Notes on Linear Gossip Algorithms: Potentials and Challenges*

Dedicated to Peter E. Caines in honour of his 80th birthday

GERENCSÉR Balázs · GERENCSÉR László

DOI: 10.1007/s11424-025-4534-4
Received: 30 October 2024

Abstract The authors consider the problem of reaching consensus over a communication network via asynchronous interaction between pairs of agents. A well-known method is the linear gossip algorithm due to Tsitsiklis (1984). Extension of this, allowing the selection of a strictly stationary sequence of communicating pairs, was given in Picci and Taylor (2013). Extension of the linear gossip algorithm to directed communication networks, retaining the linear dynamics, was proposed by Cai and Ishii (2012), later extended by Silvestre, et al. (2018). A definite novelty of these algorithms is that L_2-convergence with exponential rate can be established. The authors attend the above issues, extending the result of Picci and Taylor (2013) motivated by features of algorithms for directed networks. The authors present and discuss the algorithm of Silvestre, et al. (2018), together with systematic simulation results based on 5M randomly chosen parameter settings. The core of the proposed mathematical technology is a set of simple observations, presented with a tutorial aspect, by which the authors can conveniently establish various results on the almost sure convergence of products of strictly stationary sequences of matrices to a rank-1 matrix.

Keywords Consensus, gossip algorithms, Lyapunov-exponents, spectral gap.

1 Introduction

"A good traveler has no fixed plans and is not intent on arriving."— Lao Tzu.

GERENCSÉR Balázs
HUN-REN Alfréd Rényi Institute of Mathematics, Budapest 1052*, Hungary and Eötvös Loránd University, Department of Probability and Statistics, Budapest* 1518*, Hungary.* Email: gerencser.balazs@renyi.hun-ren.hu.
GERENCSÉR László (Corresponding author)
HUN-REN SZTAKI, Institute for Computer Science and Control, Budapest 1111*, Hungary.*
Email: gerencser.laszlo@sztaki.hun-ren.hu.
*This research was supported by the Ministry of Innovation and Technology, NRDI (National Research, Development and Innovation Office) under Grant No. KKP 137490 and its Autonomous Systems National Laboratory Program (ARNL).
◇ *This paper was recommended for publication by Guest Editor ZHANG Ji-Feng.*

Salute. We have met Peter in person in January 1989 arriving to Montreal, where I (the second author) was offered a position of a visiting professor at the Department of Electrical Engineering, McGill University. My co-author, my son, was 4 years old at that time. We have left a country behind the political structures of which were about to fall apart, and found ourselves in a cultural and professional atmosphere hardly known back home. The enthusiasm and generosity of Peter was certainly a key factor in shaping the next three and half years into a most memorable and productive period of my life. In spite of being a newcomer I had the privilege to teach a course on his then freshly published book, Linear Stochastic Systems, republished as [1]. This course has become since then my favorite point of return in my teaching activity. John Littlewood wrote: "Mathematics is the art of vague thinking". One of the excellent qualities of Peter is his ability to articulate his/our position in face of vagueness. I was also impressed by his innovative spirit encouraging his students to enter into new fields of research, such as discrete event systems or mean-field theory. I should also remember here Jorma Rissanen for having connected me with Peter. In addition, their joint paper [2] is partially responsible for my never ceasing interest system identification. Peter and his wife, Anne were also a solid background for our whole family both in needs and on festive occasions. Let me thank to you for all these, and wish you peaceful years to come.

The subject matter of this paper is a mathematical technology relevant for the analysis of consensus algorithms defined over a network. The objective of these algorithms is to compute the average of values associated with each node of the network, denoted by x^i for node i, via local, asynchronous communication between the agents. Such problems arise in a number of applications such as sensor networks[3], platooning[4], smart grids[5] or opinion dynamics[6].

For historical perspective we note that a pioneering work on the subject is due to Tsitsiklis[7], with the introduction of the classical gossip algorithm. In this simplest setting the communication links, modeled as the edges of a graph, are not directed, and if i and j communicate, then they update their values simultaneously and symmetrically by taking the same convex combination of their respective current values. The dynamics of the vector of values is thus defined in terms of a multiplication by a random, doubly stochastic matrix. For this reason this algorithm is also referred to as a *linear gossip* algorithm.

In its classical form, it is also assumed that the pairs (i, j) are chosen independently with fixed probabilities following the tick of a global clock. Convergence of variants of the gossip algorithms have been studied in [8].

Considerable *extension* of the classical linear gossip algorithms has been presented by Picci and Taylor[9], allowing a strictly stationary, ergodic choice of communicating pairs, in particular allowing a sequential choice of communicating nodes, rendering the algorithm truly distributed. In their key result they prove almost sure convergence with geometric rate.

If the communication links are *directed*, and communicating agents i and j update their values differently then the gossip algorithm may fail to converge to the mean. However, an alternative approach, variants of which has been introduced under the names push-sum[10], weighted gossip[11] or ratio consensus algorithm[12], does succeed in reaching consensus.

A convenient illustration of this approach is that various amounts of a liquid containing a chemical in various concentrations are given at the nodes. The objective is to determine the concentration of the grand total using local mixtures between communicating pairs.

In [10] it is assumed that the communication pairs (i, j), with i sending a message to j, are chosen *independently* with fixed probabilities following the tick of a global clock. Almost sure (a.s.) convergence of variants of the algorithm has been established.

Significant *extension* of the classical setup has been presented in [11], allowing strictly stationary, ergodic communication protocols. They proved a.s. convergence raising the question of identifying its rate. This open problem has been settled in [12] for a variety of settings. It may come as a pleasing surprise that the established a.s. rate is identical with that proved in [9] for linear gossip algorithms.

In passing we note that the methodology of [12] also allows the extended analysis of the effect of packet losses for push-sum algorithms, initially studied in [13].

The results of the present paper complement and extend the results in [9], and closely related, mathematically sophisticated previous results of [14] and [15].

The rate of convergence in [12] is given in terms of *spectral gap* $\lambda_1 - \lambda_2$, where λ_1, λ_2 denote the first two Lyapunov exponents associated to products of random matrices. However, it is known that in general computing the spectral gap can be NP-hard[16, 17]. On the other hand, extending the arguments of [18], computable upper bounds were given in [19, 20].

A valid reservation against ratio consensus algorithms is that the mean is being estimated by ratios of two random variables, as a consequence up to now L_2-convergence has not been established.

An ingenious method to circumvent this shortcoming was proposed by [21], defining a linear gossip algorithm for networks with directed communication links, by introducing an auxiliary variable for each node, interpreted as a surplus or *account*, the overall effect of which is that the sum of the variables remains constant. Their values may be negative, hence we prefer to refer to these variables as accounts for bookkeeping. A key feature of their construction is that the matrices describing the joint dynamics of extended variables have a common fixed left and right eigenvector, mimicking the situation of doubly stochastic matrices.

The motivation of the present paper is to discuss some of the mathematical delicacies relevant to the analysis of the linear gossip algorithm proposed in [22], the convergence of which has been established for the case of independent identically distributed (i.i.d.) sequence of matrices, under a specific choice of parameter values $\alpha = \beta = \gamma = 1/2$, defining the algorithm. Simulation results suggest, though, that the algorithm may not converge for arbitrary choices of the parameters, as suggested by the authors, see Section 7. This deficiency justifies the analysis of some of the mathematical building blocks that are relevant in the construction of the linear gossip algorithm of [22].

An essential novel aspect of the present paper is that we consider communication protocols defined by a *strictly stationary ergodic sequence* of matrices, defined below, that can be conveniently handled via Oseledec's multiplicative ergodic theorem[23]. In particular, our technical results apply for sequential communication protocols, reflecting the case of a truly distributed consensus algorithm. It is hoped that these results will contribute to the full-scale analysis of an appropriately defined extension of the linear gossip algorithm of [22].

The structure of the paper is as follows. In Section 2 we present the consensus problem for undirected and directed networks, and state two basic convergence results for classic linear gossip and for ratio consensus methods. Then we briefly present the algorithm proposed by [22] with a novel interpretation. We also present a preliminary application of Theorem 5.1 for a sequence of random matrices sharing some features with those in the algorithm of [22], see Theorem 2.4.

In Section 3, we supply the mathematical background material necessary to formulate and prove our results. In Section 4 we collect a set of simple results on random sequences of matrices with a common eigenvector, with a tutorial aspect in mind.

The *main technical result* of our paper, establishing a.s. rank-1 limit of the product of a strictly stationary, ergodic sequence of matrices, sharing a common left and right eigenvector with common eigenvalue, is given in Section 5, formulated as Theorem 5.1. This result is a significant extension of [9, Theorem 5.2], although under a slightly stronger condition, with a proof, relying on the preliminary results of Section 4, that is fairly transparent.

A simple corollary of this result, with a setup in close kinship with that of the linear gossip algorithm of [22], is given as Theorem 2.4. We also present a method for constructing a random sequence of matrices satisfying the conditions of Theorem 5.1, a special case of which, Theorem 5.5, is designed to match some of the features of the linear gossip algorithm of [22].

In Section 6 we revisit the linear gossip algorithm of [22], provide its detailed interpretation and formal description, together with a critical assessment, in the light of [21].

Finally, in Section 7 we present the summary of systematic simulation results based on 5M randomly chosen parameter values α, β, γ, defining the algorithm of [22], resulting in a remarkable dichotomy with respect to (w.r.t.) empirical stability. We also provide simulation results for a sequential communication protocol, outperforming the above methods.

A retrospect view of the paper is provided in a brief discussion.

2 Technical Setup of Consensus Problems

We consider a communication network represented by a directed graph $G = (V, E)$, with $|V| = p$. To each node i a real number x^i is associated, called the values. The problem is then to compute the ratio $\overline{x} := \sum_i x^i/p$ at all nodes recursively, using only local asynchronous interactions allowed by $G = (V, E)$. This is called the *average consensus* problem.

2.1 The Linear Gossip Algorithm

In the classic linear gossip algorithm, proposed in [7] the graph $G = (V, E)$ is not directed, and at any time n the communicating nodes i, j, with $(i, j) \in E$, update their old values, denoted by x^i_{n-1} and x^j_{n-1}, simultaneously. Letting i be the sender and j the receiver, node i sends a fixed deterministic fraction of its value to node j, denoted by α_{ji}, with $0 < \alpha_{ji} < 1$. The updating rule is symmetric, $\alpha_{ji} = \alpha_{ij}$. Then the current values at nodes i, j are updated to the convex combinations

$$x_n^j = \alpha_{ji} x_{n-1}^i + (1-\alpha_{ji}) x_{n-1}^j, \tag{1}$$
$$x_n^i = \alpha_{ij} x_{n-1}^j + (1-\alpha_{ij}) x_{n-1}^i. \tag{2}$$

As mentioned in the Introduction, in the classical form of the linear gossip algorithm the pairs (i, j) are chosen independently with fixed probabilities following the tick of a global clock. The communication link active at time n will be denoted by (i_n, j_n).

In order to describe the overall dynamics in matrix-vector notation let $x_n = (x_n^1, \cdots, x_n^p)^{\mathrm{T}}$ denote the vector of values at time n with $x_0 = x$. Construct a $p \times p$ matrix A_n obtained from the identity matrix I_p as follows: Replace the 2×2 minor I_2 of I_p corresponding to the row and column indices $i = i_n$ and $j = j_n$ by the 2×2 matrix $\overline{A}$ defined as

$$\overline{A} := \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}, \tag{3}$$

with $0 < \alpha < 1$, acting on (x_n^i, x_n^j). Obviously, A_n is a doubly stochastic matrix, having right and left eigenvectors $u = v = \mathbf{1} = (1, 1, \cdots, 1)^{\mathrm{T}}$:

$$A_n \mathbf{1} = \mathbf{1}, \quad \mathbf{1}^{\mathrm{T}} A_n = \mathbf{1}^{\mathrm{T}}. \tag{4}$$

Then the dynamics of the above linear gossip algorithms is described by the equation for $n \geq 1$,

$$x_n = A_n x_{n-1}, \quad x_0 = x. \tag{5}$$

At this point we have to interrupt our discussion with a *technical reminder*. Let $\xi = (\xi_n)$, $n \geq 0$ be a one-sided stochastic process defined over the probability space $(\Omega, \mathcal{F}, P)$, taking its values in a Euclidean space. The process ξ is *strictly stationary* if for any fixed set of integer indices $\{0 \leq i_1 < i_2 < \cdots < i_m\}$ the joint distributions of the m-tuples $(\xi_{i_1+\tau}, \xi_{i_2+\tau}, \cdots, \xi_{i_m+\tau})$ are identical for all $0 \leq \tau \in \mathbb{Z}$. A classic way of constructing ξ is as follows:

$$\xi_n(\omega) = \xi_0(T^n \omega),$$

for all $n \geq 0$, $\omega \in \Omega$, where $T : \Omega \to \Omega$ is a measurable, measure-preserving mapping, meaning that for all $C \in \mathcal{F}$ we have $P(T^{-1}C) = P(C)$. Any one-sided strictly stationary process ξ can be represented in this way, in a weak sense, meaning distributional equivalence, over an appropriate probability space $(\Omega', \mathcal{F}', P')$. We may choose $\Omega' = \{\omega' : \omega' = (\xi_0(\omega), \xi_1(\omega), \cdots), \omega \in \Omega\}$, the set of semi-infinite trajectories of ξ, and define T' as the left shift, deleting $\xi_0(\omega)$.

