Push-sum Algorithm on Time-varying Random Graphs

by

Pouya Rezaeinia

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements for
the degree of Master of Applied Science

Queen's University
Kingston, Ontario, Canada
August 2017

Abstract

In this thesis, we study the problem of achieving average consensus over a random time-varying sequence of directed graphs by extending the class of so-called push-sum algorithms to such random scenarios. Provided that an ergodicity notion, which we term the directed infinite flow property, holds and the auxiliary states of nodes are uniformly bounded away from zero infinitely often, we prove the almost sure convergence of the evolutions of this class of algorithms to the average of initial states. Moreover, for a random sequence of graphs generated using a time-varying B-irreducible sequence of probability matrices, we establish convergence rates for the proposed push-sum algorithm.

Acknowledgments

I would first like to thank my thesis advisors, Professors Tamás Linder and Bahman Gharessifard at Queen's University for admitting me to their group. I appreciate their knowledge, financial and educational support, encouragement and assistance at every stage of this thesis.

In addition to my advisors, I would like to thank Professor Behrouz Touri at the University of California San Diego for his contribution.

I would also like to express my gratitude to my family for their continuous support, and being my role models and a source of inspiration.

Pouya Rezaeinia

Contents

Chapter 1

Introduction

Many distributed algorithms, executed with limited information over a network of nodes, rely on estimating the average value of the initial state of the individual nodes. These include the distributed optimization protocols [22, 16, 26, 14, 23, 15, 11, 24, 25, 7, 13], distributed regret minimization algorithms in machine learning [1], and dynamics for fusion of information in sensor networks [21]. There is a large body of work devoted to the average consensus problem, starting with the pioneering work [12], where the so-called *push-sum algorithm* is first introduced. The key differentiating factor of the push-sum algorithm from consensus dynamics is that it takes advantage of a paralleled scalar-valued agreement dynamics, initiated uniformly across the nodes, that tracks the imbalances of the network and adjusts for them when estimating the consensus value.

In addition to the earlier work [12], several recent papers have studied the problem of average consensus, see for example [5], where other classes of algorithms based on weight adaptation are considered, ensuring convergence to the average on fixed directed graphs. The study of convergence properties of push-sum algorithms on time-varying deterministic sequences of directed graphs, to the best of our knowledge, was initiated in [3] and extended in [13], where push-sum protocols are intricately utilized to prove the convergence of a class of distributed optimization protocols on a sequence of time-varying directed graphs. The key assumption in [13] is the B-connectedness of the sequence, which means that in any window of size B the union of the underlying directed graphs over time is strongly connected. As we demonstrate, a by product of our work in deterministic settings is the generalization of the sequences on which the convergence of the push-sum algorithms is valid to the ones which satisfy the infinite flow property; in this sense, this extension mimics the properties required for the convergence of consensus dynamics, along the lines of [18].

1.1 Contribution

This thesis is concerned with the problem of average consensus for scenarios where communication between nodes is time-varying and possibly random. The convergence properties of consensus dynamics on random sequences of directed graphs are by this time well-established, see for example [18, 19, 20]. Average consensus on random graphs has also been studied in [3], under the assumption that the corresponding random sequence of stochastic matrices is *stationary* and ergodic with positive diagonals and irreducible expectation. One of our main objectives in this work is to extend these results to more general sequences of random stochastic matrices, in particular, beyond stationary. More importantly, to the best of our knowledge, we establish for the first time convergence rates for the push-sum algorithm on random sequences of directed graphs.

1.2 Organization

The remainder of this thesis is organized as follows. Chapter 2 contains mathematical preliminaries. In Chapter 3, we give a formal description of our consensus problem and then we describe the push-sum algorithm. Chapter 4 studies the convergence of products of matrices, and Chapter 5 contains our main convergence results. In Chapter 6, we derive convergence rates for the push-sum algorithm for a class of random column-stochastic matrices. Finally, we gather our conclusions and ideas for future directions in Chapter 7.

Chapter 2

Mathematical Preliminaries

We start with introducing some notational conventions.

2.1 Basic Notions

Let $\mathbb{R}$ and $\mathbb{Z}$ denote the set of real and integer numbers, respectively, and let $\mathbb{R}_{\geq 0}$ and $\mathbb{Z}_{\geq 0}$ denote the set of non-negative real numbers and integers, respectively. For a set $\mathbb{A}$, we write $S \subset \mathbb{A}$ if S is a proper subset of $\mathbb{A}$, and we call the empty set and $\mathbb{A}$ trivial subsets of $\mathbb{A}$. The complement of S is denoted by $\bar{S}$. Let $|S|$ denote the cardinality of a finite set S. We view all vectors in $\mathbb{R}^n$ as column vectors, where $n \in \mathbb{Z}_{\geq 0} - \{0\}$. We denote by $\|\cdot\|$, $\|\cdot\|_1$ and $\|\cdot\|_\infty$, the standard Euclidean norm, the 1-norm, and the infinity norm on $\mathbb{R}^n$, respectively. The ith unit vector in $\mathbb{R}^n$, whose ith component is 1 and all other components are 0, is denoted by e_i. The notation A' and v' will refer to the transpose of the matrix A and the vector v, respectively. We will use the short-hand notation $\mathbf{1}_n = (1, \ldots, 1)'$ and $\mathbf{0}_n = (0, \ldots, 0)' \in \mathbb{R}^n$. A vector v is stochastic if its elements are nonnegative real numbers that sum to 1. We use $\mathbb{R}^{n\times n}_{\geq 0}$ to denote the set of $n \times n$ non-negative real-valued matrices. A matrix $A \in \mathbb{R}^{n\times n}_{\geq 0}$ is row-stochastic (column-stochastic) if each of its rows (columns) sums to 1. For a given $A \in \mathbb{R}_{\geq 0}^{n\times n}$ and any nontrivial $S \subset [n] = \{1, \ldots, n\}$, we let $A_{S\bar{S}} = \sum_{i\in S, j\in\bar{S}} A_{ij}$. A positive matrix is a real matrix all of whose elements are positive. Finally, A_i denotes the ith row of matrix A and A^j denotes the jth column of A.

2.2 Graph theory

A (weighted) *directed graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ consists of a node set $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, an edge set $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}$, and a weighted *adjacency matrix* $A \in \mathbb{R}_{\geq 0}^{n\times n}$, with $A_{ji} > 0$ if and only if $(v_i, v_j) \in \mathcal{E}$, in which case we say that v_i is connected to v_j. Similarly, given a matrix $A \in \mathbb{R}_{\geq 0}^{n\times n}$, one can associate to A a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $(v_i, v_j) \in \mathcal{E}$ if and only if $A_{ji} > 0$, and hence A is the corresponding adjacency matrix for $\mathcal{G}$. The in-neighbors and the out-neighbors of v_i are the set of nodes $\mathcal{N}_i^{\text{in}} = \{j \in [n] : A_{ij} > 0\}$ and $\mathcal{N}_i^{\text{out}} = \{j \in [n] : A_{ji} > 0\}$, respectively. The out-degree of v_i is $d_i^{\text{out}} = |\mathcal{N}_i^{\text{out}}|$. A path is a sequence of nodes connected by edges. A directed graph is *strongly connected* if there is a path between any pair of nodes. A directed graph is *complete* if every pair of distinct vertices is connected by an edge. If the directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$ is strongly connected, we say that A is irreducible.

2.3 Sequences of random stochastic matrices

Let $\mathcal{S}_n^+$ be the set of $n \times n$ column-stochastic matrices that have positive diagonal entries, and let $\mathcal{F}_{\mathcal{S}_n^+}$ denote the Borel σ-algebra on $\mathcal{S}_n^+$. Given a probability space $(\Omega, \mathcal{B}, \mu)$, a measurable function $W : (\Omega, \mathcal{B}, \mu) \to (\mathcal{S}_n^+, \mathcal{F}_{\mathcal{S}_n^+})$ is called a random column-stochastic matrix, and a sequence $\{W(t)\}$ of such measurable functions on $(\Omega, \mathcal{B}, \mu)$ is called a random column-stochastic matrix sequence; throughout, we assume that $t \in \mathbb{Z}_{\geq 0}$. Note that for any $\omega \in \Omega$, one can associate a sequence of directed graphs $\{\mathcal{G}(t)(\omega)\}$ to $\{W(t)(\omega)\}$, where $(v_i, v_j) \in \mathcal{E}(t)(\omega)$ if and only if $W_{ji}(t)(\omega) > 0$. This in turn defines a sequence of random directed graphs on $\mathcal{V} = \{v_1, \ldots, v_n\}$, which we denote by $\{\mathcal{G}(t)\}$.

Chapter 3

Problem Statement

In this chapter we describe the consensus problem.

