip (blue).

VIII. CONCLUSION

We have proven upper bounds for the almost sure exponential convergence rate of i.i.d. ratio consensus algorithms inspired by the approach of [1] and by the analysis in [2]. The quantity $\eta_{2k} = \log \rho(\mathbb{E}(A_1^{\otimes 2k})(I - J)^{\otimes 2k})$ is indeed accessible, as it is based on the spectral description of a finite transformation of the matrix distribution describing the updates. We have shown that $\eta_{2k}/(2k)$ is non-decreasing, thus for bounding the convergence rate it is optimal to keep $k = 1$. However, our general results can be applied to provide upper bounds on the convergence rate of higher moments for linear consensus.

Through numerical examples we have observed that the bounds tend to capture well the magnitude of the rate, with an error of lower order. Also, for sparse networks few additional edges can improve efficiency significantly.

IX. APPENDIX

Proof: [of Lemma 4] Define $\gamma : [0, \infty) \rightarrow \{0, 1\}$ simply as the indicator of being positive. This naturally extends to matrices element-wise.

The primitivity of \mathcal{A} is characterized by the primitivity of $\gamma(\mathcal{A})$ as only positivity is needed through the process, without focus on the actual value, and we are working with non-negative matrices.

To investigate the support of the push-forward measure $\gamma_*\mu$, choose any $A \in \mathcal{A}$ and its projection $A^0 = \gamma(A)$. Let $\varepsilon = \min_{i,j} \{A^{i,j} \mid A^{i,j} > 0\}$. For the small ball $B(A, \varepsilon/2)$ we must have $\mu(B(A, \varepsilon/2)) > 0$ as A is in the support by assumption. Observe there are two type of matrices in the ball: \mathcal{B}_1 with some with negative elements, not playing a role, and \mathcal{B}_2 with non-negative matrices where positive elements appear at least where A has them. Therefore $\mu(\mathcal{B}_1) = 0, \mu(\mathcal{B}_2) > 0$, and for any matrix of $B \in \mathcal{B}_2$ we have $\gamma(B) \geq A^0$. This means $\gamma(\mathcal{B}_2)$ maps to a set with positive $\gamma_*\mu$ probability, with all matrices bounded below by A^0 . Consequently, the support of $\gamma(\mathcal{A})$ is majorated by the support of $\gamma_*\mu$, in the sense that for any matrix $A^0 \in \gamma(\mathcal{A})$ there exists $B^0 \in \text{supp}(\gamma_*\mu)$ such that $A^0 \leq B^0$. Note that the two supports are not necessary equal.

Consider now any sequence $A_{k_l} \cdots A_{k_1} > 0$ with $A_{k_i} \in \mathcal{A}$ presenting the primitivity of \mathcal{A} . We know it is equivalent to $\gamma(A_{k_l}) \cdots \gamma(A_{k_1}) > 0$. By the previous argument, we have matrices $\gamma(A_{k_i}) \leq B_i^0 \in \text{supp}(\gamma_*\mu)$, thus $B_l^0 \cdots B_1^0 > 0$. Being in the support on a discrete space means B_i^0 has positive probability to appear, similarly for the chosen product at any l consecutive steps, thus it will eventually occur (as we have an i.i.d. process), confirming sequential primitivity. In the meantime we rely on the matrices being allowable so that it is sufficient to find a positive product at an arbitrary starting time. ■

Proof: [of Lemma 6] We prove by contradiction, let us assume $\mathbb{E}A_1$ is reducible. Without the loss of generality we can assume that $\mathbb{E}A_1$ has the block structure $\begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$ with square blocks in the diagonal. Knowing that A_1 is non-negative, this would force A_1 to have the same structure a.s., then by stationarity for all A_n , and then for their products of any length, contradicting sequential primitivity. Thus $\mathbb{E}A_1$ indeed must be irreducible. ■

Proof: [of Lemma 10] The claim is a direct consequence of a Lemma 44 in [2], the validity of its conditions has been verified in Section IV. The cited lemma states that the product $M_n = A_n A_{n-1} \cdots A_1$ is asymptotically rank-1, more specifically for any fixed pair of rows i, j and any column k the ratio M_n^{ik}/M_n^{jk} is sub-exponential.