For a given T, a set $C \in \mathcal{F}$ is essentially invariant if $P(T^{-1}C \,\Delta\, C) = 0$, with Δ denoting the symmetric difference of two sets. The process ξ and T itself are *ergodic* if for any essentially invariant set $C \in \mathcal{F}$ we have either $P(C) = 0$ or $P(C) = 1$. Equivalently, a strictly stationary process ξ is ergodic if and only if for all non-negative measurable functions $f : \Omega \to \mathbb{R}$ we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\xi_n) = \mathbb{E}\left[f(\xi_0)\right] \quad \text{a.s.}$$

Similarly, a *two-sided* stochastic processes $\xi = (\xi_n)$, $-\infty < n < +\infty$, defined on $(\Omega, \mathcal{F}, P)$, is strictly stationary if for any fixed set of integer indices $\{i_1 < i_2 < \cdots < i_m\}$ the joint distributions of the m-tuples $(\xi_{i_1+\tau}, \xi_{i_2+\tau}, \cdots, \xi_{i_m+\tau})$ are identical for all $\tau \in \mathbb{Z}$. Setting

$$\xi_n(\omega) = \xi_0(T^n \omega),$$

for all $n \in \mathbb{Z}$, $\omega \in \Omega$, where $T : \Omega \to \Omega$ is a *one-to-one* measure-preserving mapping is a convenient way of defining ξ. Any two-sided strictly stationary process ξ can be represented in this way in a weak sense over an appropriate probability space $(\Omega', \mathcal{F}', P')$. A suitable choice may be $\Omega' = \{\omega' : \omega' = (\cdots, \xi_{-1}(\omega), \xi_0(\omega), \xi_1(\omega), \cdots), \omega \in \Omega\}$, the set of two-way-infinite trajectories of ξ, with T' defined as the left shift. ξ will be called *ergodic* if T, and hence T^{-1}, are ergodic. In what follows, the distinction between one-sided vs. two-sided ergodic processes will be made explicit or implied by the context.

Returning to [9], as pointed out in the Introduction, in this paper the classic assumption that the selection of communicating pairs is i.i.d. has been significantly relaxed. To present their result consider an underlying probability space $(\Omega, \mathcal{F}, P)$. We say that the edge-process $(i_n, j_n), n \geq 0$ or the communication protocol is strictly stationary ergodic if the associated matrix-valued process is strictly stationary ergodic. Next, we formulate a condition in the terminology of [12]. A matrix is called *allowable*, if it has no zero row or zero column. A strictly stationary process of non-negative allowable random matrices $(A_n), n \geq 1$, is called (forward) *sequentially primitive* if $M_\tau = A_\tau \cdot A_{\tau-1} \cdots A_1$ is strictly positive for some finite stopping time τ with probability 1 (w.p.1). For details see (40) below.

For the formulation of the main result of [9] we will also need the notion of Lyapunov exponents $\lambda_i = \lambda_i(A)$, $i = 1, 2, \cdots, p$, associated with a strictly stationary, ergodic sequence of matrices, such as (A_n), $n \geq 1$, to be discussed in Section 3. An upgraded form of the main result of [9], Theorem 5.2, under the stronger condition of assuming two-sided ergodicity, can now be stated as follows, with arbitrary matrix norm $\|\cdot\|$:

Theorem 2.1 *Assume that* (A_n), $-\infty < n < +\infty$, *defined above is a strictly stationary, ergodic process of random,* $p \times p$ *matrices. Assume that* (A_n) *is sequentially primitive. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| A_n \cdot A_{n-1} \cdots A_1 - \frac{\mathbf{1}\mathbf{1}^{\mathrm{T}}}{\mathbf{1}^{\mathrm{T}}\mathbf{1}} \right\| = \lambda_2(A) < 0 \quad \text{a.s.} \tag{6}$$

The above result is a direct corollary of Theorem 5.1, Section 5. As $\mathbf{1}^{\mathrm{T}}\mathbf{1} = p$ it follows that average consensus takes place with exponential rate, bounded from above by $\lambda_2(A) < 0$, improving the claim of [9, Theorem 5.2] on exponential rate with *some* rate $\lambda < 0$.

It should be noted that the condition in [9], assuming that for all edge $e \in E$ we have $P(e) > 0$, together with the assumption that the graph (G, E) is connected, implies that (A_n) is sequentially primitive, see [9, Lemma 5.1].

Note that the construction of A_n can be more formally described as follows. Let $\boldsymbol{e}_i$ be the i-th unit vector in $\mathbb{R}^p$, and let $i = i_n$, $j = j_n$. Define the $p \times 2$ selector matrices

$$S_n = S(i, j) = (\boldsymbol{e}_i, \boldsymbol{e}_j). \tag{7}$$

Then we can write

$$A_n = I_p - S_n(I_2 - \overline{A})S_n^{\mathrm{T}}. \tag{8}$$

The convenience of this representation will become plausible below, particularly in Section 6.

2.2 Ratio Consensus

We now shift our focus to the case when the communication network is represented by a directed graph $G = (V, E)$. In order to deal with this situation we introduce a set of auxiliary variables $w^i \geq 0$ called the weights.

Let $x_0 = x = (x^1, x^2, \cdots, x^p)^{\mathrm{T}}$ and $w_0 = w = (w^1, w^2, \cdots, w^p)^{\mathrm{T}}$ denote the vectors of initial values and weights, respectively, at time 0, assuming $w \geq 0, w \neq 0$. A standard choice is $w_0 = (1, 1, \cdots, 1)^{\mathrm{T}}$, which is the one needed to achieve average consensus. We update both the values and weights at time n as follows. Select a directed edge $(i, j) \in E$ randomly, representing the communicating pair. Then the sender, node i, initiates a transactions by sending a fixed deterministic fraction, say α_{ji} with $0 < \alpha_{ji} < 1$, of his/her values and weights to the receiver, node j. A classic assumption is to require that the sequence of communicating edges is i.i.d. and $\alpha_{ji} = 1/2$ for all communicating edges, leading to the celebrated push-sum method[10]. However, following [11] and [12], our focus of attention is in the case of strictly stationary ergodic sequences.

In order to describe the dynamics in matrix-vector notation, let x_{n-1} and w_{n-1} denote the p-vector of values and weights, respectively, at time $n-1$. The dynamics can then be formally described by the equations

$$x_n = A_n x_{n-1} \quad \text{and} \quad w_n = A_n w_{n-1}, \tag{9}$$

for $n \geq 1$, where A_n is a $p \times p$ random matrix obtained from the identity matrix I_p as follows. Consider the matrices describing the interaction between $i = i_n$ and $j = j_n$ with $i < j$:

$$\overline{A} := \begin{pmatrix} 1 - \alpha_{ji} & 0 \\ \alpha_{ji} & 1 \end{pmatrix}. \tag{10}$$

Then A_n is obtained by replacing the minor of I_p defined by the indices i, j by $\overline{A}$. Note that A_n is column-stochastic for all n. Using the selector matrices $S_n = S(i, j)$ defined above with $i = i_n$, $j = j_n$, we may express A_n formally as

$$A_n = I_p - S_n(I_2 - \overline{A})S_n^{\mathrm{T}}. \tag{11}$$

A remarkable property of the ratio consensus algorithm has been established in [12, Theorem 19], recast and specialized below for the current setup as follows:

Theorem 2.2 *Let (A_n), $n \geq 1$, defined as in (11), be a strictly stationary, ergodic process of random, $p \times p$ matrices. Assume that (A_n) is (forward) sequentially primitive and for the forward index of sequential primitivity ψ_n, see below (40), we have $\mathbb{E}\psi_n < \infty$. Then for any vector of initial values $x \in \mathbb{R}^p$, and any non-negative vector of initial weights $w \in \mathbb{R}^p_+$ such that $w \neq 0$ we have for all i,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \left| \frac{x_n^i}{w_n^i} - \frac{\mathbf{1}^{\mathrm{T}} x}{\mathbf{1}^{\mathrm{T}} w} \right| \leq \lambda_2 < 0 \quad \text{a.s.}$$

Choosing $w = \mathbf{1}$, Theorem 2.2 implies that average consensus is achieved: For all agents i the values x_n^i / w_n^i will converge to the same non-random limit $\overline{x} = \sum_{i=1}^p x_0^i / p$, with at least the given rate.

A by-product of the proof of the theorem above given in [12] we obtain that the matrix product $A_n \cdots A_1$ can be approximated by a rank-1 matrix with exponentially decaying error: There exist a sequence of random vectors $(u_n), n \geq 1$ such that almost surely

$$\limsup_{n\to\infty} \frac{1}{n} \log \left| A_n \cdot A_{n-1} \cdots A_1 - \frac{u_n \mathbf{1}^{\mathrm{T}}}{\mathbf{1}^{\mathrm{T}} u_n} \right| \leq \lambda_2 < 0. \tag{12}$$

For a justification of the above see Corollary 3.3 in Section 3.

It may come as a pleasing surprise that the a.s. rate of convergence for ratio consensus algorithms provided by Theorem 2.2 is identical with the a.s. rate of convergence of linear gossip algorithms in Theorem 2.1.

Remark 2.3 The condition $\mathbb{E}\psi_n < \infty$ does not show up in Theorem 2.1. This is not accidental: This condition is needed only for the ratio consensus algorithm, see [12, Lemma 44], easily implying that $1/w_n^i$ is sub-exponential, see [19, Lemma 10], thus not deteriorating the convergence rate.

It should be noted, that in analogy with the assumption for the linear gossip algorithm given in [9], assuming that for all edge $e \in E$ we have $P(e) > 0$, together with the assumption that the graph (G, E) is strongly connected, implies that (A_n) is sequentially primitive.

On the sideline: A stochastic process $\xi_n, n \geq 1$ is called sub-exponential, if for any $\varepsilon > 0$ we have for all n, with finitely many exceptions, a.s. $|\xi_n| \leq \exp(\varepsilon n)$. We will use the notation $\xi_n = \exp(o(1)n)$. Equivalently, $\xi_n, n \geq 1$ is called sub-exponential if $\limsup_{n\to\infty} \frac{1}{n} \log |\xi_n| \leq 0$.

In the case when $(A_n), n \geq 1$ form an i.i.d. sequence, a computable upper bound for the rate λ_2 has been given in [19]: Setting $B_1 := A_1 \left(I - \mathbf{1}\mathbf{1}^{\mathrm{T}} / \mathbf{1}^{\mathrm{T}}\mathbf{1} \right)$ we have

$$\lambda_2 \leq \frac{1}{2} \log \rho \left(\mathbb{E} \left[B_1 \otimes B_1 \right] \right),$$

with $\rho(\cdot)$ denoting the spectral radius. Extensive simulations indicate that the above upper bound is surprisingly tight.

2.3 Linear Gossip Algorithm for Directed Networks

As briefly mentioned in the Introduction, [22] presented a completely different approach to solving the average consensus problem for directed networks $G = (V, E)$. The communication protocol is assumed to be i.i.d. using edges of a strongly connected graph with positive probabilities. The idea is that along with the values x^i we introduce auxiliary variables y^i for each node serving as an account for bookkeeping, initialized as $y_0^i = 0$ for all i.