3.1 Average Consensus

Consider a network of nodes $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, where node $v_i \in \mathcal{V}$ has an initial state (or opinion) $x_i(0) \in \mathbb{R}$; the assumption that this initial state is a scalar is without loss of generality, and our treatment can easily be extended to the vector case. The objective of each node is to achieve *average consensus*; that is to compute the average $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i(0)$ with the constraint that only limited exchange of information between nodes is permitted. The communication layer between nodes at each time $t \geq 0$ is specified by a sequence of random directed graphs $\{\mathcal{G}(t)\}$, where $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t), W(t))$. Specifically, at each time t, node v_i updates its value based on the values of its in-neighbors $v_j \in \mathcal{N}_i^{\text{in}}(t)$, where $\mathcal{N}_i^{\text{in}}(t) = \{v_j \in \mathcal{V} : W_{ij}(t) > 0\}$. One standing assumption throughout this thesis is that each node knows its out-degree at every time t; this assumption is indeed necessary, as shown in [10]. Our main objective is to show that the class of so-called push-sum algorithms can be used to achieve average consensus at every node, under the assumption that the communication network is random. This key point distinguishes our work from the existing results in the literature [12], [13], [5]. Another key objective that we pursue in this thesis is to obtain rates of convergence for such algorithms. We start our treatment with reviewing the push-sum algorithm.

3.2 Random Push-Sum

Consider a network of nodes $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, where node $v_i \in \mathcal{V}$ has an initial state (or opinion) $x_i(0) \in \mathbb{R}$. The push-sum algorithm, proposed originally in [12], is defined as follows. Each node v_i maintains and updates, at each time $t \geq 0$, two state variables $x_i(t)$ and $y_i(t)$. The first state variable is initialized to $x_i(0)$ and the second one is initialized to $y_i(0) = 1$, for all $i \in [n]$. At time $t \geq 0$, node v_i sends $\frac{x_i(t)}{d_i^{\text{out}}(t)}$ and $\frac{y_i(t)}{d_i^{\text{out}}(t)}$ to its out-neighbors in the random directed graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t), W(t))$, which we assume to contain self-loops at each node for all $t \geq 0$. At time $(t+1)$, node v_i updates its state variables according to

$$
\begin{aligned}
x_i(t+1) &= \sum_{j \in \mathcal{N}_i^{\text{in}}(t)} \frac{x_j(t)}{d_j^{\text{out}}(t)}, \\
y_i(t+1) &= \sum_{j \in \mathcal{N}_i^{\text{in}}(t)} \frac{y_j(t)}{d_j^{\text{out}}(t)}.
\end{aligned}
\tag{3.1}
$$

It is useful to define another auxiliary variable $z_i(t+1) = \frac{x_i(t+1)}{y_i(t+1)}$; as we will show later, $z_i(t+1)$ is the estimate by node v_i of the average $\bar{x}$. One can rewrite this algorithm in a vector form; let the column-stochastic matrix $W(t)$ to be a function of $\mathcal{E}(t)$ with entries

$$W_{ij}(t) = \begin{cases} \frac{1}{d_j^{\text{out}}(t)} & \text{if } j \in \mathcal{N}_i^{\text{in}}(t), \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Using these weighted adjacency matrices, for every $t \geq 0$, we can rewrite the dynamics (3.1) as

$$\begin{aligned} x(t+1) &= W(t)x(t), \\ y(t+1) &= W(t)y(t), \end{aligned} \tag{3.3}$$

where

$$\begin{aligned} x(t) &= (x_1(t), \ldots, x_n(t))', \\ y(t) &= (y_1(t), \ldots, y_n(t))'. \end{aligned}$$

Note that in the push-sum algorithm the weights are not necessarily defined as in (3.1) and (3.2). The results in this thesis can easily be extended to the case where the $W(t)$ are column-stochastic and there exists a scalar $\gamma > 0$ such that $W_{ij}(t) \geq \gamma$ whenever $W_{ij}(t) > 0$. Metropolis weights [27], for instance, satisfy the mentioned conditions.

3.3 Implications to Distributed Optimization

As we mentioned in Chapter 1, estimating the average of the (initial) value of the individual nodes is an essential part of many distributed algorithms. Distributed optimization, for instance, has recently received a lot of interest. Here the problem is to distributively minimizing the sum of the cost functions of a network of nodes with the constraint that each node has only access to its own cost function. In the following we describe the subgradient-push algorithm, a broadcast-based algorithm that was introduced in [13], to steer every node to an optimal value.

Consider a set of nodes $\mathcal{V} = \{v_1, v_2, ..., v_n\}$ such that each node has a convex cost function $f_i : \mathbb{R}^d \to \mathbb{R}$ for some $d \in \mathbb{Z}_{\geq 0} - \{0\}$, where $i \in [n]$. The goal is to minimize the sum of the cost functions f_i:

$$\min \sum_{i \in [n]} f_i(x)$$

with the constraint that node v_i has only access to f_i and assuming that the set of optimal solutions is not empty. Every node v_i maintains vector variables $\mathbf{x}_i(t) \in \mathbb{R}^d$ and $\mathbf{w}_i(t) \in \mathbb{R}^d$, and a scalar variable $y_i(t)$, initiated arbitrarily at $\mathbf{x}_i(0) \in \mathbb{R}^d$, and with $y_i(0) = 1$. At each time t, node v_i sends $\frac{\mathbf{x}_i(t)}{d_i^{\text{out}}}$ and $\frac{y_i(t)}{d_i^{\text{out}}}$ to its out-neighbors in some directed graph $\mathcal{G}(t)$. Node v_i at time $(t+1)$ updates its variables according to

$$\begin{aligned}
\mathbf{w}_i(t+1) &= \sum_{j \in \mathcal{N}_i^{\text{in}}(t)} \frac{\mathbf{x}_j(t)}{d_j^{\text{out}}(t)}, \\
y_i(t+1) &= \sum_{j \in \mathcal{N}_i^{\text{in}}(t)} \frac{y_j(t)}{d_j^{\text{out}}(t)}, \\
\mathbf{z}_i(t+1) &= \frac{\mathbf{w}_i(t+1)}{y_i(t+1)}, \\
\mathbf{x}_j(t+1) &= \mathbf{w}_i(t+1) - \theta(t+1)\mathbf{g}_i(t+1),
\end{aligned} \tag{3.4}$$

where $\mathbf{w}_i(t) \in \mathbb{R}^d$ is an auxiliary variable used for computations, $\mathbf{g}_i(t+1)$ is the subgradient of the function $f_i(\mathbf{z})$ at $\mathbf{z} = \mathbf{z}_i(t+1)$ and $\theta(t+1)$ is the step-size. For all $i \in [n]$ the subgradients $\mathbf{g}_i$ are assumed to be uniformly bounded, i.e., there exists $L_i < \infty$ such that for all $\mathbf{z} \in \mathbb{R}^d$, $\|\mathbf{g}_i\| \leq L_i$ for all subgradients $\mathbf{g}_i$ of f_i at $\mathbf{z}$. Moreover, the step-sizes $\theta(t)$ satisfy

$$\begin{aligned}\sum_{t=1}^{\infty} \theta(t) =& \infty,\\ \sum_{t=1}^{\infty} \theta^2(t) <& \infty,\end{aligned}$$

and $\theta(t+1) \leq \theta(t)$ for all $t \geq 1$.

As mentioned in [13], without the subgradient term in (3.4), the subgradient-push protocol would be equivalent to (3.3). The averaging term is to ensure that every node receives an equal weighting after all the linear combinations and ratios have been taken. The subgradient terms are to steer the consensus point towards the optimal set, while the push-sum updates steer the vectors $\mathbf{z}_i(t+1)$ towards each other.

In order to solve this optimization problem, the subgradient term is considered to be a perturbation to the push-sum algorithm (3.1) as follows:

$$\begin{aligned}\mathbf{w}_i(t+1) &= \sum_{j \in \mathcal{N}_i^{\text{in}}(t)} \frac{\mathbf{x}_j(t)}{d_j^{\text{out}}(t)},\\ y_i(t+1) &= \sum_{j \in \mathcal{N}_i^{\text{in}}(t)} \frac{y_j(t)}{d_j^{\text{out}}(t)},\\ \mathbf{z}_i(t+1) &= \frac{\mathbf{w}_i(t+1)}{y_i(t+1)},\\ \mathbf{x}_j(t+1) &= \mathbf{w}_i(t+1) + \epsilon_i(t+1),\end{aligned} \tag{3.5}$$

where $\epsilon_i(t+1) = \theta(t+1)\mathbf{g}_i(t+1)$ is the perturbation at time $(t+1)$. In the perturbed push-sum algorithm, for a sequence $\mathcal{G}(t)$ that is B-strongly-connected, the l'th components of the sequences $\mathbf{z}_i(t+1)$ track the average of the l'th components of the $\mathbf{x}_i(t)$'s increasingly well as t increases and the nodes reach a consensus which is proven to be an optimal point of the aforementioned optimization problem. A sequence of deterministic graphs $\{\mathcal{G}(t)\}$ is called B-strongly-connected if for some integer $B > 0$ and all $t \geq 0$, $\cup_{t'=tB}^{(t+1)B-1}\mathcal{G}(t')$ is strongly connected.

