Computable Convergence Rate Bound for Ratio Consensus Algorithms

Balázs Gerencsér

Abstract—The objective of this letter 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. from 2013 on similar bounds that have been obtained in a more restrictive setup with limited conclusions. The present paper complements results by Gerencsér and Gerencsér from 2022, 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 this letter will be compared to the actual rate of almost sure convergence experimentally on a range of modulated random geometric graphs with random local interactions.

Index Terms—Agents-based systems, randomized algorithms, stochastic systems.

I. INTRODUCTION

INITIALLY ratio consensus algorithms were proposed in a special form by Kempe et al. [1] under the name push-sum, with its scope being extended later in [2] 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 [3], the spectral analysis of a network [4] or distributed optimization [5], 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 [6], [7], 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.

The exponential rate of convergence in mean square sense for gossip algorithms, with i.i.d. selection of communicating pairs, has been determined in [8]. A significant advance, assuming strictly stationary edge selection was presented in [9] 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, we refer to the survey [10].

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 [1] or [2], even for the case of communication protocols with bounded communication delays [11]. However, the question on the exact rate of a.s. convergence, raised back in 2010, see [2], 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. [12], providing an upper bound along an unspecified, infinite subset of the timeline. More recently, this letter of Gerencsér and Gerencsér [13] 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, this generalized spectral gap is known to be uncomputable in general [14]. Thus there is a mismatch in case one is after a jointly provable and accessible convergence guarantee for application efficiency.

The purpose of this letter 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 [12]. The main contribution is formulated in Section IV, Theorem 1. This result is obtained by combining arguments of [12], which we simplify and extend with the results of [13].
a technical tool borrowed from [13] we provide a transparent and self-contained proof.

II. TECHNICAL SETUP

Let us set some notation used through this letter. For a column vector \(x \in \mathbb{R}^p\), matrix \(A \in \mathbb{R}^{p \times p}\), \(\|x\|\) denotes the 2-norm, \(\|A\|\) the induced norm and \(\|A\|_F\) the Frobenius norm, whereas \(\rho(A)\) stands for the spectral radius. Let \(x^t\) denote the \(t\)th entry of \(x\) and \(A^{i,j}\) the \((i,j)\) element of \(A\). The Kronecker product of \(X \in \mathbb{R}^{p \times q}, Y \in \mathbb{R}^{p' \times q'}\) is defined as \(X \otimes Y \in \mathbb{R}^{p \times q \times p' \times q'}\) with \((X \otimes Y)(i-1)p'+j-1, m-1) = X^{i,j} \cdot Y^{m,n}\), the indices running according to the dimensions of \(X, Y\). Let \(e_i\) be the unit vector with a single 1 at position \(i\) of appropriate dimension. Define \(\log^+(z) = \max(\log z, 0)\), \(\log^-(z) = \min(\log z, 0)\).

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 \(x_0 \in \mathbb{R}^p\) be a column vector composed of the initial values associated with the nodes in some prefixed order. Our objective is to compute the average \(\bar{x} = \sum_{i=1}^p \frac{x_i}{p}\). Let \(w_0 = 1 \in \mathbb{R}^p\) an auxiliary vector, the components of which are called weights. At any time \(n \geq 1\) the transmission of an (identical) fraction of values and their weights results in updated value and weight vectors as follows:

\[
x_n = A_n x_{n-1}, \quad w_n = A_n w_{n-1},
\]

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 \(x_n^i/w_n^i\). Note that we can write \(\bar{x} = 1^\top x_n/p = 1^\top w_0 / (1^\top w_0)\). Observe also that

\[
1^\top x_n = 1^\top A_n A_{n-1} \cdots A_1 x_0 = 1^\top x_0
\]

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

\[
\limsup_{n \to \infty} \frac{1}{n} \log \left| \frac{x_n^i}{w_n^i} - \bar{x} \right|.
\]

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 [13, Ths. 12-19], thus settling an open problem raised back in 2010, see [2].

III. SEQUENTIAL PRIMITIVITY

In what follows we present the basic necessary connectivity assumptions. It is intuitively clear that in order to get convergence to the average at all nodes, we need to ensure all-to-all influence. For the application of the results of [13], a weak quantitative assumption will be also needed. 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 [15] as follows.