To describe the overall dynamics on the network in a compact, preliminary form let x_{n-1} and y_{n-1} denote the p-vectors of values and accounts, respectively, at time $n-1$, and let

$$z_{n-1} := \begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix}, \tag{13}$$

with $y_{n-1} = \mathbf{0}$. The authors of [22] then constructed an i.i.d. sequence of $2p \times 2p$ matrices (F_n), defined in terms of three parameters α, β, γ, characterizing the strength of communication between an active pair of sender and receiver, and their accounts. Specific details will be given in Section 6. The construction is deliberately such that the matrices F_n have a common fixed right and left eigenvector u and v with eigenvalue 1, in analogy with the linear gossip algorithm, see (4), namely for all n

$$F_n u = u \quad \text{and} \quad v^{\mathrm{T}} F_n = v^{\mathrm{T}}. \tag{14}$$

However, in the setting of [22] we have

$$u = \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} \quad \text{and} \quad v = \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \end{pmatrix}, \tag{15}$$

with $\mathbf{1} = \mathbf{1}_p$ and $\mathbf{0} = \mathbf{0}_p$. Note that the condition $F_n u = u$ virtually implies that F_n has at least one negative element, excluding the possibility of applying the arsenal of the theory of non-negative matrices, such as [24].

The dynamics of the system can be compactly written as

$$z_n = F_n z_{n-1}, \tag{16}$$

with z_0 containing initial data at the nodes of the network, with $y_0 = 0$. It is claimed in [22, Theorem 2] that with the specific choice of $\alpha = \beta = \gamma = 1/2$ exponentially fast a.s. convergence of z_n to average consensus takes place:

$$\limsup_{n \to \infty} \frac{1}{n} \log |F_n \cdot F_{n-1} \cdots F_1\, z_0 - \overline{x}\, u| < 0. \tag{17}$$

The above result is actually stated as a simple corollary of exponentially fast L_2-convergence, the latter being the main point of attraction of [22, Theorem 2]. Notably, L_2-convergence for ratio consensus is still an open problem.

The proof is intricate and problem-specific, raising issues of validation and simplification. The reader may be puzzled by the claim, briefly mentioned in [22], that the above convergence results may hold true for any choice of $0 < \alpha, \beta, \gamma < 1$. Systematic simulations indicate that this claim or conjecture might not be true in general, although it may be true for a fairly large set of parameters, see Section 7.

It may be of interest to provide a transparent analysis of the linear gossip algorithm proposed in [22] even for the i.i.d. case. A more ambitious project may be to extend such an analysis to strictly stationary ergodic processes (F_n).

A partial progress in this direction, resembling Theorem 2.1, can be formulated as follows:

Theorem 2.4 *Assume that $F = (F_n)$, $-\infty < n < +\infty$ is a strictly stationary, ergodic process of random, $2p \times 2p$ matrices, satisfying* (14) *with u, v given in* (15)*. Let $\mathbb{E} \log^+ \|F_n\| <$*

∞. *Assume that the top Lyapunov exponent of* $F = (F_n)$ *is* 0*, or* $\lambda_1(F) = 0$*, and the spectral gap is positive, i.e.,* $\lambda_1(F) - \lambda_2(F) > 0$*. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| F_n \cdot F_{n-1} \cdots F_1 - \frac{uv^{\mathrm{T}}}{u^{\mathrm{T}}v} \right\| = \lambda_2(F) \quad a.s. \tag{18}$$

The above result is a direct corollary of Theorem 5.1 of Section 5. A generic method for constructing random sequences of matrices (F_n), satisfying the conditions of the theorem, will be also given in Section 5.

This theorem clarifies the rational behind the specifications of (F_n) in the light of the average consensus problem. Namely, choosing an initial value z_0 with 0-s for all accounts, i.e.,

$$z_0 := \begin{pmatrix} x_0 \\ \mathbf{0} \end{pmatrix}, \tag{19}$$

the limiting effect of the transactions, provided by the theorem above, is

$$\frac{uv^{\mathrm{T}}}{u^{\mathrm{T}}v} z_0 = u\,(v^{\mathrm{T}} z_0)/p = u\,(\mathbf{1}_p{}^{\mathrm{T}} x_0)/p = \overline{x}u. \tag{20}$$

Thus we arrive at the following conclusion:

Corollary 2.5 *Under the conditions of Theorem* 2.4 *we have almost surely,*

$$\limsup_{n\to\infty} \frac{1}{n} \log |F_n \cdot F_{n-1} \cdots F_1\, z_0 - \overline{x}\, u| \le \lambda_2 < 0. \tag{21}$$

Thus, average consensus, restricted to the values x^i, takes place with exponential rate. In addition, the accounts y^i are ultimately reset to 0 with exponential rate.

A common generalization of Theorems 2.1 and 2.4 will be stated and proved in Section 5, see Theorem 5.1.

Note, however, that the validity of (21) does not imply the validity of (18), since the range of initial values z_0 for (21) is restricted to a p-dimensional subspace.

While the above corollary is fairly general, it remains an open problem if its conditions can be verified even for the i.i.d. sequence (F_n) specified in [22, Theorem 2]. An alternative challenge is to construct a sequence of matrices (F_n) from scratch, satisfying the conditions of Theorem 2.4, reflecting a truly distributed computing algorithm.

Following [22], we may wish to impose the constraint that F_n has to reflect one-directional transactions between the sender i and the receiver j and their accounts at any time n. Thus F_n has to be obtained from the unit matrix I_{2p} by replacing one of its 4×4 minors, corresponding to the values and accounts at i and j, by a fixed 4×4 matrix $\overline{F}$, defining a fixed pattern of transactions.

A key challenge is then to construct a matrix $\overline{F}$ and a communication protocol such that $0 = \lambda_1(F) > \lambda_2(F)$. The arguments of [21] and [22] suggest that selecting i.i.d. communicating edges may yield a sequence (F_n) with the above property, but a rigorous statement requires additional meticulous considerations.

3 Technical Preliminaries

3.1 Product of Random Matrices

In this subsection we provide a summary of two fundamental results on products of random matrices, based on [12]: The Fürstenberg-Kesten theorem[26], and Oseledec's multiplicative ergodic theorem[23], an improved version of which has been given in [25]. Let $|\cdot|$ denote an arbitrary norm in $\mathbb{R}^p$, and let $\|\cdot\|$ denote the induced operator norm for matrices.

Proposition 3.1 (Fürstenberg-Kesten's theorem) *Let $A := (A_n)$, $n \geq 1$ be a strictly stationary, ergodic random sequence of $p \times p$ matrices over a complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$, such that $\mathbb{E} \log^+ \|A_1\| < \infty$. Then the a.s. limit*

$$\lambda_1(A) = \lim_{n\to\infty} \frac{1}{n} \log \|A_n \cdot A_{n-1} \cdots A_1\| < \infty, \tag{22}$$

allowing $\lambda_1(A) = -\infty$, exists, and it is equal to

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E} \log \|A_n \cdot A_{n-1} \cdots A_1\| = \inf_{n\geq 1} \frac{1}{n} \mathbb{E} \log \|A_n \cdot A_{n-1} \cdots A_1\|. \tag{23}$$

A more refined characterization of $A_n \cdot A_{n-1} \cdots A_1$ is given by Oseledec's theorem. For a start we make a brief detour in the elementary theory of Lyapunov exponents. Let $A = (A_n), n \geq 1$ be a fixed sequence of $p \times p$ matrices. For any $x \in \mathbb{R}^p$ define the Lyapunov exponent of x w.r.t. A as

$$\lambda(x) = \lambda(x, A) := \limsup_{n\to\infty} \frac{1}{n} \log |A_n \cdot A_{n-1} \cdots A_1 x|. \tag{24}$$

Then, for any extended real number $-\infty \leq \mu \leq +\infty$ define

$$L_\mu = L_\mu(A) := \{x \in \mathbb{R}^p : \lambda(x, A) \leq \mu\}. \tag{25}$$

It is easily seen that L_μ is a linear subspace of $\mathbb{R}^p$, for $\mu < \mu'$ we have $L_\mu \subseteq L_{\mu'}$, and L_μ is continuous from the right. It follows that there is a finite number of possible Lyapunov exponents, $+\infty \geq \mu_1 > \mu_2 > \cdots > \mu_q \geq -\infty$, such that

$$\mathbb{R}^p = L_{\mu_1} \supsetneq L_{\mu_2} \cdots \supsetneq L_{\mu_q} \supsetneq \{0\} =: L_{\mu_{q+1}}, \tag{26}$$

and L_μ is a piecewise constant function of μ with points of discontinuity exactly at μ_i. It follows that for $1 \leq r \leq q$,

$$x \in L_{\mu_r} \setminus L_{\mu_{r+1}} \quad \text{implies} \quad \lambda(x) = \mu_r. \tag{27}$$

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 co-dimension of L_{μ_r} relative to $L_{\mu_{r+1}}$ is $i_r - i_{r+1}$, which is interpreted as the multiplicity of the Lyapunov exponent μ_r. Accordingly, we define the full set of Lyapunov exponents $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ by setting for $1 \leq i \leq p$

$$\lambda_i = \mu_r, \quad \text{if} \quad i_r \geq i > i_{r+1}. \tag{28}$$

The *multi-set* of Lyapunov-exponents will be denoted by $\Lambda(A)$. It is readily seen that scaling $A = (A_n)$ by $\eta \in \mathbb{R}$, i.e., defining $\overline{A}_n = \eta A_n$ will yield

$$\Lambda(\overline{A}) = \Lambda(A) + \log|\eta|. \tag{29}$$

If $(A_n) = (A_n(\omega))$ is the realization of a strictly stationary ergodic process, then the above observations can be extended to the fascinating result below, stated first in [23], and proved under weaker condition in [25]. We will use the notation $M_n := A_n \cdot A_{n-1} \cdots A_1$.

Proposition 3.2 (Oseledec's theorem) *Assume that (A_n), $n \geq 1$ 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 limit below exists:*

$$\lambda(x) = \lim_{n\to\infty} \frac{1}{n} \log |A_n \cdot A_{n-1} \cdots A_1 x|. \tag{30}$$

Moreover, the Lyapunov exponents $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, 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. In addition, the following almost sure limit exists:

$$M^* = \lim_{n\to\infty} \left(M_n^{\mathrm{T}} M_n\right)^{1/2n}. \tag{31}$$

An alternative notation for the Lyapunov-exponents $\lambda(x)$, spelling out its dependency on $A = (A_n)$ and ω will be

$$\lambda(x) = \lambda(x, A, \omega). \tag{32}$$

The proof given in [25] implies the following remarkable fact on the singular value decomposition (SVD) of M_n:

$$M_n = U_n \Sigma_n V_n, \tag{33}$$

where U_n, V_n are orthonormal matrices, and Σ_n is diagonal with entries $\sigma_n^1 \geq \sigma_n^2 \geq \cdots \geq \sigma_n^p \geq 0$, such that

$$\lambda_k = \lim_{n\to\infty} \frac{1}{n} \log \sigma_n^k \quad \text{a.s.} \quad k = 1, \cdots, p. \tag{34}$$

Therefore we have

$$\Sigma_n = \operatorname{diag}(\mathrm{e}^{(\lambda_k + o(1))n}) \quad \text{a.s.} \tag{35}$$

Surprisingly, the orthonormal matrices V_n will also converge almost surely in a restricted sense. For the current paper we present only a special case. In particular, if $\lambda_1 > \lambda_2$, then, by Lemma 5 of [25], for the first row of V_n, denoted by $v_n^{1\cdot}$ we have for any $\varepsilon > 0$

$$v_n^{1\cdot} - v^{1\cdot} = O(\mathrm{e}^{-(\lambda_1 - \lambda_2 + o(1))n}) \quad \text{a.s.} \tag{36}$$

for some random $v^{1\cdot}$. Writing

$$M_n = u_n^{\cdot 1} v^{1\cdot} \sigma_n^1 + \sum_{k=2}^{p} u_n^{\cdot k} v^{k\cdot} \sigma_n^k, \tag{37}$$

we get the following corollary:

Corollary 3.3 *Under the conditions of Proposition* 3.2*, assuming $\lambda_1 > \lambda_2$, we have for some fixed random $v^{1\cdot}$*

$$M_n = u_n^{\cdot 1} v^{1\cdot} \sigma_n^1 + O(\mathrm{e}^{(\lambda_2 + o(1))n}). \tag{38}$$

In the special case when $A_n = A$ for all n, arranging the eigenvalues of A, say ρ_i, according to their absolute values in non-increasing order, we have for all i

$$\lambda_i = \log |\rho_i|. \tag{39}$$

This is easily seen from the definition of λ_i under (28) and the decomposition of $\mathbb{R}^p$ as a direct sum of invariant subspaces corresponding to distinct real or conjugate pairs of complex eigenvalues.