In [13], the communication layer is deterministic. In large networks the communication links between nodes occur at random, which raises the question, whether the subgradient-push algorithm can find an optimal value of the optimization problem when the communication layer occurs at random. An special case of this problem is when $f_i(t) = 0$ for all $i \in [n]$ and $t \in \mathbb{R}$, which boils down to our problem, "Push-sum Algorithm on Time-varying Random Graphs."

Chapter 4

Ergodicity

In this chapter, we establish some important auxiliary results regarding the convergence of products (ergodicity) of matrices which satisfy the so-called directed infinite flow property (c.f. Definition 4.1.4). We study the products of a class of matrices in a deterministic setting, which we then use to study the push-sum algorithm in the next chapter.

We start by some definitions.

4.1 Ergodicity of Row-stochastic Matrices

We start with the definitions of consensus and ergodicity of a sequence of matrices that are based on the asymptotic behavior of the products of these matrices.

Definition 4.1.1 (Consensus [18])**.** *Let $\{A(t)\}$ be a sequence of row-stochastic matrices, and for $t \geq s \geq 0$, let $A(t:s)$ denote the product*

$$A(t:s) = A(t)A(t-1)\cdots A(s), \tag{4.1}$$

where $A(s:s) = A(s)$. The sequence $\{A(t)\}$ is said to admit consensus if

$\lim_{k\to\infty} A(t:0) = \mathbf{1}_n v'$ *for some stochastic vector* $v \in \mathbb{R}^n$.

Definition 4.1.2 (Ergodicity [4], [18])**.** *Let* $\{A(t)\}$ *be a sequence of row-stochastic matrices. The sequence* $\{A(t)\}$ *is said to be weakly ergodic, if for all* $i, j, l \in [n]$ *and any* $s \geq 0$, $\lim_{t\to\infty} (A_{il}(t:s) - A_{jl}(t:s)) = 0$. *The sequence is said to be strongly ergodic if* $\lim_{t\to\infty} A(t:s) = \mathbf{1}_n v'(s)$ *for any* $s \geq 0$, *where* $v(s) \in \mathbb{R}^n$ *is a stochastic vector.*

It can be shown that weak ergodicity and strong ergodicity are equivalent [4, Theorem 1]. We will simply call such a sequence of row-stochastic matrices ergodic. It is worth mentioning that ergodicity of a sequence of matrices has a close relationship with the theory of ergodic Markov chains [18, Section 5.2].

Note that a sequence of matrices that is ergodic admits consensus; however, a sequence of matrices that admits consensus is not necessarily ergodic.

We first establish a sufficient condition for ergodicity of a sequence of row-stochastic matrices, Proposition 4.1.6, which we subsequently use in our convergence result for the push-sum algorithm. For this reason, we consider the following dynamical system:

$$x(t+1) = A(t)x(t), \qquad \text{for all } t \geq 0. \tag{4.2}$$

Let us start by two key definitions.

Definition 4.1.3 (Strong Aperiodicity [18])**.** *We say that a sequence of matrices* $\{A(t)\}$ *is strongly aperiodic if there exists* $\gamma > 0$ *such that* $A_{ii}(t) \geq \gamma$, *for all* $t \geq 0$ *and* $i \in [n]$.

Motivated by the *infinite flow property* [18, Definition 3.2.], we provide the following definition.

Definition 4.1.4 (Directed Infinite Flow Property). *We say that a sequence of matrices $\{A(t)\}$ has the directed infinite flow property if for any non-trivial $S \subset [n]$, $\sum_{t=0}^{\infty} A_{S\bar{S}}(t) = \infty$.*

Consider now a sequence of matrices $\{A(t)\}$ that is strongly aperiodic and has the directed infinite flow property. Let $k_0 = 0$, and for any $q \geq 1$, define

$$k_q = \text{argmin}_{t' > k_{q-1}} \left(\min_{S \subset [n]} \sum_{t=k_{q-1}}^{t'-1} A_{S\bar{S}}(t) > 0 \right). \tag{4.3}$$

Note that k_q is the minimal time instance after k_{q-1}, such that there is nonzero information flow between any non-trivial subset of $\mathcal{V}$ and its complement; consequently, the directed graph associated with the product $A(k_q - 1)A(k_q - 2) \cdots A(k_{q-1})$ is strongly connected; otherwise, one can find a non-trivial $S \in [n]$ such that $\sum_{t=k_{q-1}}^{t'-1} A_{S\bar{S}}(t) = 0$, which contradicts the definition of k_q.

Proposition 4.1.5. *If a sequence of matrices $\{A(t)\}$ has the directed infinite flow property, k_q is finite for all $q \geq 0$.*

Proof. Suppose that k_q is not finite for some $q \geq 0$. Then, using (4.3), there exists a non-trivial subset $S \subset [n]$ such that $\sum_{t=k_{q-1}}^{\infty} A_{S\bar{S}}(t) = 0$. This implies that $\sum_{t=0}^{\infty} A_{S\bar{S}}(t) < \infty$, which contradicts the assumption that $\{A(t)\}$ has the directed infinite flow property. □

To establish convergence results for the products of row-stochastic matrices satisfying Definition 4.1.4, we argue that in each time window where the underlying directed graph becomes strongly connected for n times, i.e., after $k_{qn} - k_{(q-1)n}$ time steps for some q, *significant mixing* will occur. To formalize this statement, let $\ell_0 = 0$

and

$$\ell_q = k_{qn} - k_{(q-1)n}, \tag{4.4}$$

for $q \geq 1$. For $t > s \geq 0$, we also define

$$\mathbb{Q}_{t,s} = \{q : s \leq k_{(q-1)n}, k_{qn} \leq t\}.$$

We are now ready to state our first result.

Proposition 4.1.6. *Consider the dynamics* (4.2)*, where the sequence of row-stochastic matrices* $\{A(t)\}$ *is such that* $A'(t)$ *satisfies* (3.2)*. Suppose, additionally, that* $\{A(t)\}$ *is strongly aperiodic and has the directed infinite flow property. Then,*

(i) there is a vector $\phi(s) \in \mathbb{R}^n$ *such that, for all* $i, j \in [n]$ *and* $t \geq s$*,*

$$\left| [A(t:s)]_{ij} - \phi_j(s) \right| \leq \Lambda_{t,s},$$

where $\Lambda_{t,s} = \prod_{q \in \mathbb{Q}_{t,s}} \lambda_q$ *and* $\lambda_q = \left(1 - \frac{1}{n^{\ell_q}}\right) \in (0,1)$*;*

(ii) if, for the sequence $\{\ell_q\}$ *associated with* $\{A(t)\}$*, we have*

$$\sum_{q=1}^{\infty} \frac{1}{n^{\ell_q}} = \infty, \tag{4.5}$$

then the sequence $\{A(t)\}$ *is ergodic.*

Proof. We start by proving the first statement. By definition of k_q, we know that for all $q \geq 0$, $A(k_{q+1} - 1 : k_q)$ is irreducible. Since each $A(t)$ is strongly aperiodic, by

Lemma A.2, the matrix

$$\begin{aligned} &A(k_{n(q+1)}-1:k_{nq}) \\ &\quad = A(k_{n(q+1)}-1:k_{n(q+1)-1}) \times \cdots \times A(k_{nq+2}-1:k_{nq+1}) \times A(k_{nq+1}-1:k_{nq}), \end{aligned}$$

which is the product of n irreducible matrices, is positive for all $q \geq 0$. Hence, by Lemma A.1 (ii) in the Appendix, for all $i, j \in [n]$, we have

$$[A(k_{n(q+1)}-1:k_{nq})]_{ij} \geq \frac{1}{n^{k_{n(q+1)}-k_{nq}}} = \frac{1}{n^{l_{q+1}}}.$$

Now, since $A(t:s) = A(t:s)I_n$ and for all $j \in [n]$, $\max_{i\in[n]}[I_n]_{ij} - \min_{i\in[n]}[I_n]_{ij} = 1$, using Lemma A.3, we obtain

$$\max_{i\in[n]}[A(t:s)]_{ij} - \min_{i\in[n]}[A(t:s)]_{ij} \leq \Lambda_{t,s}. \tag{4.6}$$

Note that if we let $\phi_j(s) = \min_{i\in[n]} A_{ij}(t:s)$ for all $j \in [n]$, we have

$$\Big|[A(t:s)]_{ij} - \phi_j(s)\Big| \leq \max_{i\in[n]}[A(t:s)]_{ij} - \min_{i\in[n]}[A(t:s)]_{ij}. \tag{4.7}$$

Using (4.6) and (4.7), we conclude that

$$\Big|[A(t:s)]_{ij} - \phi_j(s)\Big| \leq \Lambda_{t,s},$$

for all $i, j \in [n]$.