**Definition 1:** 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 [16]. 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 2:** A strictly stationary process \((A_n), n \geq 1\) of \(p \times p\) non-negative allowable matrices is **sequentially primitive** if \(A_n A_{n-1} \cdots A_1\) is strictly positive for a finite stopping time \(\tau\).

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

**Lemma 1:** 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 \(\mu\) be a fully supported distribution on \(\mathcal{A}\). Consider the i.i.d. sequence \((A_n), n \geq 1\), distributed according to \(\mu\). Then \((A_n)\) is sequentially primitive.

The proof will be given in the Appendix. For the sake of historical perspective we mention that the following simple conditions of [2] also imply sequential primitivity, as it was shown there.

**Proposition 1:** Let \((A_n), n \geq 1\) be a strictly stationary ergodic 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 [12] the above conditions of [2] are assumed to be satisfied for an i.i.d. sequence and in addition \(|\mathcal{A}| < \infty\) is assumed. It is easily seen that \(\mathbb{E} A_1\) being irreducible is in fact needed for sequential primitivity.

**Lemma 2:** 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.

**Lemma 3:** 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.

The examples above indicate that sequential primitivity may be a fundamental concept, and this is indeed fully justified in [13], see in particular Theorem 19, restated as Proposition 2 below. The conditions of the latter, to be used throughout this letter, can be reformulated as follows.

**Assumption 1:** Let \(\mathcal{A}\) be a set of \(p \times p\) matrices, and let \((A_n), n \geq 1\) be an \(\mathcal{A}\)-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 [12], which also requires positive diagonals and finite \( \mathcal{A} \).

IV. MAIN RESULT

In order to clarify the main result to be stated, let us revisit the linear algebraic arguments of [12] in preparation for the analysis of \( x_n'/w_n' - \bar{x} \). Let \( M_n = A_nA_{n-1} \cdots A_1 \) 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 \( J = I^{1/p} \). Note that \( x_n = M_nx_0 \) can be decomposed as

\[
x_n = M_nx_0 + M_n(I - J)x_0 = M_n1\bar{x} + M_n(I - J)x_0 = w_n\bar{x} + M_n(I - J)x_0.
\]

Thus, at agent \( i \), the ratio consensus algorithm will yield

\[
x_{ni}^{'} = \frac{x_{ni}'}{w_n'} = \bar{x} + \frac{e_{i}^{\top}M_n(I - J)x_0}{w_n'}.
\]

It follows that the error of \( x_{ni}'/w_n' \) is largely controlled by behavior of the matrix \( M_n = A_n(I - J) \). We can get a useful alternative expression by noting that \( A_m \) being column-stochastic implies

\[
(I - J)A_m(I - J) = A_m(I - J).
\]

Applying this repeatedly for \( N_n \) we get the expression

\[
N_n = (A_n(I - J)) \cdot (A_{n-1}(I - J)) \cdots (A_1(I - J)).
\]

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

\[
\limsup_{n \to \infty} \frac{1}{n} \log \sum_{i=1}^{p} \left| \frac{x_{ni}'}{w_n'} - \bar{x} \right| \leq \frac{\eta_2}{2} < 0.
\]

The technical results of this letter 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 [8], [9]. For the current statement see Corollary 2.

V. TIGHT BOUNDS FOR A.S. CONVERGENCE

In this section we highlight the relevant conditions of [13], verify them based on Assumption 1, then restate the corresponding results 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 [13, Proposition 1 and 2], and used throughout that paper, is automatically satisfied for column-stochastic matrices. Following these fundamental results, let \( \lambda_1 \) and \( \lambda_2 \) be the first and second largest Lyapunov exponents associated with \( (A_n), n \geq 1 \).

The conditions of [13, Th. 8], 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 1. Condition 11 of [13], imposing a kind of lower bound on the strictly positive elements of \( A_n \) is required by Assumption 1 in identical form. Finally, the condition \( \lambda_1 - \lambda_2 > 0 \), required in the first main result of [13], stated as Theorem 12, is in fact implied by Assumption 1, see [13, Th. 36].