3.2 Sequentially Primitive Non-Negative Matrix Processes

We will now briefly revisit the concept of sequentially primitivity in a systematic exposition.

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

$$\psi_n = \min\{\psi \geq 1 : A_{n+\psi-1} \cdot A_{n+\psi-2} \cdots A_n > 0\}. \tag{40}$$

Since by assumption A_n is row-allowable we will have $M_n > 0$ with strict inequality for all $n \geq \psi_1$.

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

$$\rho_n = \min\{\rho \geq 1 : A_n \cdot A_{n-1} \cdots A_{n-\rho+1} > 0\}. \tag{41}$$

It is easily seen that a two-sided strictly stationary sequence (A_n) is forward sequentially primitive if and only if it is backward sequentially primitive. Moreover, the indices of forward and backward sequential primitivity, ψ_n and ρ_n, have the same distributions, see Lemma 5 of [12].

Furthermore, Theorem 36 of [12] implies:

Lemma 3.5 *Let $A = (A_n)$, $n \geq 1$ be a strictly stationary, ergodic process of non-negative allowable matrices such that $\mathbb{E} \log^+ \|A_1\| < +\infty$, and A is (forward) sequentially primitive. Then $\lambda_1(A) - \lambda_2(A) > 0$.*

A direct corollary is the following lemma, reinforcing Theorem 2.2:

Lemma 3.6 *Let $A = (A_n)$ be as in Theorem 2.2. Then we have $\lambda_1(A) = 0$ and $\lambda_2(A) < 0$.*

4 A Common Eigenvector of (G_n)

In this section we discuss a few remarkable properties of sequences of random matrices (G_n) having a *common right* eigenvector u, needed subsequently, and presented with a tutorial aspect in mind. These auxiliary results will be used for the proof of extensions of Theorems 2.1

and 2.4, and also for arguments that may be relevant for the further analysis of linear gossip algorithms on directed networks, proposed in [21] and [22].

Having said all this consider a strictly stationary ergodic sequence of $p \times p$ matrices (G_n) with a common right eigenvector $u \in \mathbb{R}^p$, having a common eigenvalue $\eta \in \mathbb{R}$:

$$G_n u = \eta u \quad \text{w.p.1} \tag{42}$$

for all n. It is easily seen that under this condition $\log|\eta|$ is an element of the Lyapunov-spectrum of (G_n), i.e., $\log|\eta| \in \Lambda(G)$, see Lemma 4.2 below. But first:

Condition 4.1 *$G = (G_n)$, $-\infty < n < +\infty$ is a* two-sided *strictly stationary, ergodic sequence of $p \times p$ random matrices defined over $(\Omega, \mathcal{F}, P)$. Moreover $\mathbb{E}\log\|G_1\|^+ < \infty$.*

Lemma 4.2 *Let $G = (G_n)$ satisfy Condition 4.1. Assume that the matrices G_n have a common right eigenvector u with the same eigenvalue $\eta \in \mathbb{R}$, so that $G_n u = \eta u$ for all n w.p.1. Then $\log|\eta|$ belongs to the multi-set of Lyapunov-exponents of G: $\log|\eta| \in \Lambda(G)$.*

Remark 4.3 Similarly, if the G_n-s have a common left eigenvector v^{T} such that $v^{\mathrm{T}} G_n = \eta v^{\mathrm{T}}$ then $\log|\eta| \in \Lambda(G^{\mathrm{T}})$. A standard example is when the G_n-s are *column stochastic* matrices, with $v^{\mathrm{T}} = \mathbf{1}^{\mathrm{T}}$ and $\eta = 1$.

Proof First assume that $\eta \neq 0$. Letting $\overline{G}_n = G_n/\eta$, we have $\overline{G}_n u = u$ and the statement of the lemma reduces to $0 \in \Lambda(\overline{G})$. Once this is shown, the claim $\log|\eta| \in \Lambda(G)$ follows by (29). Let us therefore assume that $\eta = 1$.

Define the deterministic one-dimensional subspace

$$U := \{x : x = cu,\ c \in \mathbb{R}^p\}.$$

Recalling Oseledec's theorem, stated as Proposition 3.2, consider the set $\Omega' \subset \Omega$ given there, with $P(\Omega') = 1$. Let $\omega \in \Omega'$ and consider the random sub-spaces $L_\mu = L_\mu(\omega)$ defined under (25). Then for any $\varepsilon > 0$ we have

$$U \cap L_{-\varepsilon}(\omega) = \{0\} \quad \text{and} \quad U \subset L_\varepsilon(\omega),$$

hence the subspace-valued piece-wise constant function $L_\mu(\omega)$ is discontinuous at $\mu = 0$ a.s. Recalling the definition of Lyapunov exponents, see (26), it follows that $\mu = 0$ belongs to the multi-set of Lyapunov-exponents of $\overline{G} = (\overline{G}_n)$, as stated.

The case of $\eta = 0$ can be settled by noting first that $G_n u = 0$ for all n w.p.1 implies $\lambda(u, G, \omega) = -\infty$ w.p.1. Hence $U \subset L_{-\infty}(\omega)$ w.p.1 and thus $L_{-\infty}(\omega)$ is a non-trivial subspace w.p.1. and consequently $-\infty \in \Lambda(G)$. ∎

Remark 4.4 For an exercise: We note that the last conclusion $-\infty \in \Lambda(G)$ holds under the much weaker condition that $P(G_1 \text{ is singular}) > 0$.

For a constant sequence of matrices $G_n \equiv G$, with slight abuse of notations we have

$$\Lambda(G) = \log|\mathrm{Sp}(G)|. \tag{43}$$

Since $\mathrm{Sp}(G) = \mathrm{Sp}(G^{\mathrm{T}})$ we have $\Lambda(G^{\mathrm{T}}) = \Lambda(G)$. However, for a general sequence $G = (G_n)$, satisfying Condition 4.1, we may have $\Lambda(G^{\mathrm{T}}) \neq \Lambda(G)$.

In spite of this, a generalization of the identity $\mathrm{Sp}(G) = \mathrm{Sp}(G^{\mathrm{T}})$ can be established as follows. Let us define the stochastic process $\overleftarrow{G} = (\overleftarrow{G}_n)$ by time-reversal as

$$\overleftarrow{G}_n := G_{-n}, \quad \text{for } -\infty < n < +\infty. \tag{44}$$

Lemma 4.5 *Assume that Condition* 4.1 *holds for the two-way infinite process* $G = (G_n)$. *Then we have* $\Lambda(\overleftarrow{G}^{\mathrm{T}}) = \Lambda(G)$.

Proof Note that Condition 4.1 implies the conditions of Oseledec's theorem given as Proposition 3.2. Recall that for any $1 \leq k \leq p$,

$$\lambda_1(G) + \lambda_2(G) + \cdots + \lambda_k(G) = \lambda_1(G \wedge G \wedge \cdots \wedge G),$$

where the number of terms within the bracket is k. Similarly,

$$\lambda_1(\overleftarrow{G}^{\mathrm{T}}) + \lambda_2(\overleftarrow{G}^{\mathrm{T}}) + \cdots + \lambda_k(\overleftarrow{G}^{\mathrm{T}}) = \lambda_1(\overleftarrow{G}^{\mathrm{T}} \wedge \overleftarrow{G}^{\mathrm{T}} \wedge \cdots \wedge \overleftarrow{G}^{\mathrm{T}}).$$

Setting $H = G \wedge G \wedge \cdots \wedge G$, and noting that wedge product and time reversal combined with transposition are exchangeable, the right hand sides of the above equalities can be written as $\lambda_1(H)$ and $\lambda_1(\overleftarrow{H}^{\mathrm{T}})$, respectively. Thus to prove the lemma it is sufficient to show that for any k we have $\lambda_1(H) = \lambda_1(\overleftarrow{H}^{\mathrm{T}})$.

Obviously, G itself can take the role of H thus it is sufficient to show that

$$\lambda_1(\overleftarrow{G}^{\mathrm{T}}) = \lambda_1(G). \tag{45}$$

Recalling the definition of $\lambda_1(\cdot)$ and the shift-invariance of the process G the above equality is readily obtained:

$$\lambda_1(G) = \lim_n \frac{1}{n} \mathbb{E} \log \|G_n \cdot G_{n-1} \cdots G_1\|,$$
$$\lambda_1(\overleftarrow{G}^{\mathrm{T}}) = \lim_n \frac{1}{n} \mathbb{E} \log \|G_{-n}^{\mathrm{T}} \cdot G_{-n+1}^{\mathrm{T}} \cdots G_{-1}^{\mathrm{T}}\| = \lim_n \frac{1}{n} \mathbb{E} \log \|G_{-1} \cdot G_{-2} \cdots G_{-n}\|.$$

The proof is completed. ∎

Remark 4.6 From Lemma 4.5 it follows that if $G = (G_n)$, $-\infty < n < +\infty$ is i.i.d., then $\Lambda(G^{\mathrm{T}}, \omega) = \Lambda(G, \omega)$, by noting that, in this case, $\overleftarrow{G}$ and G have the same distributions.

Let the orthogonal complement of U be denoted by $S = S_u$:

$$S = S_u := \{x : x \in \mathbb{R}^p,\, u^{\mathrm{T}} x = 0\}. \tag{46}$$

Now, S is an invariant subspace w.r.t. G_n^{T} for all n. Let the restriction of G_n^{T} to S be denoted by G_{nS}^{T}. To clarify the notations take a matrix representation of G_n in a fixed coordinate system in $\mathbb{R}^p$ for all n, such that u is the first unit vector, $u = (1, 0, \cdots, 0)^{\mathrm{T}}$ and the rest of the coordinate vectors being chosen from S. Then we get, identifying the linear operators $G_n, G_n^{\mathrm{T}}, G_{nS}^{\mathrm{T}}$ with their matrix representations:

$$G_n = \begin{pmatrix} \eta & g_n^{\mathrm{T}} \\ \mathbf{0} & G_{nS} \end{pmatrix} \quad \text{and} \quad G_n^{\mathrm{T}} = \begin{pmatrix} \eta & \mathbf{0}^{\mathrm{T}} \\ g_n & G_{nS}^{\mathrm{T}} \end{pmatrix}. \tag{47}$$

Let the random sequence (G_{nS}^{T}), $-\infty < n < +\infty$, read with increasing indices, be denoted by G_S^{T}. Obviously, G_S^{T} satisfies Condition 4.1, implying, by Oseledec's theorem, Proposition 3.2, among others that the Lyapunov spectrum $\Lambda(G_S^{\mathrm{T}})$ is constant w.p.1. Not surprisingly we have the following result:

Lemma 4.7 *Under the conditions of Lemma* 4.2 *we have*

$$\Lambda(G_S^{\mathrm{T}}) \subset \Lambda(G^{\mathrm{T}}). \tag{48}$$

The interpretation of inclusion for multi-sets, $A \subseteq B$, is that every element of A occurs in B with greater or equal multiplicity.