We next prove part (ii); since $\lambda_q \in (0,1)$ for all $q \geq 1$, we have that $\ln(\lambda_q) \leq \frac{-1}{n^{\ell_q}}$, where we have used the fact that $\ln(\zeta) \leq \zeta - 1$ for all $\zeta > 0$. This implies

$$\sum_{q=1}^{\infty} \ln(\lambda_q) \leq -\sum_{q=1}^{\infty} \frac{1}{n^{\ell_q}}. \tag{4.8}$$

On the other hand, we have

$$\lim_{t\to\infty} \Lambda_{t,0} = \lim_{t\to\infty} \prod_{q\in\mathbb{Q}_{t,0}} \lambda_q = \lim_{t\to\infty} \exp\left(\sum_{q\in\mathbb{Q}_{t,0}} \ln(\lambda_q)\right).$$

The definition of the sets $\mathbb{Q}_{t,s}$ implies that we can write the right hand side as $\exp\left(\sum_{q=1}^{\infty} \ln(\lambda_q)\right)$, which gives

$$\lim_{t\to\infty} \Lambda_{t,0} = \exp\left(\sum_{q=1}^{\infty} \ln(\lambda_q)\right) = 0,$$

where the last equality follows from (4.8) and the assumption $\sum_{q=0}^{\infty} \frac{1}{n^{\ell_q}} = \infty$. Using the fact that $\lim_{t\to\infty} \Lambda_{t,0} = 0$, we have that $\lim_{t\to\infty} \Lambda_{t,s} = 0$, for any $s > 0$. Hence, by Proposition 4.1.6, part (i), we conclude that $\{A(t)\}$ is weakly (and thus strongly) ergodic. □

4.2 Product of Column-stochastic Matrices

Following similar steps as in Proposition 4.1.6 we obtain the following result for sequences of column-stochastic matrices of the form (3.2).

Proposition 4.2.1. *Consider the dynamics* (4.2) *and assume that sequence of matrices* $\{A(t)\}$ *is strongly aperiodic and has the directed infinite flow property, where the* $A(t)$ *are weighted adjacency matrices in the form of* (3.2)*. Then,*

(i) *there is a vector* $\phi(t) \in \mathbb{R}^n$ *such that, for all* $i, j \in [n]$ *and* $t \geq s$,

$$\Big|[A(t:s)]_{ij} - \phi_i(t)\Big| \leq \Lambda_{t,s},$$

where $\Lambda_{t,s} = \prod_{q \in \mathbb{Q}_{t,s}} \lambda_q$ *and* $\lambda_q = \left(1 - \frac{1}{n^{\ell_q}}\right)$;

(ii) *for the sequence* $\{\ell_q\}$ *associated with* $\{A(t)\}$, *if*

$$\sum_{q=1}^{\infty} \frac{1}{n^{\ell_q}} = \infty,$$

then for all $j \in [n]$, $\lim_{t\to\infty}\Big|[A(t:s)]_{ij} - \phi_i(t)\Big| = 0$.

It is worth pointing out that in Proposition 4.1.6, since the $A(t)$ are row-stochastic, $x(t)$ approaches a vector with identical entries. However, in Proposition 4.2.1 the $x(t)$ does not necessarily approach a fixed vector.

Chapter 5

Convergence of Push-Sum

With all the pieces in place, we are now ready to study the behavior of the push-sum algorithm in a random setting.

5.1 Convergence of the Push-sum Algorithm

In the following theorem we investigate the behavior of the push-sum algorithm when the communication layer is time-varying with random changes.

Theorem 5.1.1. *Consider the push-sum algorithm* (3.3) *and suppose that the sequence of random column-stochastic matrices* $\{W(t)\}$ *has the directed infinite flow property, almost surely. Then, we have*

$$|z_i(t+1) - \bar{x}| \leq \frac{2\|x(0)\|_1}{y_i(t+1)} \Lambda_{t,0},$$

almost surely, where $\Lambda_{t,0} = \prod_{q \in \mathbb{Q}_{t,0}} \lambda_q$ *and* $\lambda_q = \left(1 - \frac{1}{n^{\ell_q}}\right) \in (0,1)$.

Proof. Define

$$D(t:s) \triangleq W(t:s) - \phi(t)\mathbf{1}_n',$$

where $\phi(t)$ is a (random) vector from part (i) of Proposition 4.2.1. In addition, under the push-sum algorithm we have that

$$\begin{aligned} x(t+1) &= W(t:0)x(0), \\ y(t+1) &= W(t:0)y(0), \end{aligned}$$

for all $t \geq 0$. Hence, for every $t \geq 0$ and all $i \in [n]$, we have

$$\begin{aligned} z_i(t+1) - \bar{x} &= \frac{x_i(t+1)}{y_i(t+1)} - \frac{\mathbf{1}_n' x(0)}{n} \\ &= \frac{[W(t:0)x(0)]_i}{[W(t:0)y(0)]_i} - \frac{\mathbf{1}_n' x(0)}{n} \\ &= \frac{[D(t:0)x(0)]_i + \phi_i(t)\mathbf{1}_n' x(0)}{[D(t:0)y(0)]_i + \phi_i(t)\mathbf{1}_n' y(0)} - \frac{\mathbf{1}_n' x(0)}{n}. \end{aligned}$$

Using the fact that $y(0) = \mathbf{1}_n$ and by bringing the fractions to a common denominator, we have

$$\begin{aligned} z_i(t+1) - \bar{x} =& \frac{[D(t:0)x(0)]_i + \phi_i(t)\mathbf{1}_n' x(0)}{[D(t:0)\mathbf{1}_n]_i + n\phi_i(t)} - \frac{\mathbf{1}_n' x(0)}{n} \\ =& \frac{n[D(t:0)x(0)]_i + n\phi_i(t)\mathbf{1}_n' x(0)}{n([D(t:0)\mathbf{1}_n]_i + n\phi_i(t))} - \frac{[D(t:0)\mathbf{1}_n]_i \mathbf{1}_n' x(0) + n\phi_i(t)\mathbf{1}_n' x(0)}{n([D(t:0)\mathbf{1}_n]_i + n\phi_i(t))} \\ =& \frac{n[D(t:0)x(0)]_i - [D(t:0)\mathbf{1}_n]_i \mathbf{1}_n' x(0)}{n([D(t:0)\mathbf{1}_n]_i + n\phi_i(t))}. \end{aligned}$$

Note that the denominator in the last equation is equal to $ny_i(t+1)$. Hence, for all

$i \in [n]$ and $t \geq 1$ we have

$$\begin{aligned}|z_i(t+1) - \bar{x}| \leq & \frac{\|x(0)\|_1}{y_i(t+1)} \left(\max_j |[D(t:0)]_{ij}| \right) + \frac{|\mathbf{1}_n' x(0)|}{n y_i(t+1)} \left(\max_j |[D(t:0)]_{ij}| \right) n \\ = & \frac{|\mathbf{1}_n' x(0)| + \|x(0)\|_1}{y_i(t+1)} \left(\max_j |[D(t:0)]_{ij}| \right),\end{aligned}$$

where the inequality follows from the triangle inequality. Since $|\mathbf{1}_n' x(0)| \leq \|x(0)\|_1$, we have that

$$|z_i(t+1) - \bar{x}| \leq \frac{2\|x(0)\|_1}{y_i(t+1)} \left(\max_j |[D(t:0)]_{ij}| \right).$$

Using the upper bound in part (i) of Proposition 4.2.1, we obtain

$$|z_i(t+1) - \bar{x}| \leq \frac{2\|x(0)\|_1}{y_i(t+1)} \Lambda_{t,0}. \tag{5.1}$$

□

The upper bound for the deviation from the initial average at each time, derived in Theorem 5.1.1, not only depends on the 1-norm of the initial states vector $x(0)$, but also on $\Lambda_{t,0}$, which is an indicator of the sparsity of communication links. In addition, the bound also depends on the inverse of $y_i(t)$, which can decrease to zero if the outgoing information flow dramatically exceeds the incoming flow for a set of nodes.