Now we are in a position to restate [13, Th. 19] in a specialized form, the reference result for identifying the convergence rate of ratio consensus, with \( w = 1 \) as follows:

Proposition 2: Let Assumption 1 be satisfied. Then for an arbitrary vector of initial values \( x \in \mathbb{R}^p \) and initial weights \( w = 1 \), we have for all \( i = 1, \ldots, p \)

\[
\limsup_{n \to \infty} \frac{1}{n} \log \frac{\|e_i^\top M_nx\|}{\|e_i^\top M_n\|} - \bar{x} \leq \lambda_2 < 0 \quad \text{a.s.}
\]

In the current development will need a critical auxiliary technical result on the evolution of the weight vector \( w_n \).

Lemma 4: Let Assumption 1 be satisfied. Then \( 1/\min_i w_n' \) is sub-exponential:

\[
\limsup_{n \to \infty} \frac{1}{n} \log \frac{1}{\min_i w_n'} \leq 0.
\]

The proof will be given in the Appendix.

VI. RESTRICTED CONTRACTION OF \( A_nA_{n-1} \cdots A_1 \)

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

\[
\|S\|^2_F = L(S \otimes S).
\]

Lemma 5: Under Assumption 1 we have with \( \eta_{2k} = \log \rho(\mathbb{E}(B_1^{\otimes 2k})) \)

\[
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E} \left\| N_n^{\otimes k} \right\|_F^2 \leq \eta_{2k} < 0.
\]

The lemma above was given in [12] for the case \( k = 1 \) with a proof, relying on another paper of the authors. Lemma 5 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 (2), followed by taking expectation, recalling that \( (A_n) \) is i.i.d., we get

\[
\mathbb{E}(N_n^{\otimes 2k}) = \mathbb{E}(B_1^{\otimes 2k}) \cdots \mathbb{E}(B_1^{\otimes 2k}) = \left(\mathbb{E}(B_1^{\otimes 2k})\right)^n.
\]

From here using (3) we get

\[
\mathbb{E} \left\| N_n^{\otimes k} \right\|_F^2 = \mathbb{E} \left( L(N_n^{\otimes k}) \right) = L \mathbb{E}(N_n^{\otimes 2k}) = L \left( \mathbb{E}(B_1^{\otimes 2k}) \right)^n.
\]

Let \( L \) is a fixed linear functional, thus

\[
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E} \left\| N_n^{\otimes k} \right\|_F^2 \leq \limsup_{n \to \infty} \frac{1}{n} \log \left(\|L\| \cdot \mathbb{E}(B_1^{\otimes 2k})^n\right) = \eta_{2k},
\]

using a standard expression of the spectral radius. This confirms the first inequality in (4).

For the second part of the inequality, note that the expectation of the column-stochastic \( A_1^{\otimes 2k} \) is itself column-stochastic
For any positive \( k \in \mathbb{N} \) ensure a single maximal eigenvalue with left eigenvector \( \tilde{\mathbf{v}} \) and the primitivity assumptions provide irreducibility by Lemma 1 and 2. Therefore the Perron-Frobenius theorem ensures a single maximal eigenvalue with left eigenvector \( \mathbf{1} \) of \( \mathbf{1} \otimes 2^k \). Therefore multiplying \( \mathbb{E}(\mathbf{A}^{2^k}) \) by the projection \( P_1 = (\mathbf{I} \otimes 2^k - \mathbf{A}^{2^k}) \), which maps \( \mathbb{R} \otimes 2^k \) into the orthogonal complement of \( \mathbf{1} \otimes 2^k \) and acts as identity there will result in the stable matrix \( \tilde{B} = \mathbb{E}(\mathbf{A}^{2^k})P_1 \). By the same observation, \( P_1(I-J)^{2^k} = (I-J)\mathbf{1} \otimes 2^k \). Consequently we may express the log spectral radius of interest as

\[
\limsup_{n \to \infty} \frac{1}{n} \log \left\| \left( \mathbb{E}(\mathbf{A}^{2^k})P_1(I-J)^{2^k} \right)^n \right\| \quad (5)
\]

Note that (1) can be extended to the tensor power, also inserting \( P_1 \) using the invariance observed above, i.e.,

\[
(I-J)^{2^k} \tilde{B}(I-J)^{2^k} = \tilde{B}(I-J)^{2^k}.
\]

Repeatedly applying this to the product inside the expression of (5) to eliminate \((I-J)^{2^k}\) terms we arrive at

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

**Corollary 1:** Under Assumption 1,

\[
\limsup_{n \to \infty} \frac{1}{n} \log \left\| N_n^{2^k} \right\|_F^2 \leq \eta_{2^k} \quad \text{a.s.}
\]

**Proof:** Given the moment bound of Lemma 5, by a standard combination of the Chernoff-inequality and the Borel-Cantelli lemma, for any fixed \( l \in \mathbb{Z}^+ \) the event \( \frac{1}{n} \log \| N_n^{2^k} \|_F \geq \eta_{2^k} + \frac{l}{2} \) occurs finitely many times a.s. which then combined for all \( l \in \mathbb{Z}^+ \) confirms the claim.