Proof Let $-\infty \le \mu < \infty$ be arbitrary. Let $\Omega' \subset \Omega$ with $P(\Omega') = 1$, defined as the intersection of the subsets of Ω, associated with G_S^{T} and G^{T}, as given in Oseledec's theorem, Proposition 3.2. Then the invariance of S under G_n^{T} implies for $\omega \in \Omega'$

$$L_\mu(G_S^{\mathrm{T}}, \omega) = L_\mu(G^{\mathrm{T}}, \omega) \cap S. \tag{49}$$

It follows that if μ is a point of discontinuity of $L_\mu(G_S^{\mathrm{T}}, \omega)$ then, necessarily, it is also a point of discontinuity of $L_\mu(G^{\mathrm{T}}, \omega)$. Letting μ_r denote the points of discontinuity of $L_\mu(G_S^{\mathrm{T}}, \omega)$ in strictly decreasing order, we also have, omitting the argument ω,

$$L_{\mu_r}(G_S^{\mathrm{T}}) \setminus L_{\mu_{r-1}}(G_S^{\mathrm{T}}) = \big(L_{\mu_r}(G^{\mathrm{T}}) \setminus L_{\mu_{r-1}}(G^{\mathrm{T}})\big) \cap S.$$

Thus the multiplicity of any μ_r in $\Lambda(G_S^{\mathrm{T}})$ can not exceed its multiplicity in $\Lambda(G^{\mathrm{T}})$, hence the claim follows. ∎

Applying the lemma with $\overleftarrow{G}^{\mathrm{T}}$ replacing G^{T} we get:

Corollary 4.8 *Under the conditions of Lemma* 4.2 *we have*

$$\Lambda(\overleftarrow{G}_S^{\mathrm{T}}) \subset \Lambda(\overleftarrow{G}^{\mathrm{T}}) = \Lambda(G). \tag{50}$$

The cardinalities of the multisets $\Lambda(\overleftarrow{G}_S^{\mathrm{T}})$ and $\Lambda(G)$ are $p-1$ and p, respectively, thus the question remains which element of $\Lambda(G)$ has to be removed to get $\Lambda(G_S^{\mathrm{T}})$.

Let $\mu_+ \in \Lambda(\overleftarrow{G}^{\mathrm{T}})$ denote the Lyapunov-exponent defined by

$$\{\mu_+\} := \Lambda(\overleftarrow{G}^{\mathrm{T}}) \setminus \Lambda(\overleftarrow{G}_S^{\mathrm{T}}). \tag{51}$$

Recall the notation for the Lyapunov-exponent of $x \in \mathbb{R}^p$ w.r.t. G as $\lambda(x, G, \omega)$, see (32). Let $\Omega' \in \Omega$ with $P(\Omega') = 1$, described by Oseledec's theorem, Proposition 3.2.

By Lemma 4.2 we have $\lambda(u, G, \omega) = \log|\eta|$ for $\omega \in \Omega'$, motivating the lemma below.

Lemma 4.9 *Let the conditions of Lemma* 4.2 *be satisfied. Then we have*

$$\Lambda(\overleftarrow{G}_S^{\mathrm{T}}) = \Lambda(G) \setminus \{\log|\eta|\}.$$

Proof Note that $\log|\eta| \in \Lambda(G)$, and by Lemma 4.5 $\Lambda(G) = \Lambda(\overleftarrow{G}^{\mathrm{T}})$. First assume $\log|\eta| \notin \Lambda(\overleftarrow{G}_S^{\mathrm{T}})$. Then $\log|\eta| \in \Lambda(G) = \Lambda(\overleftarrow{G}^{\mathrm{T}})$ implies the claim of the lemma.

The question is how to complete the proof for the case $\eta \in \Lambda(\overleftarrow{G}_S^{\mathrm{T}})$. We may be tempted to use an appropriate result of the extensive literature on the continuity of Lyapunov-exponents. However, we may proceed by some kind of simple interpolation argument.

Define a class of processes $G(c)$ by a modification of (47) so that the $(1,1)$ element η is replaced by an arbitrary $c \in \mathbb{R}$:

$$G_n(c) = \begin{pmatrix} c & g_n^{\mathrm{T}} \\ \mathbf{0} & G_{nS} \end{pmatrix} \quad \text{and} \quad G_n^{\mathrm{T}}(c) = \begin{pmatrix} c & \mathbf{0}^{\mathrm{T}} \\ g_n & G_{nS}^{\mathrm{T}} \end{pmatrix}.$$

Obviously, S remains invariant for $G_n^{\mathrm{T}}(c)$ for any c, actually $G_{nS}^{\mathrm{T}}(c) = G_{nS}^{\mathrm{T}}$. Now, by direct inspection we get that for all $\omega \in \Omega$ the Lyapunov-exponent $\lambda(u, G(c), \omega)$ becomes $\log|c|$, thus $\log|c| \in \Lambda(G(c))$. Thus we get

$$\Lambda(G(c)) = \Lambda(G) \cup \{\log|c|\} \setminus \{\log|\eta|\}. \tag{52}$$

Following (51), write $\{\mu_+(c)\} := \Lambda(\overleftarrow{G}^{\mathrm{T}}(c)) \setminus \Lambda(\overleftarrow{G}_S^{\mathrm{T}})$ for any $c \in \mathbb{R}$. With this notation we have

$$\Lambda(G) = \Lambda(G(\eta)) = \Lambda(\overleftarrow{G}^{\mathrm{T}}(\eta)) = \Lambda(\overleftarrow{G}_S^{\mathrm{T}}) \cup \{\mu_+(\eta)\}. \tag{53}$$

Substituting this into (52) we get

$$\Lambda(G(c)) = \Lambda(\overleftarrow{G}_S^{\mathrm{T}}) \cup \mu_+(\eta) \cup \{\log|c|\} \setminus \{\log|\eta|\}. \tag{54}$$

Let us now choose c such that $\log|c| \notin \Lambda(\overleftarrow{G}_S^{\mathrm{T}})$. By the first part of the proof we have $\mu_+(c) = \log|c|$. Thus we have

$$\Lambda(G(c)) = \Lambda(\overleftarrow{G}^{\mathrm{T}}(c)) = \Lambda(\overleftarrow{G}_S^{\mathrm{T}}) \cup \{\mu_+(c)\} = \Lambda(\overleftarrow{G}_S^{\mathrm{T}}) \cup \{\log|c|\}. \tag{55}$$

Matching the right hand sides of the above two equalities it follows that

$$\Lambda(\overleftarrow{G}_S^{\mathrm{T}}) \cup \{\mu_+(\eta)\} \cup \{\log|c|\} \setminus \{\log|\eta|\} = \Lambda(\overleftarrow{G}_S^{\mathrm{T}}) \cup \{\log|c|\}, \tag{56}$$

and from here we conclude $\mu_+(\eta) = \log|\eta|$ as stated. ∎

The above lemma and its proof implies the following corollary:

Corollary 4.10 *Under the conditions of Lemma* 4.2 *we have*

$$\Lambda(\overleftarrow{G}^{\mathrm{T}}) = \Lambda(\overleftarrow{G}_S^{\mathrm{T}}) \cup \{\log|\eta|\}.$$

A remarkable feature of the above corollary is that $\Lambda(\overleftarrow{G^{\mathrm{T}}})$, and hence $\Lambda(G)$, is completely determined by $\Lambda(\overleftarrow{G}_S^{\mathrm{T}})$ and $\lambda(u, G, \omega) = \log|\eta|$, allowing a certain degree of freedom in extending G_S^{T} to G^{T}, that is choosing $g_n \in \mathbb{R}^{p-1}$ in (47), subject to the conditions imposed onto G itself in Condition 4.1. This phenomenon will be explored below in Lemmas 5.2 and 5.3.

Tilting our discussion towards the proof of the upcoming Theorem 5.1 we show that that we can enforce the condition $G_n u = 0$ for all n by modifying G_n as follows:

Lemma 4.11 *Let $G = (G_n)$ be as in Lemma 4.2, and let $z \in \mathbb{R}^p$ be such that $u^{\mathrm{T}} z \neq 0$. Define the modified sequence of matrices $H = (H_n)$ by*

$$H_n = G_n - \eta \cdot uz^{\mathrm{T}}/u^{\mathrm{T}}z. \tag{57}$$

Then

$$\Lambda(H) = \Lambda(G) \setminus \{\log|\eta|\} \cup \{-\infty\}. \tag{58}$$

Proof Note that $H_n u = 0$ for all n. Thus u is a common eigenvector of H_n-s with eigenvalue $\eta_0 = 0$. Obviously, S is an invariant subspace for H_n^{T}, and $H_{nS}^{\mathrm{T}} = G_{nS}^{\mathrm{T}}$ for all n. Indeed, for $x \in S$,

$$x^{\mathrm{T}} H_n = x^{\mathrm{T}}(G_n - \eta \cdot uz^{\mathrm{T}}/u^{\mathrm{T}}z) = x^{\mathrm{T}} G_n \tag{59}$$

implying $H_n^{\mathrm{T}} x \in S$. Thus, by Lemma 4.9,

$$\Lambda(H) = \Lambda(\overleftarrow{H}_S^{\mathrm{T}}) \cup \{\log \eta_0\} = \Lambda(\overleftarrow{G}_S^{\mathrm{T}}) \cup \{-\infty\}. \tag{60}$$

Applying Lemma 4.9 once again, and writing $\Lambda(\overleftarrow{G}_S^{\mathrm{T}}) = \Lambda(G) \setminus \{\log|\eta|\}$ we get the claim. ∎

In the special case when $\log|\eta|$ is the maximal Lyapunov-exponent we get the following corollary:

Corollary 4.12 *Under the conditions of Lemma 4.11, and the additional condition that $\log|\eta| = \lambda_1(G)$ we have*

$$\lambda_1(H) = \lambda_2(G). \tag{61}$$

A special case of the above is the following corollary:

Corollary 4.13 *Let $A = (A_n)$ satisfy the conditions of Theorem 2.2, and define $B = (B_n)$ by*

$$B_n = A_n - \mathbf{1}\mathbf{1}^{\mathrm{T}}/\mathbf{1}^{\mathrm{T}}\mathbf{1}. \tag{62}$$

Then $\lambda_1(B) = \lambda_2(A) < 0$.

Remark 4.14 It is worth mentioning that some of the elementary arguments of this section can be abridged by the application of a not quite elementary result saying that if G_n is upper block-triangular for all n, with square diagonal blocks A_n, D_n of fixed size, say

$$G_n = \begin{pmatrix} A_n & B_n \\ 0 & D_n \end{pmatrix}, \tag{63}$$

then

$$\Lambda(G) = \Lambda(A) \cup \Lambda(D). \tag{64}$$

5 Rank-1 Limits of Random Products

In this section we present an extension of Theorems 2.1 and 2.4 together with their proofs, and based on this extension we provide an attractive technical tool that may be relevant for the further analysis of the linear gossip algorithm proposed in [22].

We consider a strictly stationary ergodic sequence of $p \times p$ matrices $G = (G_n)$, $-\infty < n < +\infty$ with common right and left eigenvectors u and v^{T}, both in $\mathbb{R}^p$, with eigenvalues $\eta \in \mathbb{R}$: For all $n \in \mathbb{Z}$,

$$G_n u = \eta u \quad \text{and} \quad v^{\mathrm{T}} G_n = \eta v^{\mathrm{T}} \quad \text{w.p.1.} \tag{65}$$

Our objective is to establish a.s. convergence of the products of $G_n \cdots G_1$ to a rank-1 matrix. If $\eta \neq 0$, then we can reduce the problem to the case $\eta = 1$ by scaling G_n, see below. The following theorem extends Theorems 2.1 and 2.4:

Theorem 5.1 *Let $G = (G_n)$, $-\infty < n < +\infty$ be as in Condition* 4.1*, satisfying* (65) *with eigenvalues $\eta = 1$. Assume that* 0 *is the top Lyapunov exponent in $\Lambda(G)$, i.e., $\lambda_1(G) = 0$, and the spectral gap is positive: $\lambda_2(G) < 0$. Then*

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| G_n \cdot G_{n-1} \cdots G_1 - \frac{uv^{\mathrm{T}}}{u^{\mathrm{T}} v} \right\| = \lambda_2(G) < 0 \quad \text{a.s.} \tag{66}$$

Apart from the condition that (G_n) is a two-sided process, the theorem above is a significant extension of the main result of [9], Theorem 5.2, inasmuch the individual matrices G_n can be very general: No structure is imposed and non-negativity is not assumed. The above theorem also significantly complements the main result of [22], Theorem 2, establishing rank-1 limit almost surely and in L_2-sense for a class of i.i.d. processes (G_n) with very special structures for the individual matrices G_n, see Section 6 and in particular Proposition 6.1.