Proposition 5.1.2. *Consider the push-sum algorithm* (3.3) *and suppose that the sequence of random column-stochastic matrices* $\{W(t)\}$ *has the directed infinite flow property, almost surely. Moreover, suppose that the sequence* $\{\ell_q\}$ *associated with* $\{W(t)\}$ *satisfies* (4.5)*, almost surely. If there exists* $\delta > 0$*, such that for any* $t \geq 0$*,* *there is* $t' \geq t$ *such that* $y_i(t') \geq \delta$ *for all* $i \in [n]$*, then*

$$\lim_{t\to\infty} |z_i(t+1) - \bar{x}| = 0, \quad \textit{almost surely.}$$

Remark 5.1.3. *In the next chapter we exhibit a class of random matrix sequences* $\{W(t)\}$ *that satisfy the conditions of Proposition 5.1.2 and thus admit average consensus almost surely.*

Proof. Proof of this proposition is similar to the proof of Theorem 4.1 in [3], where the sequence $\{W(t)\}$ is assumed to be stationary; however, since we do not assume stationarity, we provide a proof. By Proposition 4.2.1 part (ii), for any $\varepsilon > 0$ there is a time t_ε such that for all $t \geq t_\varepsilon$ and $i \in [n]$,

$$\sum_{j=1}^{n} \left| [W(t:0)]_{ij} - \frac{1}{n}\sum_{k=1}^{n}[W(t:0)]_{ik} \right| < \delta\varepsilon.$$

By assumption, there exists $t'_\varepsilon \geq t_\varepsilon$ such that $y(t'_\varepsilon) \geq \delta$, which implies that $f(t'_\varepsilon) < \varepsilon$, where $f(t)$ is defined as in Lemma A.5. Since by Lemma A.5, $f(t)$ is non-increasing, $f(t) < \varepsilon$ for all $t \geq t'_\varepsilon$, meaning that $f(t)$ converges to zero as $t \to \infty$ and hence, $\lim_{t\to\infty} |z_i(t+1) - \bar{x}| = 0$, almost surely. □

Chapter 6

B-Irreducible Sequences

In this chapter we characterize a class of random column-stochastic matrices that admits average consensus and we provide a rate of convergence of the push-sum algorithm for this class. To achieve this, we restrict the class of random matrices that we consider; as we will point out later, this restricted class still includes many interesting sequences of random matrices.

6.1 B-irreducible Sequences of Column-stochastic Matrices

In the following discussion, we assume that the push-sum dynamics is generated by a column-stochastic matrix sequence $\{W(t)\}$ where

$$W_{ij}(t) = \frac{R_{ij}(t)}{\sum_{i=1}^{n} R_{ij}(t)}, \tag{6.1}$$

for all $i, j \in [n]$, where $R_{ij}(t)$ is 1 with probability $P_{ij}(t)$, and is 0 with probability $1 - P_{ij}(t)$ such that $\{R_{ij}(t) : i, j \in [n], t \geq 0\}$ are independent random variables. In other words, there is a random communication link between node v_j and v_i at time t with probability $P_{ij}(t)$. Note that $\{W(t)\}$ is a sequence of independent random column-stochastic matrices.

Furthermore, for the probability matrix sequence $\{P(t)\}_{t\geq 0}$, we assume that the following holds.

Assumption 6.1.1. *$\{P(t)\}_{t\geq 0}$ is a sequence of $n\times n$ matrices with $P_{ij}(t)\in[0,1]$. Additionally, we assume that $P_{ii}(t)=1$, for all $v_i\in\mathcal{V}$. Also, for some constant $\epsilon>0$, we assume that $P_{ij}(t)\geq\epsilon$ for all $i,j\in[n]$ and all $t\geq 0$ such that $P_{ij}(t)\neq 0$. Finally, we assume that the sequence $\{P(t)\}_{t\geq 0}$ is B-irreducible, i.e. for some integer $B>0$,*

$$\sum_{t'=tB}^{(t+1)B-1} P(t)$$

is irreducible for all $t\geq 0$.

We next state the main result of this chapter.

6.2 Main Results

Theorem 6.2.1. *Consider the push-sum algorithm* (3.3) *and let $\{W(t)\}$ be a sequence of random column-stochastic matrices defined by* (6.1)*, where $\{P(t)\}$ satisfies Assumption 6.1.1. Let $p=\epsilon^{2(n-1)}$. Then, for any $t\geq B+\frac{2nB}{p}$, where $n\geq 2$*

$$\mathbb{E}\left[\ln\left(|z_i(t+1)-\bar{x}|\right)\right]\leq c_0-c_1t$$

where

$$\begin{aligned} c_0 &= \ln\left(2\|x(0)\|_1\right)+\ln(n)\left(\frac{nB}{p}+B\right)+\ln(15),\\ c_1 &= -\frac{p}{2nB}\ln\left(1-\frac{1}{n^{\frac{4nB}{p}}}\right). \end{aligned}$$

The proof relies on the following results.

Lemma 6.2.2. *Let $\{W(t)\}$ be a sequence of random column-stochastic matrices defined by* (6.1)*, where $\{P(t)\}$ satisfies Assumption 6.1.1. Let $\{k_q\}$ and $\{\ell_q\}$ be the sequences defined, respectively, in* (4.3) *and* (4.4) *along each sample path. Then*

(i) the sequence $\{W(t)\}$ has the directed infinite flow property almost surely, and

(ii) for the sequence $\{\ell_q\}$, we have

$$\sum_{q=0}^{\infty} \frac{1}{n^{\ell_q}} = \infty, \quad \textit{almost surely.}$$

Proof. We start by proving (i). For any $t \geq 0$, let us define the sequence of events

$$\mathcal{A}_t = \Big\{ \sum_{t'=tB}^{(t+1)B-1} W(t') \text{ is irreducible} \Big\}. \tag{6.2}$$

Note that for all $t \geq 0$, the events $\{\mathcal{A}_t\}_{t \geq 0}$ are independent and that $\mathcal{A}_t$ implies $\sum_{t'=tB}^{(t+1)B-1} W_{S\bar{S}}(t') > 0$, for any non-trivial $S \subset [n]$. Since $\min_{i,j \in [n]: P_{ij}(t) > 0} P_{ij}(t) > \epsilon > 0$, for all $t \geq 0$, we have

$$\mathbb{P}(\mathcal{A}_t) \geq \epsilon^{2(n-1)}.$$

This follows from Lemma A.4 and the fact that $\{P(t)\}$ is B-irreducible and hence, there is at least a subset of size $2(n-1)$ of the edges (v_j, v_i) that form a strongly connected graph and $P_{ij}(t') \geq \epsilon$ for some $t' \in [tB, (t+1)B - 1]$.

Since the events $\mathcal{A}_t$ are independent, hence, by the second Borel-Cantelli lemma [6, Theorem 2.3.6], $\sum_{t'=tB}^{(t+1)B-1} W_{S\bar{S}}(t') > 0$ infinitely often, almost surely. Moreover, since every positive entry of $W(t)$ is bounded below by $\frac{1}{n}$, for any non-trivial $S \subset [n]$, $\sum_{t=0}^{\infty} W_{S\bar{S}}(t) = \infty$, almost surely, implying that $\{W(t)\}$ has the directed infinite flow property, almost surely. This also implies that k_q and ℓ_q are finite for all q, almost surely. This completes the proof of (i).

To prove (ii), let us define, for all $t \geq 0$, the sequence of events

$$\mathcal{C}_t = \bigcap_{t'=tn}^{(t+1)n-1} \mathcal{A}_{t'}, \tag{6.3}$$

where $\mathcal{A}_t$ is defined in (6.2). Since the $\mathcal{A}_t$ are independent, $\mathbb{P}(\mathcal{C}_t) = \prod_{t'=tn}^{(t+1)n-1} \mathbb{P}(\mathcal{A}_{t'}) \geq \epsilon^{2n(n-1)}$ for all $t \geq 0$. This implies that $\sum_{t=0}^{\infty} \mathbb{P}(\mathcal{C}_t) = \infty$. Again, since the $\mathcal{C}_t$ are independent, by the Borel-Cantelli lemma, $\mathcal{C}_t$ occurs infinitely often, almost surely. This implies that $\ell_q \leq nB$ infinitely often, almost surely. Hence, $\sum_{q=1}^{\infty} \frac{1}{n^{\ell_q}} = \infty$, almost surely. □

Lemma 6.2.3. *In the push-sum algorithm* (3.3)*, let* $\{W(t)\}$ *be a sequence of random column-stochastic matrices corresponding to the sequence* $\{P(t)\}$ *satisfying Assumption 6.1.1. Then for all* $t \geq 0$ *there exists* $t' \geq t$ *such that for all* $i \in [n]$*,* $y_i(t') \geq \frac{1}{n^{nB}}$.

Proof. Consider the event $\mathcal{C}_t$ defined in (6.3). At any time $\mathcal{C}_t$ occurs, by Lemma A.2, the product $W(tnB + nB - 1 : tnB)$ is positive; moreover, by Lemma A.1, $W_{ij}(tnB + nB - 1 : tnB) \geq \frac{1}{n^{nB}}$ for all $i, j \in [n]$. Since $W(t)$ is column-stochastic, we have $W_{ij}(tnB + nB - 1 : 0) \geq \frac{1}{n^{nB}}$. By Lemma 6.2.2, $\mathcal{C}_t$ occurs infinitely often, almost surely; therefore, for all $t \geq 0$ there exists $t' \geq t$ such that for all $i \in [n]$, $y_i(t') \geq \frac{1}{n^{nB}}$. □

The preceding two lemmas and Proposition 5.1.2 imply the following.