**Proof of Theorem 1:** We perform a slight rearrangement so that we can introduce the \( 2^k \)-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

\[
\limsup_{n \to \infty} \frac{1}{n} \log \left\| \sum_{i=1}^p \left| \frac{x_i}{w_n} \right|^2 - \mathbf{1} \right\| \leq \limsup_{n \to \infty} \frac{1}{2kn} \log \max_i \left| \frac{x_i}{w_n} \right|^2 \eta_{2^k}^{2k}.
\]

(6)

To get a hand on this quantity, recall that the denominator is sub-exponential by Lemma 4, thus it does not alter the rate. For the numerator, there holds for some \( c_{2^k} > 0 \)

\[
\left| e_i^T N_n x_0 \right|^{2^k} \leq c_{2^k} \left| e_i^T \right|^{2^k} \| N_n \|_F^{2k} \| x_0 \|_F^{2k} = c_{2^k} \left| e_i^T \right|^{2k} \| N_n \|_F^{2k} \| x_0 \|_F^{2k}.
\]

(7)

via the equivalence of norms and where \( \| N_n \|_F^{2k} \), \( \| N_n \|_F^{2k} \) both express the sum of all \( k \)-fold products of the squared elements of \( N_n \) and are thus equal. Plugging this back to (6) and using the result of Corollary 1 we get the upper bound of \( \eta_{2^k}/2^k \) on the rate. Set \( k = 1 \) to conclude.

Combining Lemma 5 with (7) above we get a \( 2^k \)-moment convergence rate bound for linear consensus.

**Corollary 2:** Under Assumption 1 further requiring \( A_n \) to be doubly stochastic there holds

\[
\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{E} \left\| M_n x_0 - \mathbf{1} \right\|_F^{2^k} \leq \eta_{2^k}.
\]

**VII. Optimizing the Tensor Exponent \( k \)**

As we have seen in the proof of Theorem 1, the main term in (7) quantifying the error becomes \( \| N_n \|_F^{2k} \) once \( k \) is chosen, which is then further bounded with the tools obtained before. By directly examining \( \| N_n \|_F^{2k} \) we would get the so-called \( s = 2^k \)-th mean Lyapunov exponent that could be defined for any \( s > 0 \), see [17], as

\[
\lambda_s = \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left\| \tilde{B}_n \cdots \tilde{B}_1 \right\|_F^s.
\]

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

The core of the (mid-point) convexity of \( \eta_{2^k} \) is a Cauchy-Schwartz type comparison for tensor products. Various versions are available in the literature, see, e.g., [18].

**Proposition 3:** Let us consider random matrices \( X \) and \( Y \). Then there is a constant \( C > 0 \) depending on the dimensions such that

\[
\mathbb{E} \|X \otimes Y\|_2 \leq C \sqrt{\mathbb{E} \|X \otimes X\|_2} \cdot \sqrt{\mathbb{E} \|Y \otimes Y\|_2}.
\]

For square matrices we further have

\[
\rho(\mathbb{E} \|X \otimes Y\|) \leq \sqrt{\rho(\mathbb{E} \|X \otimes X\|)} \cdot \sqrt{\rho(\mathbb{E} \|Y \otimes Y\|)}.
\]

We can conveniently apply the above in our context:

**Lemma 6:** In the setting of Assumption 1, \( \eta_{2^k} \) is (mid-point) convex in \( k \). Also, \( \eta_{2^k}/(2^k) \) is non-decreasing.