Proof First note that

$$G_n \cdot G_{n-1} \cdots G_1 - \frac{uv^{\mathrm{T}}}{u^{\mathrm{T}} v} = \left(G_n - \frac{uv^{\mathrm{T}}}{u^{\mathrm{T}} v} \right) \cdots \left(G_1 - \frac{uv^{\mathrm{T}}}{u^{\mathrm{T}} v} \right), \tag{67}$$

as easily seen by induction. Define the matrices H_n by

$$H_n = G_n - \frac{uv^{\mathrm{T}}}{u^{\mathrm{T}} v}. \tag{68}$$

Obviously, u and v^{T} are right and left eigenvectors of H_n with 0 eigenvalue for all n. It is also obvious that the random sequence $H = (H_n)$ satisfies Condition 4.1.

By the Fürstenberg-Kesten theorem[26] supplemented by Oseledelec' theorem[25], restated above as Propositions 3.1 and 3.2, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \| H_n \cdot H_{n-1} \cdots H_1 \| = \lambda_1(H) \quad \text{a.s.} \tag{69}$$

But by Corollary 4.12 and our assumption on the spectral gap we have $\lambda_1(H) = \lambda_2(G) < 0$, as stated. ∎

Assume for a moment that $G_n \equiv G$ is fixed, and represented in analogy with (47) as

$$G = \begin{pmatrix} \eta & g^{\mathrm{T}} \\ \mathbf{0} & G_S \end{pmatrix} \quad \text{and} \quad G^{\mathrm{T}} = \begin{pmatrix} \eta & \mathbf{0}^{\mathrm{T}} \\ g & G_S^{\mathrm{T}} \end{pmatrix}. \tag{70}$$

It is readily seen that the spectrum $\mathrm{Sp}(G)$ is not affected by changing the value of g^{T}. In a coordinate-free form this amounts to adding a matrix ux^{T} with $x \in S = S_u$. A remarkable extension of this simple fact is given in the following lemma.

Lemma 5.2 *Let (G_n^*) be a random sequence of matrices satisfying the conditions of Lemma* 4.2. *Let us set, as usual, $S = S_u := \{x : x \in \mathbb{R}^p,\ u^{\mathrm{T}}x = 0\}$. Let $x_n \in S_u$ be a random sequence such that (G_n^*, x_n) is strictly stationary, and $\mathbb{E}\log^+|x_n| < \infty$. Define $G_n = G_n^* + ux_n^{\mathrm{T}}$. Then*

$$\Lambda(G) = \Lambda(G^*). \tag{71}$$

Proof We have $G_n u = G_n^* u + ux_n^{\mathrm{T}}u = \eta u$ for all n. Furthermore, $G_{nS}^{\mathrm{T}} = G_{nS}^{*\mathrm{T}}$ for all n. Indeed, for any $x \in S_u$ we have $x^{\mathrm{T}}G_n = x^{\mathrm{T}}G_n^* + x^{\mathrm{T}}ux_n^{\mathrm{T}} = x^{\mathrm{T}}G_n^*$. Thus Lemma 4.9 implies the claim. ∎

A reciprocal result holds when the matrices (G_n^*) have a common left eigenvector:

Lemma 5.3 *Let (G_n^*), $-\infty < n < +\infty$ be a random sequence of matrices satisfying Condition* 4.1 *with a common left eigenvector $v \neq 0$, so that $v^{\mathrm{T}}G_n^* = \eta v^{\mathrm{T}}$ for all n. Let us set $S = S_v := \{y : y \in \mathbb{R}^p,\ v^{\mathrm{T}}y = 0\}$. Let $y_n \in S_v$ be a random sequence such that (G_n^*, y_n) is strictly stationary, and $\mathbb{E}\log^+|y_n| < \infty$. Define $G_n = G_n^* + y_n v^{\mathrm{T}}$. Then*

$$\Lambda(G) = \Lambda(G^*). \tag{72}$$

Proof The random sequence of matrices $(\overleftarrow{G}_n^{*\mathrm{T}})$ satisfies the conditions of Lemma 5.2 with the common right eigenvector v. Likewise, $(\overleftarrow{G}_n^{*\mathrm{T}}, y_n)$ satisfies the conditions of the lemma with y_n replacing x_n and S_v replacing S_u. Lemma 5.2 implies

$$\Lambda(\overleftarrow{G}^{\mathrm{T}}) = \Lambda(\overleftarrow{G}^{*\mathrm{T}}). \tag{73}$$

The lemma now follows in virtue of Lemma 4.5. ∎

The above lemma opens an avenue to constructing a random sequence of matrices (G_n) with a prescribed common right eigenvector u having the same eigenvalue η, so that $G_n u = \eta u$ for all n, while keeping $\Lambda(G) = \Lambda(G^*)$.

Lemma 5.4 *Let (G_n^*) be a random sequence of matrices satisfying the conditions of Lemma* 5.3. *Let $u \neq 0$ be a prescribed common right eigenvector, having the same eigenvalue η as v^{T}, such that $v^{\mathrm{T}}u \neq 0$. Then there exists a unique sequence (y_n), given by $y_n = (\eta I - G_n^*)u/v^{\mathrm{T}}u$, such that the modified random sequence*

$$G_n = G_n^* + y_n v^{\mathrm{T}} \tag{74}$$

satisfies

$$v^{\mathrm{T}}G_n = \eta v^{\mathrm{T}} \quad \textit{and} \quad G_n u = \eta u. \tag{75}$$

The sequence (G_n) satisfies Condition 4.1 *and $\Lambda(G) = \Lambda(G^*)$.*

Proof Consider the equations

$$G_n u = (G_n^* + y_n v^{\mathrm{T}})u = \eta u.$$

This is a linear equation for y_n in $p-1$ unknown, since $y_n \in S_v$, while the number of equations is also $p-1$, due to scale invariance of u. The unique solution is given by

$$y_n = (\eta I - G_n^*)u/v^{\mathrm{T}}u. \tag{76}$$

It is readily seen that $v^{\mathrm{T}} y_n = 0$, or $y_n \in S_v$. Obviously, (G_n^*, y_n) is strictly stationary, and $\mathbb{E} \log^+ |y_n| < \infty$, as required. Thus $\Lambda(G) = \Lambda(G^*)$ follows from Lemma 5.3. ∎

A nice application of Lemma 5.4 that is potentially relevant for linear gossip algorithms proposed in [22] is as follows:

Theorem 5.5 *Let (G_n^*) be a random sequence of $2p \times 2p$ matrices satisfying the conditions of Lemma 5.3. Let*

$$v = \begin{pmatrix} \mathbf{1}_p \\ \mathbf{1}_p \end{pmatrix} \quad \textit{and} \quad u = \begin{pmatrix} \mathbf{1}_p \\ \mathbf{0}_p \end{pmatrix}. \tag{77}$$

Assume that v is a common left eigenvector of G_n^*-s with common eigenvalues $\eta = 1$, and let u be a prescribed common right eigenvector, with common eigenvalues $\eta = 1$ for a suitable modification of (G_n^*). Assume that (G_n^*) is non-negative and (forward) sequentially primitive. Define

$$G_n = G_n^* + (I_{2p} - G_n^*)\, u v^{\mathrm{T}} / p. \tag{78}$$

Then for (G_n) we have $\Lambda(G) = \Lambda(G^*)$, and almost surely

$$\lim_{n \to \infty} \frac{1}{n} \log \left\| G_n \cdot G_{n-1} \cdots G_1 - u v^{\mathrm{T}} / p \right\| = \lambda_2(G^*) < 0.$$

The novelty of the above result compared to Theorem 5.1 is that the existence of common left and right eigenvectors, with eigenvalue 1, is guaranteed by construction, together with other technical conditions, as the proof will show.

Proof By Lemma 5.4, writing $\mathbf{1}$ for $\mathbf{1}_p$ and $\mathbf{0}$ for $\mathbf{0}_p$, the matrices G_n satisfy

$$(\mathbf{1}^{\mathrm{T}},\ \mathbf{1}^{\mathrm{T}}) G_n = (\mathbf{1}^{\mathrm{T}},\ \mathbf{1}^{\mathrm{T}}) \quad \text{and} \quad G_n \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix}, \tag{79}$$

furthermore $\Lambda(G) = \Lambda(G^*)$. Now, by the Perron-Frobenius theorem and (43) the top-Lyapunov-exponent of (G_n^*) is 0. Furthermore, by Lemma 3.5 the spectral gap of (G_n^*) is positive, hence $\lambda_2(G^*) < 0$. Thus, by Theorem 5.1 we get that $\lambda_2(G^*)$ is equal to the a.s. limit

$$\lim_{n \to \infty} \frac{1}{n} \log \left\| G_n \cdot G_{n-1} \cdots G_1 - \begin{pmatrix} \mathbf{1} \\ \mathbf{0} \end{pmatrix} (\mathbf{1}^{\mathrm{T}}, \mathbf{1}^{\mathrm{T}}) / p \right\|.$$

The proof is completed. ∎

The above construction is mathematically attractive, but the resulting matrices are unfortunately not sparse, in contrast to the construction presented in [22]. A partial ramification of this may be setting up a fixed point problem for $\overline{x}$, but this is beyond the scope of this paper.

Let us thus restrict attention to column-stochastic matrices, in particular, we assume $v^{\mathrm{T}} G_n^* = v$ with $v = \mathbf{1}_{2p}$. Then the boundedness of the products $G_n \cdot \cdots \cdot G_1$, $n \geq 1$ can be established:

Lemma 5.6 *Consider a sequence of column-stochastic matrices G_n^*, satisfying the conditions of Lemma 5.4. Define G_n, following (74), by*

$$G_n = G_n^* + y_n v^{\mathrm{T}} = G_n^* + (I - G_n^*) u v^{\mathrm{T}} / v^{\mathrm{T}} u. \tag{80}$$

Then the products $G_n \cdot \dots \cdot G_1$, $n \geq 1$ stay bounded.

Proof First we show that

$$G_n \cdot G_{n-1} \cdots G_1 = G_n^* \cdot G_{n-1}^* \cdots G_1^* + \sum_{m=1}^{n} G_n^* \cdot G_{n-1}^* \cdots G_{m+1}^* y_m v^{\mathrm{T}}.$$

Proceed by induction. The proposition is certainly true for $n = 1$. Let us assume that it is true for some n. Then

$$\begin{aligned} &G_{n+1} \cdot G_n \cdots G_1 \\ =&(G_{n+1}^* + y_{n+1} v^{\mathrm{T}}) \cdot G_n \cdots G_1 \\ =&(G_{n+1}^* + y_{n+1} v^{\mathrm{T}}) \cdot G_n^* \cdots G_1^* + (G_{n+1}^* + y_{n+1} v^{\mathrm{T}}) \cdot \sum_{m=1}^{n} G_n^* \cdot G_{n-1}^* \cdots G_{m+1}^* \, y_m v^{\mathrm{T}}. \end{aligned}$$

Noting that $v^{\mathrm{T}} G_m^* = v^{\mathrm{T}}$ and that $v^{\mathrm{T}} y_m = 0$ for all m the claim follows by direct inspection.

Now, the sum on right hand side becomes

$$\sum_{m=1}^{n} G_n^* \cdot G_{n-1}^* \cdots G_{m+1}^* \, (I - G_m^*) u v^{\mathrm{T}} / v^{\mathrm{T}} u,$$

which is a telescopic sum, yielding

$$G_n \cdot G_{n-1} \cdots G_1 = G_n^* \cdot G_{n-1}^* \cdots G_1^* - G_n^* \cdot G_{n-1}^* \cdots G_1^* \, u v^{\mathrm{T}} / v^{\mathrm{T}} u + u v^{\mathrm{T}} / v^{\mathrm{T}} u.$$

Since the products $G_n^* \cdot G_{n-1}^* \cdots G_1^*$ are bounded, the claim follows. ∎

Corollary 5.7 *Let the conditions of Lemma* 5.6 *be satisfied, and let* $0 = \lambda_1(G^*) > \lambda_2(G^*)$. *Then it holds in* $L_2(\Omega, \mathcal{F}, P)$:

$$\lim_{n \to \infty} \mathbb{E} \, \| G_n \cdot G_{n-1} \cdots G_1 - u v^{\mathrm{T}} / u^{\mathrm{T}} v \|^2 = 0. \tag{81}$$

Unfortunately, with the techniques used above, the rate of L_2-convergence can not be established in this generality.

6 Revisiting Linear Gossip

6.1 An Informal Description

In this section we provide the details of the linear gossip algorithm for directed networks proposed in [22], discuss the challenges faced in its analysis, and propose a reasonable heuristics for rectifying some of the deficiencies of the original algorithm. We begin with an informal description of the algorithm of [22] highlighting its logic.