As $w_n = M_n \mathbf{1}$, each w_n^i/w_n^j is easily seen to be a convex combination of the corresponding quotients M_n^{ik}/M_n^{jk} , $k = 1, \dots, p$. Thus w_n^i/w_n^j is also sub-exponential. Recall that A_m is column stochastic for all m , and hence M_n is also

column stochastic for all n . Thus we have $\mathbf{1}^\top w_n = p$ for all n . Summation of w_n^i/w_n^j through i , with j fixed, yields p/w_n^j . Since each term is sub-exponential, it follows that p/w_n^j , and its maximum over j , is sub-exponential as well. ■

REFERENCES

- [1] F. Iutzeler, P. Ciblat, and W. Hachem, "Analysis of sum-weight-like algorithms for averaging in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2802–2814, 2013.
- [2] B. Gerencsér and L. Gerencsér, "Tight bounds on the convergence rate of generalized ratio consensus algorithms," *IEEE Transactions on Automatic Control*, 2021. Available online.
- [3] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proceedings of 44th Annual IEEE Symposium on Foundations of Computer Science*, pp. 482–491, 2003.
- [4] F. Bénézit, V. Blondel, P. Thiran, J. N. Tsitsiklis, and M. Vetterli, "Weighted gossip: Distributed averaging using non-doubly stochastic matrices," in *Proceedings of 2010 IEEE International Symposium on Information Theory (ISIT)*, pp. 1753–1757, 2010.
- [5] A. Giridhar and P. Kumar, "Toward a theory of in-network computation in wireless sensor networks," *IEEE Communications Magazine*, vol. 44, no. 4, pp. 98–107, 2006.
- [6] D. Kempe and F. McSherry, "A decentralized algorithm for spectral analysis," *Journal of Computer and System Sciences*, vol. 74, no. 1, pp. 70–83, 2008.
- [7] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2014.
- [8] J. N. Tsitsiklis, *Problems in decentralized decision making and computation*. PhD thesis, Massachusetts Institute of Technology, 1984.
- [9] V. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," in *44th IEEE Conference on Decision and Control*, pp. 2996–3000, 2005.
- [10] P. Frasca and J. M. Hendrickx, "Large network consensus is robust to packet losses and interferences," in *2013 European Control Conference (ECC)*, pp. 1782–1787, IEEE, 2013.
- [11] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.
- [12] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.
- [13] G. Picci and T. J. Taylor, "Almost sure exponential convergence to consensus of random gossip algorithms," *International Journal of Robust and Nonlinear Control*, vol. 23, no. 9, pp. 1033–1045, 2013.
- [14] A. H. Sayed, "Adaptation, learning, and optimization over networks," *Foundations and Trends® in Machine Learning*, vol. 7, no. 4-5, pp. 311–801, 2014.
- [15] C. N. Hadjicostis and T. Charalambous, "Average consensus in the presence of delays in directed graph topologies," *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 763–768, 2014.
- [16] J. N. Tsitsiklis and V. Blondel, "The Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate," *Mathematics of Control, Signals and Systems*, vol. 10, no. 1, pp. 31–40, 1997.
- [17] V. Y. Protasov and A. Voynov, "Sets of nonnegative matrices without positive products," *Linear Algebra and its Applications*, vol. 437, no. 3, pp. 749–765, 2012.
- [18] E. Seneta, *Non-negative matrices and Markov chains*. Springer Science & Business Media, 2006.
- [19] L. Arnold, W. Kliemann, and E. Oeljeklaus, "Lyapunov exponents of linear stochastic systems," in *Lyapunov exponents*, pp. 85–125, Springer, 1986.
- [20] L. Gerencsér and Z. Orlovits, "Lq-stability of products of block-triangular stationary random matrices," *Acta Scientiarum Mathematicarum*, vol. 74, pp. 927–944, 2008.
- [21] G. Pisier, *Introduction to operator space theory*, vol. 294. Cambridge University Press, 2003.
- [22] U. Haagerup, "The Grothendieck inequality for bilinear forms on C*-algebras," *Advances in Mathematics*, vol. 56, no. 2, pp. 93–116, 1985.

- [23] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, 2017.
- [24] J. Fairbanks, M. Besançon, S. Schölly, J. Hoffman, N. Eubank, and S. Karpinski, "JuliaGraphs/Graphs.jl: an optimized graphs package for the Julia programming language," 2021.
- [25] M. Penrose, *Random geometric graphs*, vol. 5. OUP Oxford, 2003.
- [26] P. Gupta and P. R. Kumar, *Critical Power for Asymptotic Connectivity in Wireless Networks*, pp. 547–566. Boston, MA: Birkhäuser Boston, 1999.