**Proof:** By applying Proposition 3 with \( X = B^{\otimes 2^k-1} \) and \( Y = B^{\otimes 2^k+1} \) for \( k \geq 1 \) we directly get

\[
\exp(\eta_{2^k}) \leq \exp(\eta_{2^k-2}/2) \exp(\eta_{2^k+2}/2),
\]

showing the convexity of \( \eta_{2^k} \). To complete the sequence, extend to \( \eta_0 = 0 \). Proposition 3 still provides convexity at \( k = 1 \), using the identity matrix when necessary, this easily implies that \( \eta_{2^k}/(2^k) \) must be non-decreasing for \( k \geq 1 \).

**VIII. Numerical Results**

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

\[
\dot{x}_2 := \frac{1}{n} \log \left( \frac{1}{\sqrt{p}} \mathbb{E} \left[ \text{diag}(w_n)^{-1} M_n (I-J) \right] \right),
\]

aggregating the empirical rate of convergence of the process for the full range of starting vectors 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 [19], [20].

For the underlying network, we consider a model based on Random Geometric Graphs (RGG) [21] with a simple perturbation where a dependence among the positions of the agents can be introduced. We interpolate between a grid and uniform random placement. 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: $z'_i$, a unique point on the $\sqrt{p_0} \times \sqrt{p_0}$ square grid fitted into $[0, 1]^2$ and $z'_i$, a uniform random position in $[0, 1]^2$. The final position is then declared as $z_i = c z'_i + (1 - c) z'_i$.

It remains to define the graph on the points obtained in 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 [22] 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 Fig. 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 $p \approx 58$.

The reference dynamics is asynchronous directed gossip: every step a uniformly chosen node communicates towards a single uniformly chosen neighbor, sending $1/2$ fraction of its value and weight. Fig. 2 presents $\hat{\lambda}_2$ according to (8) together with $\eta_2/2$ for 500 simulations for various values of $c$, we see the two moving together despite the wild randomness of graph instances. We also see the difference of the two, showing that $\eta_2/2$ is a 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 $1/2$ to be sent between them. Note that the conditions of Theorem 1, Assumption 1 still holds, here $\mathcal{A}$ is a union of segments in the space of non-negative matrices. We denote the empirical rate and the computed bound by $\hat{\lambda}_{tw}$, $\eta_{tw}/2$, respectively. Second, the reference gossip is modified to send only a fraction of $1/4$ to a single recipient, but we allow twice as many steps to take place, we will name this slowed gossip for convenience. Now the corresponding notation for empirical and bounding rates are $2\hat{\lambda}_{sg}$, $\eta_{sg}/2$, which include the rescaling due to the change in running time. Fig. 3 presents the comparison of rates of the modified strategies. For the empirical rates we observe no consistent ordering of the three strategies, however, the difference is an order of magnitude smaller than the variance caused by the graph variability, see the range in Fig. 2. The two-way gossip has a stronger computed bound than the reference process, even more for the slowed gossip.

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), at each step, we evaluate $\hat{\lambda}_2$, $\eta_2/2$, which include the rescaling due to the change in running time. Fig. 3 presents the comparison of rates of the modified strategies. For the empirical rates we observe no consistent ordering of the three strategies, however, the difference is an order of magnitude smaller than the variance caused by the graph variability, see the range in Fig. 2. The two-way gossip has a stronger computed bound than the reference process, even more for the slowed gossip.
then the reference. Note however, this ordering is not fully present in the earlier sparse, less interconnected phases.

IX. 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 [12] and by the analysis in [13]. The quantity \( \eta_{2k} = \log \rho(\mathbf{A} \otimes (I - J) \otimes (I - J)) \) 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.

APPENDIX

**Proof of Lemma 1:** Confirming its primitivity, there exist \( \tilde{A}_m \in A, m = 1, \ldots, 1 \) such that \( A \tilde{A}_1 \cdots \tilde{A}_1 \) is strictly positive. Let \( N \) denote the cone of non-negative \( p \times p \) matrices and define \( \gamma : [0, \infty) \to [0, 1) \) as the indicator of being positive, which naturally extends element-wise to \( N \).