Following [22], once an ordered pair (i, j) is selected, with i being the sender and j being the receiver, the transformation of values and accounts at these nodes is carried out following an a priori fixed communication protocol. In the first step of the transaction at time n node i communicates its value x_{n-1}^i to node j without changing its own value, and at the same clears its account y_{n-1}^i to 0 by transferring it to node j.

The receiver node j updates its own value in several steps. First, having learnt the value of node i, he/she updates its own value to the convex combination $\alpha x^i_{n-1} + (1-\alpha)\, x^j_{n-1}$, with a fixed $0 < \alpha < 1$. Equivalently, he/she corrects its own value by adding $\alpha(x^i_{n-1} - x^j_{n-1})$.

Next, he/she then transfers a fixed fraction γ of the received account y^i_{n-1}, with $0 < \gamma < 1$, to its own value, while transferring the rest to its own account. Similarly, he/she then transfers a fixed fraction β of its own account y^j_{n-1} to its own value, with $0 < \beta < 1$, while leaving the rest on its own account. The last two actions play the role of feedback.

Finally, node j updates its account by compensating the first step, when correcting of its value by $\alpha(x^i_{n-1} - x^j_{n-1})$: The opposite amount $-\alpha(x^i_{n-1} - x^j_{n-1})$ that would have been moved to node i in the setup of the classical, linear, bidirectional gossip algorithm, a move prohibited in the current setup, is moved to the account of node j.

6.2 A Formal Description

A formal description of the above algorithm, in terms of the current variables at the ordered pair of nodes (i, j), is given as follows:

$$
\begin{aligned}
x^i_n &= x^i_{n-1}, \quad x^j_n = \alpha x^i_{n-1} + (1-\alpha) x^j_{n-1} + \gamma y^i_{n-1} + \beta y^j_{n-1}, \\
y^i_n &= 0, \quad y^j_n = -\alpha x^i_{n-1} + \alpha x^j_{n-1} + (1-\gamma) y^i_{n-1} + (1-\beta) y^j_{n-1}.
\end{aligned}
$$

It is easily seen that, in complete analogy with (11) in [21] for the case $\gamma = 0$, the last equation can be written as

$$
y^j_n = y^j_{n-1} + y^i_{n-1} - (x^j_n - x^j_{n-1}), \tag{82}
$$

reflecting the role of the accounts y^k to keep the the overall sum of the variables x^k, y^k unchanged.

Note that if pairs are chosen sequentially, the receiver at any time becoming the sender at the next time moment, then, at the end of a transaction at time n, we will have $y^k_n = 0$ for all k except for k being equal to the index of the latest receiver.

We may wish to rewrite the algorithm using matrix-vector notations. Assume $i < j$. To describe the algebra of the effect a transaction from node i to node j we introduce the vectors

$$
\overline{z}_{n-1} := \begin{pmatrix} \overline{x}_{n-1} \\ \overline{y}_{n-1} \end{pmatrix} := \begin{pmatrix} x^i_{n-1} \\ x^j_{n-1} \\ y^i_{n-1} \\ y^j_{n-1} \end{pmatrix}, \tag{83}
$$

and the time-independent matrices

$$
\overline{F} := \begin{pmatrix} \overline{A} & \overline{B} \\ \overline{C} & \overline{D} \end{pmatrix} := \begin{pmatrix} 1 & 0 & 0 & 0 \\ \alpha & 1-\alpha & \gamma & \beta \\ 0 & 0 & 0 & 0 \\ -\alpha & \alpha & 1-\gamma & 1-\beta \end{pmatrix}.
$$

We note in passing that in [21] the authors have chosen $\beta = 0$.

Then the vector of new values and auxiliary variables at nodes i, j will be given by

$$\overline{F}\overline{z}_{n-1} = \begin{pmatrix} \overline{A} & \overline{B} \\ \overline{C} & \overline{D} \end{pmatrix} \cdot \begin{pmatrix} \overline{x}_{n-1} \\ \overline{y}_{n-1} \end{pmatrix}. \tag{84}$$

Let

$$\overline{v} = \begin{pmatrix} \mathbf{1}_2 \\ \mathbf{1}_2 \end{pmatrix} \quad \text{and} \quad \overline{u} = \begin{pmatrix} \mathbf{1}_2 \\ \mathbf{0}_2 \end{pmatrix}. \tag{85}$$

Not surprisingly, the matrix $\overline{F}$ satisfies

$$\overline{v}^{\mathrm{T}}\overline{F} = \overline{v}^{\mathrm{T}} \quad \text{and} \quad \overline{F}\overline{u} = \overline{u}. \tag{86}$$

6.3 Random Positioning

In order to describe the overall dynamics in matrix-vector notation, recall the definition of the state z_n in (13), and find the dynamics it follows, see (16).

To capture the algebra of constructing F_n from $\overline{F}$ we proceed as in Section 2, considering now the problem of positioning the fixed square matrix $\overline{F}$ of size 4×4 into a unit matrix I_{2p}, by replacing one of its randomly selected 4×4 minor I_4, corresponding to the row and column indices $i, j, i+p, j+p$, by $\overline{F}$.

Let the indices of the sender and the receiver at time n be $i = i_n$, $j = j_n$. Recalling the definition of a $p \times 2$ selector matrix $S(i,j) = (\boldsymbol{e}_i, \boldsymbol{e}_j)$, see (7), and setting $S_n := S(i,j)$, define the $2p \times 4$ selector matrix by

$$I_2 \otimes S(i,j) = I_2 \otimes S_n. \tag{87}$$

Then we blow up the 4×4 matrix $\overline{F}$ to a $2p \times 2p$ matrix, multiplying it by the above selector and its transpose from the left and the right, and express F_n, defined in [22], as follows:

$$F_n = I_{2p} - (I_2 \otimes S_n)(I_4 - \overline{F}_n)(I_2 \otimes S_n)^{\mathrm{T}}. \tag{88}$$

It is readily seen that the matrices F_n satisfy

$$F_n u = u \quad \text{and} \quad v^{\mathrm{T}} F_n = v^{\mathrm{T}}, \tag{89}$$

for all n, as postulated in (14) and (15).

The *key problem* is then to formulate conditions for $\overline{F}$ and the selection process S_n ensuring $0 = \lambda_1(F) > \lambda_2(F)$.

6.4 Convergence

While the heuristics of the above algorithm is appealing, simulation results indicate that, with arbitrary choice of α, β, γ, it may fail to converge to average consensus, see Section 7. On the other hand, a.s. and L_2-convergence is stated for the case $\alpha = \beta = \gamma = 1/2$ in [22, Theorem 2], as follows:

Proposition 6.1 *Consider the linear gossip algorithm defined above with $\alpha = \beta = \gamma = 1/2$. Let the communicating pairs (i, j) be selected independently with probabilities $p_{ij} = p_{ji} > 0$ for all $(i, j) \in E$. Assume that the network is connected. Then average consensus takes place a.s. with exponential rate, in particular, (21) holds with some $\lambda_2 < 0$. In addition, average consensus takes place also in L_2 with exponential rate.*

A distinctive feature of the above result is that L_2-convergence is established, known for the classic, symmetric linear gossip algorithms, but being an open challenge for ratio consensus. In fact, L_2-convergence with exponential rate is the main message of the theorem (and of the proof), from which a.s. convergence with exponential rate is easily derived via the Borel-Cantelli lemma.

The selection of communicating pairs being i.i.d. a key tool of [22] is the analysis of the spectrum of the matrix

$$\mathbb{E}\left[(F_1 - uv^{\mathrm{T}}/p) \otimes (F_1 - uv^{\mathrm{T}}/p)\right]. \tag{90}$$

The authors also suggest that their method can be adjusted to work for any choice of $0 < \alpha, \beta, \gamma < 1$. However, systematic simulations indicate that the extension of [22, Theorem 2] in such generality might not be true, see Section 7.

On the other hand, the same simulation results do indicate that the proposed algorithm works as expected whenever $\alpha \leq 1/2$.

It should be mentioned that a similar convergence result, for i.i.d. communication protocols, has been established in [21], under the conditions that $\beta = 0$ and γ is sufficiently small, using matrix perturbation theory for the spectral analysis of (90).

Finally, simulation results suggest, under the condition $\beta = \gamma = 1/2$, that sequential choice of communicating pairs yields superior convergence for $\alpha \leq 0.75$. However, it is an open question if $0 = \lambda_1(F) > \lambda_2(F)$ holds for these scenarios.

6.5 Closing the Loop

A potentially useful alternative viewpoint is to describe the evolution of the variables x_n, y_n individually, and thus interpreting their interaction as a feedback effect. To this end, write F_n using $p \times p$ blocks as

$$F_n = \begin{pmatrix} A_n & B_n \\ C_n & D_n \end{pmatrix}. \tag{91}$$

The dynamics of (x_n, y_n) is then given by

$$x_n = A_n x_{n-1} + B_n y_{n-1}, \tag{92}$$

$$y_n = C_n x_{n-1} + D_n y_{n-1}. \tag{93}$$

The benefit of this representation is that we may consider (92) and (93) independently as coupled linear systems with states x and y respectively. Noting that

$$F_n - uv^{\mathrm{T}}/p = \begin{pmatrix} A_n - \mathbf{1}_p \mathbf{1}_p^{\mathrm{T}}/p & B_n - \mathbf{1}_p \mathbf{1}_p^{\mathrm{T}}/p \\ C_n & D_n \end{pmatrix}, \tag{94}$$

a further benefit of this representation is that the (forward) products of state matrices $A_n - \mathbf{1}_p\mathbf{1}_p^{\mathrm{T}}/p$ and D_n are beautifully characterized by Corollary 4.13 and Lemma 6.2 below. Thus the analysis leading to these results may give a clue to the analysis of the feedback system (92)–(93).

We end this section with a piece of recreational mathematics. Let $\overline{D}$ be a fixed square matrix of size 2×2. Let us position it into a unit matrix I_p, with $p > 2$, by replacing one of its 2×2 minor by $\overline{D}$, using an appropriate $p \times 2$ selector matrix $S = S(i, j)$, see (7):

$$D(S) := I_p - S(I_2 - \overline{D})S^{\mathrm{T}}. \tag{95}$$

Lemma 6.2 *Assume that $\overline{D}$ is a contraction w.r.t. the Euclidean norm in $\|\cdot\|$. Let (S_n), $n \geq 1$ be a strictly stationary ergodic process of selector matrices defined by ordered subsets of indices (i, j) of size 2, called K_n, such that $\mathbb{P}(i \in K_n) > 0$ for all $i = 1, \cdots, p$. Then for $D = (D_n)$, $n \geq 1$ constructed by (95) with $S = S_n$, we have $\lambda_1(D) < 0$.*

The appeal of this result is that it is readily seen that

$$\lim_{n\to\infty} D_n \cdot D_{n-1} \cdots D_1\, y_0 = 0 \quad \text{a.s.}, \tag{96}$$

for any initial condition $y_0 \in \mathbb{R}^p$. A cheap way of seeing this is to refer to the fact that the matrices D_n are paracontractions w.r.t. the Euclidean norm $\|\cdot\|$.

Indeed, for all n, and for any fixed point $y^* \in \mathbb{R}^p$ of D_n, and any $y \in \mathbb{R}^p$ we have either $\|D_n y - y^*\| < \|y - y^*\|$ or $D_n y = y$. Since the matrices D_n have a unique common fixed point $y^* = 0$, (96) follows from Theorem 1 of [27].

The added value of our Lemma 6.2 is that a.s. exponential rate of convergence is established. The proof of this is left to the reader.

7 Simulation Results

Recalling that the convergence of the algorithm as stated in [22, Theorem 2] has not been established for general choice of parameters (α, β, γ), one may be curious in assessing the stability properties of the products of (F_n) defined by [22]. With this motivation in mind we have performed extensive simulation studies with 5M random choices of parameters on a specific network.