Corollary 6.2.4. *Let $\{W(t)\}$ be a sequence of random column-stochastic matrices corresponding to the sequence $\{P(t)\}$ satisfying Assumption 6.1.1. Then $\{W(t)\}$ admits average consensus, almost surely.*

Lemma 6.2.5. *Let $\{W(t)\}$ be a sequence of random column-stochastic matrices corresponding to the sequence $\{P(t)\}$ satisfying Assumption 6.1.1. Let $\{\ell_q\}$ be the sequence defined in (4.4) along each sample path. For all $t \geq B + \frac{2nB}{p}$, we have*

$$\mathbb{E}\left[\Lambda_{t,0}\right] \leq \exp\left(-\beta_t^2\left(\frac{t}{B}-2\right)\right) + 2\left(1-\frac{1}{n^{\frac{4nB}{p}}}\right)^{\frac{pt}{2nB}},$$

where $\Lambda_{t,0} = \prod_{q\in\mathbb{Q}_{t,0}}(1-\frac{1}{n^{l_q}})$, $\beta_t = \frac{p}{2} - \frac{2pB}{t}$, and $p = \epsilon^{2(n-1)}$.

Proof. Let $X_B(t)$ be the indicator of the event $\mathcal{A}_t$, i.e.,

$$X_B(t) = \begin{cases} 1 & \text{if } \sum_{t'=tB}^{(t+1)B-1} W(t') \text{ is irreducible,} \\ 0 & \text{otherwise.} \end{cases}$$

By the preceding argument, we have $\mathbb{P}(X_B(t)=1) \geq p = \epsilon^{2(n-1)} > 0$. Note that the $X_B(t)$ are independent. We let $H_B(T) = \sum_{t=0}^{T} X_B(t)$ for all $T \geq 0$, and define

$$q_t \triangleq \max\{q : k_q \leq t\}.$$

By definition of $H_B(\cdot)$ and q_t, we have that

$$q_t \geq H_B\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right). \tag{6.4}$$

Now, we have that

$$\mathbb{E}\left[\Lambda_{t,0}\right] = \mathbb{E}\left[\Lambda_{t,0} \,\middle|\, q_t \leq \frac{pt}{2B}\right] \mathbb{P}\left(q_t \leq \frac{pt}{2B}\right) + \mathbb{E}\left[\Lambda_{t,0} \,\middle|\, q_t > \frac{pt}{2B}\right] \mathbb{P}\left(q_t > \frac{pt}{2B}\right).$$

Since all terms on the right-hand side are less than or equal to 1, we have

$$\mathbb{E}\left[\Lambda_{t,0}\right] \leq \mathbb{P}\left(q_t \leq \frac{pt}{2B}\right) + \mathbb{E}\left[\Lambda_{t,0} \,\middle|\, q_t > \frac{pt}{2B}\right].$$

Using (6.4), we have

$$\mathbb{E}\left[\Lambda_{t,0}\right] \leq \mathbb{P}\left(H_B\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) \leq \frac{pt}{2B}\right) + \mathbb{E}\left[\Lambda_{t,0} \,\middle|\, q_t > \frac{pt}{2B}\right].$$

Let us consider the second term on the right-hand side. When $q_t > \frac{pt}{2B}$, we have $|\mathbb{Q}_{t,0}| \geq \left\lfloor \frac{pt}{2nB} \right\rfloor$. Using Lemma A.7 to maximize the second term on the right-hand side over the choices of ℓ_q, we obtain

$$\begin{aligned}\mathbb{E}\left[\Lambda_{t,0} \,\middle|\, q_t > \frac{pt}{2B}\right] \leq & \left(1 - \frac{1}{n^{\frac{t}{\left\lfloor \frac{pt}{2nB} \right\rfloor}}}\right)^{\left\lfloor \frac{pt}{2nB} \right\rfloor} \\ \leq & 2\left(1 - \frac{1}{n^{\frac{t}{\left\lfloor \frac{pt}{2nB} \right\rfloor}}}\right)^{\frac{pt}{2nB}}. \end{aligned} \tag{6.5}$$

To further simplify the above inequality, we show that $\frac{t}{\left\lfloor \frac{pt}{2nB} \right\rfloor} \leq \frac{4nB}{p}$. To show this, we note that for all $t \geq \frac{2nB}{p} + B$, we have $\frac{pt}{2nB} > 1$ and hence, $\left\lfloor \frac{pt}{2nB} \right\rfloor \geq 1$. Now, assume that $\xi = \left\lfloor \frac{pt}{2nB} \right\rfloor \geq 1$. We have $2nB\xi \leq pt \leq 2nB(\xi + 1)$. Therefore,

$$\frac{t}{\left\lfloor \frac{pt}{2nB} \right\rfloor} \leq \frac{2nB}{p}\left(\frac{\xi + 1}{\xi}\right) \leq \frac{4nB}{p},$$

where the last inequality follows from the fact that $\xi \geq 1$.

Using this inequality in (6.5), we get

$$\begin{aligned}\mathbb{E}\left[\Lambda_{t,0} \;\middle|\; q_t > \frac{pt}{2B}\right] \leq & 2\left(1 - \frac{1}{n^{\frac{t}{\left\lfloor \frac{pt}{2nB} \right\rfloor}}}\right)^{\frac{pt}{2nB}} \\ \leq & 2\left(1 - \frac{1}{n^{\frac{4nB}{p}}}\right)^{\frac{pt}{2nB}}. \end{aligned} \tag{6.6}$$

On the other hand, since $\mathbb{E}[X_B(t)] \geq p$ for all $t \geq B$, we have

$$\begin{aligned} & \mathbb{P}\left(H\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) \leq \frac{pt}{2B}\right) \\ & = \mathbb{P}\left(\sum_{t'=0}^{\lfloor t/B \rfloor - 1} X_B(t') - p\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) \leq -\alpha_t\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right)\right) \\ & \leq \mathbb{P}\left(\sum_{t'=0}^{\lfloor t/B \rfloor - 1} \left(X_B(t') - \mathbb{E}[X_B(t')]\right) \leq -\alpha_t\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right)\right), \end{aligned}$$

where

$$\alpha_t = \frac{p\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) - \frac{pt}{2B}}{\left\lfloor \frac{t}{B} \right\rfloor - 1}. \tag{6.7}$$

When $t \geq B + \frac{2nB}{p}$, $\alpha_t > 0$ and hence, by Lemma A.6, we obtain

$$\mathbb{P}\left(H\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) \leq \frac{pt}{2B}\right) \leq \exp\left(-\alpha_t^2\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right)\right) \leq \exp\left(-\alpha_t^2\left(\frac{t}{B} - 2\right)\right). \tag{6.8}$$

From (6.7), we have

$$\begin{aligned}\alpha_t >& \frac{p\left(\frac{t}{B}-2\right)-\frac{pt}{2B}}{\frac{t}{B}}\\ =&\frac{p}{2}-\frac{2pB}{t}.\end{aligned}$$

If we let $\beta_t = \frac{p}{2} - \frac{2pB}{t}$, using (6.6) and (6.8), we conclude that

$$\mathbb{E}\left[\Lambda_{t,0}\right] \leq \exp\left(-\beta_t^2\left(\frac{t}{B}-2\right)\right) + 2\left(1-\frac{1}{n^{\frac{4nB}{p}}}\right)^{\frac{pt}{2nB}},$$

finishing the proof. □

Lemma 6.2.5 (to the best of our knowledge) for the first time explicitly presents on expectation convergence rate for the ergodicity of the product of random stochastic matrices.

Lemma 6.2.6. *In the push-sum algorithm* (3.3) *let* $\{W(t)\}$ *be a sequence of random column-stochastic matrices corresponding to the sequence* $\{P(t)\}$ *satisfying Assumption 6.1.1. We have, for all* $i \in [n]$ *and* $t \geq 0$*,*

$$\mathbb{E}\left[\ln\left(\frac{1}{y_i(t)}\right)\right] \leq \ln(n)\left(B\frac{n}{p}+B\right).$$

Proof. By Lemma A.1, for all $t < \frac{Bn}{p} + B$ and $i \in [n]$ we have

$$[W(t:0)]_{ii} \geq \frac{1}{n^{B\frac{n}{p}+B}},$$

almost surely. This implies that

$$\mathbb{E}\left[\ln\left(\frac{1}{y_i(t)}\right)\right] \leq \ln(n)\left(B\frac{n}{p}+B\right),$$

for all $t < \frac{Bn}{p} + B$ and $i \in [n]$. If $t \geq \frac{Bn}{p} + B$, let $t = aB + b$, where $a, b \in \mathbb{Z}_{\geq 0}$ and $b < B$. Define

$$\tau_t = \begin{cases} \min\{T : \sum_{t=a-T}^{a-1} X_B(t) = n\}, & \text{if } \sum_{t=0}^{a-1} X_B(t) \geq n \\ a & \text{otherwise.} \end{cases}$$