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

For each \( \tilde{A}_m \), define the small neighborhood \( B_m = B(\tilde{A}_m, \epsilon_m/2) \cap N \) with \( \epsilon_m = \min_i,j(\tilde{A}_m^{ij} | \tilde{A}_m > 0) \). Crucially, for any \( B_m \in B_m, \gamma(B_m) \geq \gamma(A_m) \) element-wise. Also \( \mu(B_m) > 0 \) as \( \mu \) is fully supported on \( A \). Therefore the series of sets with positive probability \( B_1, \ldots, B_l \) will eventually be hit together in a block, noting the generating sequence \( (A_n), n \geq 1 \) is i.i.d. At that moment for the realizing matrices \( B_0 \in B_m \) we have \( \gamma(B_1 \cdots B_l) = \gamma(B_l) \cdots \gamma(B_1) \geq \gamma(A_l) \cdots \gamma(A_1) \) which is strictly positive. 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 2:** 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 (\( \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \)) 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 4:** The claim is a direct consequence of a [13, Lemma 44], the validity of its conditions has been verified in Section V. 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 \( M_n^i/M_n^j \) is sub-exponential.

As \( w_n^i = w_n^j, \) each \( w_n^i/w_n^j \) is easily seen to be a convex combination of the corresponding quotients \( M_n^i/M_n^j, k = 1, \ldots, p. \) 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 \( 1/w_n^i = p = n \) Summation of \( w_n^i/w_n^j \) through \( i, j \) fixed, yields \( \rho/w_n \). Since each term is sub-exponential, it follows that \( p/w_n \), and its maximum over \( j, \) is sub-exponential as well.

**References**

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

[2] F. Bénézit, V. Blondel, P. Thiran, J. N. Tsitsiklis, and M. Vetterli, “Weighted gossip: Distributed averaging using non-doubly stochastic matrices,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), 2010, pp. 1753–1757.

[3] A. Giridhar and P. R. Kumar, “Toward a theory of in-network computation in wireless sensor networks,” IEEE Commun. Mag., vol. 44, no. 4, pp. 98–107, Apr. 2006.

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

[5] A. Nedić and A. Olshevsky, “Distributed optimization over time-varying directed graphs,” IEEE Trans. Autom. Control, vol. 60, no. 3, pp. 601–615, Mar. 2015.

[6] J. N. Tsitsiklis, “Problems in decentralized decision making and computation,” Ph.D. dissertation, Dept. Electr. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 1984.

[7] V. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, “Convergence in multiaagent coordination, consensus, and flocking,” in Proc. 44th IEEE Conf. Decis. Control, 2005, pp. 2996–3000.

[8] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” IEEE Trans. Inf. Theory, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[9] 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.

[10] A. H. Sayed, “Adaptation, learning, and optimization over networks,” Found. Trends Mach. Learn., vol. 7, nos. 4–5, pp. 311–801, 2014.

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

[12] 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.

[13] B. Gerencsér and L. Gerencsér, “Tight bounds on the convergence rate of generalized ratio consensus algorithms,” IEEE Trans. Autom. Control, vol. 67, no. 4, pp. 1669–1684, Apr. 2022.

[14] 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,” Math. Control Signals Syst., vol. 10, no. 1, pp. 31–40, 1997.

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

[16] E. Seneta, Non-Negative Matrices and Markov Chains. New York, NY, USA: Springer, 2006.

[17] L. Arnold, W. Kliemann, and E. Oeljeklaus, “Lyapunov exponents of linear stochastic systems,” in Lyapunov Exponents. Heidelberg, Germany: Springer, 1986, pp. 85–125.

[18] L. Gerencsér and Z. Orlovich, “Lq-stability of products of block-triangular stationary random matrices,” Acta Scienciarum Mathematicarum, vol. 74, nos. 3–4, pp. 927–944, 2008.

[19] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, “Julia: A fresh approach to numerical computing,” SIAM Rev., vol. 59, no. 1, pp. 65–98, 2017. [Online]. Available: https://epubs.siam.org/doi/10.1137/141000671

[20] J. Fairbanks, M. Besançon, S. Schöll, J. Hoffiman, N. Eubank, and S. Karpinski, “JuliaGraphs/Graphs.jl: An Optimized Graphs Package for the Julia Programming Language.” 2021. [Online]. Available: https://github.com/JuliaGraphs/Graphs.jl

[21] M. Penrose, Random Geometric Graphs, vol. 5. Oxford, U.K.: Oxford Univ. Press, 2003.

[22] P. Gupta and P. R. Kumar, Critical Power for Asymptotic Connectivity in Wireless Networks. Boston, MA, USA: Birkhäuser, 1999, pp. 547–566. [Online]. Available: https://doi.org/10.1007/978-1-4612-1784-8_33