We consider a complete graph with $p = 50$ nodes, representing symmetric all-to-all connections. Transactions are performed with an i.i.d. choice of communicating pairs with equal probabilities. Testing various 3-tuples of (α, β, γ) for possible stability we introduce a quantitative measure by re-expressing the error of interest $F_n \cdots F_1 z_0 - \overline{x}u$ as follows. First, note that we can write

$$\overline{x}u = u\,\frac{u^{\mathrm{T}} z_0}{u^{\mathrm{T}} u} = \begin{pmatrix} J_p & 0 \\ 0 & 0 \end{pmatrix} z_0,$$

with $J_p = \mathbf{1}_p\mathbf{1}_p^{\mathrm{T}}/p$, since the y component of the initial z_0 is 0. Furthermore, thanks to the fact that $F_n u = u$ for all n, we have

$$F_n \cdot F_{n-1} \cdots F_1 \begin{pmatrix} J_p & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} J_p & 0 \\ 0 & 0 \end{pmatrix}, \tag{97}$$

from where for the error we get

$$F_n \cdot F_{n-1} \cdots F_1 z_0 - \overline{x} u = F_n \cdot \cdots \cdot F_1 \begin{pmatrix} I_p - J_p & 0 \\ 0 & 0 \end{pmatrix} z_0.$$

Using this representation of the error, we readily conclude that the validity of (21) implies that the expression above converges to 0 at exponential speed for *any* $z_0 \in \mathbb{R}^{2p}$, without requiring $y_0 = 0$. In this case it follows that a.s.

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| F_n \cdot F_{n-1} \cdots F_1 \begin{pmatrix} I_p - J_p & 0 \\ 0 & 0 \end{pmatrix} \right\| \leq \lambda_2 < 0.$$

In the current setup a sequence of matrices (F_n) will be called stable if the above inequality is satisified, i.e., the norms decay exponentially a.s.

Hence it is reasonable to quantify the error by

$$R_n = \log \left\| F_n \cdot F_{n-1} \cdots F_1 \begin{pmatrix} I_p - J_p & 0 \\ 0 & 0 \end{pmatrix} \right\|,$$

providing a measure of discrepancy from consensus. In the current set of simulations, we fix the number of steps as $n = 5000$.

We considered 5 000 000 random (α, β, γ) 3-tuples, and for each of them we checked if $R_n < 0$. The 3-tuples for which $R_n < 0$, indicating potential stability, form a cloud of points within $[0, 1]^3$. Simulation results indicate the validity of the following conjecture, essentially saying that if agents simultaneously decrease their confidence in their assessment of the values (and thus increase the influence of other agents) by increasing α, it might generate instability.

Conjecture 7.1 Assume that a triple (α, β, γ) generates a stable sequence of matrices (F_n). Then this property is inherited for any other triple by decreasing α only.

Assuming the validity of Conjecture 7.1 we may define a critical value $\alpha^*(\beta, \gamma)$ below which stability is ensured. This function of $(\beta, \gamma) \in [0, 1]^2$ is visualised via a contour plot in Figure 1, with the levels identified by the color scale on the right side of the figure.

With the current set of randomly chosen data the evaluation of the critical value $\alpha^*(\beta, \gamma)$ is based on samples with (β', γ') taking their values in a small neighborhood of the reference point (β, γ), and collecting all instances, for which $R_n < 0$, while bookkeeping the corresponding α'-s. This procedure necessarily causes random fluctuations visible on Figure 1, and also on Figures 2–3 that follow. A remarkable finding of these experiments is that that $\alpha \leq 1/2$ seems always to be stable.

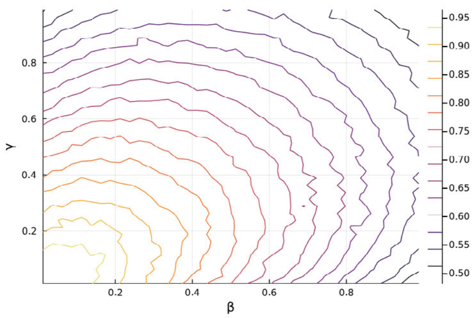

Figure 1 Contour plot of the simulated critical α values as seen after 5000 transmissions on a 50 node complete graph

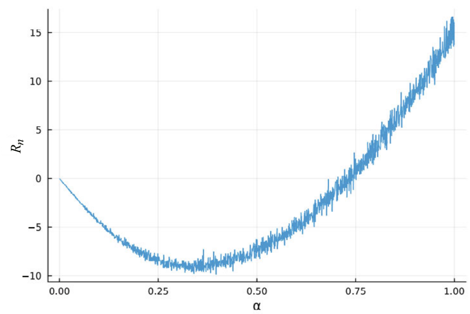

Figure 2 Simulated R_n as a function of α for $\beta, \gamma \approx 1/2$, as seen after 5000 transmissions on a 50 node complete graph

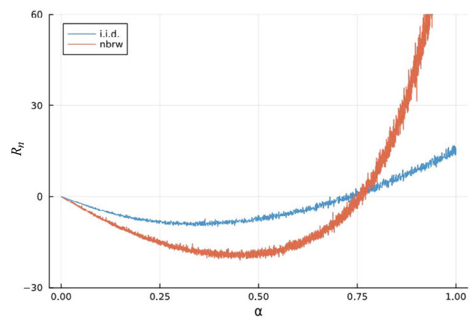

Figure 3 Simulated R_n as a function of α for i.i.d. and sequential communication, with $\beta, \gamma \approx 1/2$, as seen after 5000 transmissions on a 50 node complete graph

Figure 1 also suggests that $\alpha^*(\beta, \gamma)$ is a quasi-concave function taking on values close to 1 when β and γ tend to 0, suggesting that low confidence in the values x^i, amounting to choosing α close to 1, can be compensated by increased friction in updating the accounts y^i.

To provide another perspective to our simulation results, we may wish to fix $\beta = \gamma = 1/2$ as in [22] and follow the evolution of the error as a function of α. Processing the already available

R_n-s is carried out by filtering all instances for which $(\beta, \gamma) \in [0.49, 0.51]^2$, leading to tuples with general α and β, γ restricted to the vicinity of $1/2$. We thus forget β, γ, plot R_n against α for this collection resulting in the function that can be seen on Figure 2.

A final step, deserving attention, is to consider a communication protocol with sequential communication, according to which the receiver at any time becomes the sender at the next time instant, with backtracking prohibited. We plot R_n against α, with β, γ restricted to the vicinity of $1/2$, as in Figure 2. Overlapping the result with Figure 2 using a different scale, we get Figure 3, showing remarkable differences between the two methods: Sequential communication amplifies both stability and instability.

8 Discussion

We have considered the problem of reaching consensus over a communication network, represented by a possibly directed graph. The technical focal point of the paper was the extension of the results of Picci and Taylor[9], allowing the selection of a strictly stationary, ergodic sequence of communicating pairs, resulting in a fairly general theorem on rank-1 limits of products of random matrices, Theorem 5.1. The current investigations were motivated by our interest in extending the applicability of linear gossip algorithms, designed for directed networks, selecting i.i.d. communicating pairs, by Cai and Ishii[21] and Silvestre, et al.[22] to more general scenarios covering sequential selection of communicating pairs. The appeal of the cited papers is that they establish L_2-convergence with exponential rate, unknown for widely studied non-linear methods. The viability and limitations of the algorithm of [22] have been tested by systematic simulation based on 5M randomly chosen parameter setting, displaying a remarkable dichotomy with respect to empirical stability. We also provide simulation results for an algorithm with sequential communication protocol, outperforming the method of [22] for the specific parameter setting that is central in the cited paper.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

The authors express their gratitude to the anonymous reviewers, and to their colleague and long-time collaborator, György Michaletzky, for their useful comments on the paper.

References

[1] Caines P E, *Linear Stochastic Systems*, SIAM, Philadelphia, PA, 2018.

[2] Rissanen J and Caines P E, The strong consistency of maximum likelihood estimators for ARMA processes, *The Annals of Statistics*, 1979, **7**(2): 297–315.

[3] Dimakis A G, Sarwate A D, and Wainwright M J, Geographic gossip: Efficient aggregation for sensor networks, *Proceedings of the* 5*th International Conference on Information Processing in Sensor Networks*, 2006, 69–76.

[4] Hendrickx J M, Gerencsér B, and Fidan B, Trajectory convergence from coordinate-wise decrease of quadratic energy functions, and applications to platoons, *IEEE Control Systems Letters*, 2019, **4**(1): 151–156.

[5] Hadjicostis C N and Charalambous T, Asynchronous coordination of distributed energy resources for the provisioning of ancillary services, 2011 49*th Annual Allerton Conference on Communication, Control, and Computing* (*Allerton*)*, IEEE*, 2011, 1500–1507.

[6] Blondel V D, Hendrickx J M, and Tsitsiklis J N, Continuous-time average-preserving opinion dynamics with opinion-dependent communications, *SIAM Journal on Control and Optimization*, 2010, **48**(8): 5214–5240.

[7] Tsitsiklis J N, Problems in decentralized decision making and computation, PhD Thesis, Massachusetts Institute of Technology, Cambridge, 1984.

[8] Boyd S, Ghosh A, Prabhakar B, et al., Randomized gossip algorithms, *IEEE Transactions on Information Theory*, 2006, **52**(6): 2508–2530.

[9] Picci G and Taylor T J, Almost sure exponential convergence to consensus of random Gossip algorithms, *International Journal of Robust and Nonlinear Control*, 2013, **23**(9): 1033–1045.

[10] Kempe D, Dobra A, and Gehrke J, Gossip-based computation of aggregate information, *Proceedings of* 44*th Annual IEEE Symposium on Foundations of Computer Science*, 2003, 482–491.

[11] Bénézit F, Blondel V, Thiran P, et al., Weighted Gossip: Distributed averaging using non-doubly stochastic matrices, *Proceedings of* 2010 *IEEE International Symposium on Information Theory* (*ISIT*), 2010, 1753–1757.

[12] Gerencsér B and Gerencsér L, Tight bounds on the convergence rate of generalized ratio consensus algorithms, *IEEE Transactions on Automatic Control*, 2021, **67**(4): 1669–1684.

[13] Gerencsér B and Hendrickx J M, Push sum with transmission failures, *IEEE Transactions on Automatic Control*, 2018, **64**(3): 1019–1033.

[14] Tahbaz-Salehi A and Jadbabaie A, A necessary and sufficient condition for consensus over random networks, *IEEE Transactions on Automatic Control*, 2008, **53**(3): 791–795.

[15] Tahbaz-Salehi A and Jadbabaie A, Consensus over ergodic stationary graph processes, *IEEE Transactions on Automatic Control*, 2010, **55**(1): 225–230.

[16] Tsitsiklis J N and Blondel V D, 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*, 1997, **10**(1): 31–40.

[17] Blondel V D and Tsitsiklis J N, A survey of computational complexity results in systems and control, *Automatica*, 2000, **36**(9): 1249–1274.

[18] Iutzeler F, Ciblat P, and Hachem W, Analysis of sum-weight-like algorithms for averaging in wireless sensor networks, *IEEE Transactions on Signal Processing*, 2013, **61**(11): 2802–2814.

[19] Gerencsér B, Computable convergence rate bound for ratio consensus algorithms, *IEEE Control Systems Letters*, 2022, **6**: 3307–3312.

[20] Gerencsér B and Kornyik M, Low complexity convergence rate bounds for the synchronous Gossip

subclass of push-sum algorithms, *IEEE Control Systems Letters*, 2024, **8**: 1283–1288.

[21] Cai K and Ishii H, Average consensus on general strongly connected digraphs, *Automatica*, 2012, **48**(11): 2750–2761.

[22] Silvestre D, Hespanha J P, and Silvestre C, Broadcast and Gossip stochastic average consensus algorithms in directed topologies, *IEEE Transactions on Control of Network Systems*, 2018, **6**(2): 474–486.

[23] Oseledec V I, A multiplicative ergodic theorem: Lyapunov characteristic number for dynamical systems, *Transactions of the Moscow Mathematical Society*, 1968, **19**: 197–231.

[24] Seneta E, *Non-Negative Matrices and Markov Chains*, Springer Science & Business Media, New York, 2006.

[25] Raghunathan M S, A proof of Oseledec's multiplicative ergodic theorem, *Israel Journal of Mathematics*, 1979, **32**(4): 356–362.

[26] Furstenberg H and Kesten H, Products of random matrices, *The Annals of Mathematical Statistics*, 1960, **31**(2): 457–469.

[27] Elsner L, Koltracht I, and Neumann M, Convergence of sequential and asynchronous nonlinear paracontractions, *Numerische Mathematik*, 1992, **62**(1): 305–319.