When $\tau_t = a$, $W_{ij}(t:0) \geq \frac{1}{n^{\tau_t B+B}}$, for all $i, j \in [n]$. When $\tau_t \neq a$, by Lemma A.2, $W(aB-1:(a-\tau_t)B)$ is a positive matrix and consequently by Lemma A.1

$$W_{ij}(t:(a-\tau_t)B) \geq \frac{1}{n^{\tau_t B+B}},$$

for all $i, j \in [n]$; in addition, since the $W(t)$ are column-stochastic, we have $W_{ij}(t:0) \geq \frac{1}{n^{\tau_t B+B}}$. Therefore, for all $t \geq 0$ we have

$$\ln\left(\frac{1}{W_{ij}(t:0)}\right) \leq \ln(n)(\tau_t B + B) \quad \text{for all } i, j \in [n].$$

Consider a sequence of independent Bernoulli trials Y_t, where in each trial the probability of success is p. The number of trials until n successes occur is a negative binomial random variable Z having parameters n and p. Since $\mathbb{P}(\tau_t \leq i) \geq \mathbb{P}(Z \leq i)$ for all $i \geq n$, we have $\mathbb{E}[\tau_t] \leq \mathbb{E}[Z]$. Since $\mathbb{E}[Z] = \frac{n}{p}$, we obtain $\mathbb{E}[\tau_t] \leq \frac{n}{p}$, and hence,

$$\mathbb{E}\left[\ln\left(\frac{1}{y_i(t)}\right)\right] \leq \ln(n)\left(B\frac{n}{p}+B\right).$$

□

We are now in a position to prove Theorem 6.2.1.

Proof of Theorem 6.2.1. In (5.1), since both sides are positive, we have

$$\begin{aligned}\ln\left(|z_i(t+1)-\bar{x}|\right) &\leq \ln\left(\frac{2\|x(0)\|_1}{y_i(t+1)}\Lambda_{t,0}\right)\\ &= \ln\left(2\|x(0)\|_1\right) + \ln\left(\frac{1}{y_i(t+1)}\right) + \ln\left(\Lambda_{t,0}\right).\end{aligned}$$

By taking expectations and using Lemma 6.2.6, we obtain

$$\begin{aligned}\mathbb{E}\left[\ln\left(|z_i(t+1)-\bar{x}|\right)\right] &\leq \ln\left(2\|x(0)\|_1\right) + \ln(n)\left(\frac{nB}{p}+B\right) + \mathbb{E}\left[\ln\left(\Lambda_{t,0}\right)\right]\\ &\leq \ln\left(2\|x(0)\|_1\right) + \ln(n)\left(\frac{nB}{p}+B\right) + \ln\left(\mathbb{E}\left[\Lambda_{t,0}\right]\right), \qquad (6.9)\end{aligned}$$

where the last inequality follows from Jensen's inequality. Now by Lemma 6.2.5, we have

$$\mathbb{E}\left[\Lambda_{t,0}\right] \leq \exp\left(-\beta_t^2\left(\frac{t}{B}-2\right)\right) + 2\left(1-\frac{1}{n^{\frac{4nB}{p}}}\right)^{\frac{pt}{2nB}},$$

where $\beta_t = \frac{p}{2} - \frac{2pB}{t}$. Let us consider the first term on the right hand side; since $\beta_t \leq \frac{1}{2}$

we have

$$\begin{aligned}
\exp\left(-\beta_t^2\left(\frac{t}{B}-2\right)\right) \leq& \exp\left(-\beta_t^2\frac{t}{B}+\frac{1}{2}\right)\\
=& \exp\left(-\frac{p^2t}{4B}+2p^2+\frac{1}{2}-\frac{4p^2B}{t}\right)\\
\leq& \exp\left(-\frac{p^2t}{4B}+\frac{5}{2}\right)\\
\leq& 13\exp\left(-\frac{p^2t}{4B}\right)\\
=& 13\left(\exp\left(-\frac{pn}{2}\right)\right)^{\frac{pt}{2nB}}.
\end{aligned}$$

Since $n \geq 2$, $\exp\left(-\frac{pn}{2}\right) \leq \exp(-p)$. On the other hand, $\left(1-\frac{1}{n^{\frac{4nB}{p}}}\right) \geq \left(1-\frac{1}{2^{\frac{8}{p}}}\right)$ for all $n \geq 2$ and $B \geq 1$. It can be seen the for $p \in [0,1]$, $\exp(-p) \leq \left(1-\frac{1}{2^{\frac{8}{p}}}\right)$, and consequently $\exp\left(-\frac{pn}{2}\right) \leq \left(1-\frac{1}{n^{\frac{4nB}{p}}}\right)$. Hence

$$\mathbb{E}\left[\Lambda_{t,0}\right] \leq 15\left(1-\frac{1}{n^{\frac{4nB}{p}}}\right)^{\frac{pt}{2nB}}. \tag{6.10}$$

Using (6.9) and (6.10) we conclude that

$$\begin{aligned}
\mathbb{E}\left[\ln\left(|z_i(t+1)-\bar{x}|\right)\right] \leq& \ln\left(2\|x(0)\|_1\right)+\ln(n)\left(\frac{nB}{p}+B\right)\\
&+\ln(15)+\frac{pt}{2nB}\ln\left(1-\frac{1}{n^{\frac{4nB}{p}}}\right).
\end{aligned}$$

□

Our approach and results also provide the (known) geometric convergence rate when the push-sum algorithm is considered for a B-strongly-connected deterministic sequence of column-stochastic matrices. A sequence of deterministic matrices $\{A(t)\}$

is B-strongly-connected if for some integer $B > 0$ and all $t \geq 0$, $\sum_{t'=tB}^{(t+1)B-1} A(t')$ is irreducible. For a B-strongly-connected sequence of column-stochastic matrices, $y_i(t) \geq \frac{1}{n^{nB}}$. In addition, if we let $p = 1$ in (6.10), we have

$$\Lambda_{t,0} \leq 15 \left(1 - \frac{1}{n^{4nB}}\right)^{\frac{t}{2nB}}.$$

Using (5.1), we obtain

$$|z_i(t+1) - \bar{x}| \leq \frac{30\|x(0)\|_1}{n^{nB}} \left(1 - \frac{1}{n^{4nB}}\right)^{\frac{t}{2nB}},$$

which shows the geometric convergence for B-strongly-connected sequences.

Chapter 7

Conclusions

7.1 Summary

We have studied the convergence properties of the push-sum algorithm for average consensus on sequences of random directed graphs. We have proved that this dynamics is convergent almost surely when some mild connectivity assumptions are met and the auxiliary states of nodes are uniformly bounded away from zero infinitely often. We have shown that the latter assumption holds for sequences of random matrices constructed using a sequence of time-varying B-irreducible probability matrices. We have also obtained convergence rates for the proposed push-sum algorithm.

7.2 Future Work

Future work includes studying the behavior of the push-sum algorithm when the random graphs are dependent, or with a different weight allocation method, and the implications of our results in scenarios with link-failure and in distributed optimization on random time-varying graphs. One can also consider the average consensus problem when the number of nodes is time-varying.

Chapter 8

Appendix

Lemma A.1 (Lemma 1 [14])**.** *Consider a sequence of directed graphs $\{\mathcal{G}(t)\}$, which we assume to contain all the self-loops, with a corresponding sequence of weighted adjacency matrices $\{A(t)\}$. In addition, assume that $A_{ij}(t) \geq \gamma$ whenever $A_{ij}(t) > 0$, for some $\gamma > 0$. Then the following statements hold:*

(i) $[A(t:s)]_{ii} \geq \gamma^{t-s+1}$, *for all* $i \in [n]$ *and* $t \geq s \geq 0$;

(ii) *if* $[A(r)]_{ij} > 0$ *for some* $t \geq r \geq s \geq 0$ *and* $i, j \in [n]$, *then* $[A(t:s)]_{ij} \geq \gamma^{t-s+1}$;

(iii) *if* $[A(s)]_{ik} > 0$ *and* $[A(r)]_{kj} > 0$ *for some* $t \geq r > s \geq 0$, *then* $[A(t:s)]_{ij} \geq \gamma^{t-s+1}$.

Lemma A.2. *For $n \geq 2$, let $\{A(i)\}_{i=1}^{n-1}$ be a sequence of weighted adjacency matrices associated with the strongly connected directed graphs $\{\mathcal{G}(i)\}_{i=1}^{n-1}$ on the node set $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, where $\mathcal{G}(i) = (\mathcal{V}, \mathcal{E}(i), A(i))$ and $A(i) \in S_n^+$ for all $i \in [n-1]$. Then the matrix product $A(n-1:1)$ is positive.*

Proof. Let $\mathcal{G}(k:1) = (\mathcal{V}, \mathcal{E}(k:1))$ indicate the directed graph associated with the product $A(k:1)$, where $k \in [n-1]$. Let $\mathcal{N}_i^{\text{out}}(k:1)$ and $d_i^{\text{out}}(k:1)$ indicate the set of out-neighbors and out-degree of node $i \in [n]$ in directed graph $\mathcal{G}(k:1)$, respectively. Consider an arbitrary but fixed node $i \in [n]$. Since $A(1) \in \mathcal{S}_n^+$ and $\mathcal{G}(1)$ is strongly connected, we have

$$d_i^{\text{out}}(1) \geq 2. \tag{8.1}$$

Now consider the directed graph $\mathcal{G}(k:1)$ and assume that $d_i^{\text{out}}(k:1) \leq n-1$ for some $k \in [n-1]$; we show that $d_i^{\text{out}}(k+1:1) > d_i^{\text{out}}(k:1)$. By Lemma A.1(ii), we have $\mathcal{N}_i^{\text{out}}(k:1) \subseteq \mathcal{N}_i^{\text{out}}(k+1:1)$. Moreover, since $\mathcal{G}(k+1)$ is strongly connected and $d_i^{\text{out}}(k:1) \leq n-1$, there is $l \notin \mathcal{N}_i^{\text{out}}(k:1)$ such that $l \in \mathcal{N}_j^{\text{out}}(k+1)$ for some $j \in \mathcal{N}_i^{\text{out}}(k:1)$; otherwise, there is no path between i and l in $\mathcal{G}(k+1)$, contradicting the strong connectivity of $\mathcal{G}(k+1)$. Hence, by Lemma A.1 (iii) $l \in \mathcal{N}_i^{\text{out}}(k+1:1)$, implying that

$$d_i^{\text{out}}(k+1:1) > d_i^{\text{out}}(k:1).$$

This along with (8.1) imply that

$$d_i^{\text{out}}(k:1) \geq k+1, ,$$

for all $k \in [n-1]$, which implies that $d_i^{\text{out}}(n-1:1) = n$. Since this statement holds for any $i \in [n]$, the matrix product $A(n-1:1)$ is positive. $\square$

Lemma A.3 (Lemma 3 [9]). *For row-stochastic matrices A, B and $C = AB$, we* *have:*

$$\max_{j\in[n]}(\max_{i\in[n]} C_{ij} - \min_{i\in[n]} C_{ij}) \leq \left(1 - \min_{i,i'\in[n]} \sum_{j\in[n]} \min\left(A_{ij}, A_{i'j}\right)\right) \max_{j\in[n]}(\max_{i\in[n]} B_{ij} - \min_{i\in[n]} B_{ij}).$$

Lemma A.4 (Corollary 5.3.6 [2])**.** *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a strongly connected directed graph with n vertices. Then, there exists a subset $\tilde{\mathcal{E}} \subseteq \mathcal{E}$ with at most $2(n-1)$ edges such that the graph $\tilde{\mathcal{G}} = (\mathcal{V}, \tilde{\mathcal{E}})$ is strongly connected.*

Lemma A.5 (Lemma 4.3 [3])**.** *Consider the push-sum algorithm* (3.3)*. Define*

$$f(t) = \max_{i\in[n]} \frac{\sum_{j=1}^{n} |[W(t:0)]_{ij} - \frac{1}{n}\sum_{k=1}^{n}[W(t:0)]_{ik}|}{y_i(t)}.$$

Then, $f(t)$ is non-increasing and

$$\|z(t) - \bar{x}\mathbf{1}_n\|_\infty \leq \|x(0)\|_\infty f(t).$$

Lemma A.6 (Hoeffding's inequality [8])**.** *If $X_1, X_2, \ldots, X_n$ are independent random variables and $0 \leq X_i \leq 1$, for all $i \in [n]$, then for any $\alpha > 0$, we have*

$$\mathbb{P}\left(\sum_{i=1}^{n}\left(X_i - \mathbb{E}[X_i]\right) \leq -\alpha n\right) \leq \exp\left(-2\alpha^2 n\right).$$

Lemma A.7. *For $n > 1$ and for all $l_1, l_2, \ldots, l_q \in \mathbb{Z}_{\geq 0}$, $q > 0$, we have*

$$\prod_{i=1}^{q}\left(1 - \frac{1}{n^{l_i}}\right) \leq \left(1 - \frac{1}{n^{\frac{t}{q}}}\right)^{q},$$

where $t = l_1 + l_2 + \cdots + l_q$.

Proof. It suffices to show that

$$\frac{1}{q}\sum_{i=1}^{q}\ln\left(1-\frac{1}{n^{l_q}}\right)\leq\ln\left(1-\frac{1}{n^{\frac{t}{q}}}\right),$$

which simply follows from Jensen's inequality, since the function $g(\zeta)=\ln\left(1-\frac{1}{n^{\zeta}}\right)$ is concave. $\square$

Bibliography

[1] M. Akbari, B. Gharesifard, and T. Linder. Distributed online convex optimization on time-varying directed graphs. *IEEE Transactions on Control of Network Systems*, 2017. to appear.

[2] Jørgen Bang-Jensen and Gregory Z. Gutin. *Digraphs: Theory, Algorithms and Applications*. Springer London, 2009.

[3] F. Bénézit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli. Weighted gossip: Distributed averaging using non-doubly stochastic matrices. In *IEEE International Symposium on Information Theory Proceedings*, pages 1753–1757. IEEE, 2010.

[4] S. Chatterjee and E. Seneta. Towards consensus: Some convergence theorems on repeated averaging. *Journal of Applied Probability*, 14(1):89–97, 1977.

[5] A. D. Dominguez-Garcia and C. N. Hadjicostis. Distributed matrix scaling and application to average consensus in directed graphs. *IEEE Transactions on Automatic Control*, 58(3):667–681, 2013.

[6] R. Durrett. *Probability: Theory and Examples*. Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 4th edition, 2010.

[7] B. Gharesifard and J. Cortés. Distributed continuous-time convex optimization on weight-balanced digraphs. *IEEE Transactions on Automatic Control*, 59(3):781–786, 2014.

[8] W. Hahn. *Theory and Application of Liapunov's Direct Method*. Prentice Hall, 1963.

[9] J. Hajnal and M. S. Bartlett. Weak ergodicity in non-homogeneous markov chains. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 54, pages 233–246. Cambridge University Press, 1958.

[10] J. M. Hendrickx and J. N. Tsitsiklis. Fundamental limitations for anonymous distributed systems with broadcast communications. In *Allerton Conf. on Communications, Control and Computing*, pages 9–16, 2015.

[11] B. Johansson, M. Rabi, and M. Johansson. A randomized incremental subgradient method for distributed optimization in networked systems. *SIAM Journal on Control and Optimization*, 20(3):1157–1170, 2009.

[12] D. Kempe, A. Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *IEEE Symposium on Foundations of Computer Science*, pages 482–491, Washington, DC, October 2003.

[13] A. Nedic and A. Olshevsky. Distributed optimization over time-varying directed graphs. *IEEE Transactions on Automatic Control*, 60(3):601–615, 2015.

[14] A. Nedic and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Transactions on Automatic Control*, 54(1):48–61, 2009.

[15] A. Nedic, A. Ozdaglar, and P. A. Parrilo. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4):922–938, 2010.

[16] M. Rabbat and R. Nowak. Distributed optimization in sensor networks. In *Symposium on Information Processing of Sensor Networks*, pages 20–27, Berkeley, CA, April 2004.

[17] P. Rezaeinia, B. Gharesifard, T. Linder, and B. Touri. Push-sum on random graphs. *IEEE Transactions on Automatic Control*, submitted 2017.

[18] B. Touri. *Product of random stochastic matrices and distributed averaging.* Springer Science & Business Media, 2012.

[19] Behrouz Touri and Cedric Langbort. On endogenous random consensus and averaging dynamics. *IEEE Transactions on Control of Network Systems*, 1(3):241–248, 2014.

[20] Behrouz Touri and Angelia Nedić. Product of random stochastic matrices. *IEEE Transactions on Automatic Control*, 59(2):437–448, 2014.

[21] K. I. Tsianos, S. Lawlor, and M. G. Rabbat. Push-sum distributed dual averaging for convex optimization. In *IEEE Conf. on Decision and Control*, pages 5453–5458, Maui, HI, 2012.

[22] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, 1986.

[23] P. Wan and M. D. Lemmon. Event-triggered distributed optimization in sensor networks. In *Symposium on Information Processing of Sensor Networks*, pages 49–60, San Francisco, CA, 2009.

[24] J. Wang and N. Elia. Control approach to distributed optimization. In *Allerton Conf. on Communications, Control and Computing*, pages 557–561, Monticello, IL, October 2010.

[25] J. Wang and N. Elia. A control perspective for centralized and distributed convex optimization. In *IEEE Conf. on Decision and Control*, pages 3800–3805, Orlando, Florida, 2011.

[26] L. Xiao and S. Boyd. Optimal scaling of a gradient method for distributed resource allocation. *Journal of Optimization Theory & Applications*, 129(3):469–488, 2006.

[27] L Xiao, S. Boyd, and S. Lall. Distributed average consensus with time-varying metropolis weights. *Automatica*, 2006.