Title
Distributed Averaging Dynamics and Optimization over Random Networks

Author
Aghajan Abdollah, Adel

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Distributed Averaging Dynamics and Optimization over Random Networks

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Communication Theory and Systems)

by

Adel Aghajan Abdollah

Committee in charge:

Professor Behrouz Touri, Chair
Professor Massimo Franceschetti
Professor Philip E. Gill
Professor Piya Pal
Professor Meisam Razaviyayn
Professor Alireza Salehi Golsefidy

2021

The dissertation of Adel Aghajan Abdollah is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my parents, Azam and Hasan

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

First, I would like to express my sincere gratitude to my academic advisor, Professor Behrouz Touri, who expertly mentored me through my PhD, for his support, guidance, and patience. Many thanks are also due to the other members of my dissertation committee, Professor Massimo Franceschetti, Professor Philip E. Gill, Professor Piya Pal, Professor Meisam Razaviyayn, and Professor Alireza Salehi Golsefidy, for kindly accepting to be the members of my thesis committee and their helpful comments.

Furthermore, I would like to thank Professor Tara Javidi, for her support and guidance during my PhD. I would also like to express my sincere gratitude to my Master adviser, Professor Khosravifard, who is my first research adviser, and hence, has had a significant impact on the way I do research.

I am thankful to my friends Shekhar and Ayoob who are supportive during my PhD especially when I got Covid-19. Finally, and most importantly, I would like to thank my family, Hasan, Azam, and Masoud, for their unconditional support and encouragement.

Chapter 2, in full, is a reprint of the material as it appears in A. Aghajan and B. Touri, On ergodicity of time-varying distributed averaging dynamics, *2020 American Control Conference (ACC)*, IEEE, 2020, pp. 4417–4422. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in A. Aghajan and B. Touri, Ergodicity of continuous-time distributed averaging dynamics: A spanning directed rooted tree approach, being accepted for publication in IEEE Transactions on Automatic Control. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in A. Aghajan and B. Touri, Distributed optimization over dependent random networks, being submitted for publication. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, is a reprint of the material as it appears in A. Aghajan and B. Touri,

Geometric convergence for distributed optimization over dependent random networks, being prepared for publication. The dissertation author was the primary investigator and author of this paper.

VITA

2009	B. E. in Electrical Engineering, Isfahan University of Technology, Isfahan
2012	M. E. in Electrical Engineering, Isfahan University of Technology, Isfahan
2021	Ph. D. in Electrical and Computer Engineering, University of California San Diego

PUBLICATIONS

A. Aghajan and B. Touri, Distributed optimization over dependent random networks,arXiv preprint arXiv:2010.01956, submitted, 2020,

A. Aghajan and B. Touri, Ergodicity of continuous-time distributed averaging dynamics: A spanning directed rooted tree approach, *accepted for publication in IEEE Transactions on Automatic Control*,

A. Aghajan and B. Touri, On ergodicity of time-varying distributed averaging dynamics, in *2020 American Control Conference (ACC)*, IEEE, 2020, pp. 4417–4422.

A. Aghajan and B. Touri, Geometric convergence for distributed optimization over dependent random networks, *in preparation*.

ABSTRACT OF THE DISSERTATION

Distributed Averaging Dynamics and Optimization over Random Networks

by

Adel Aghajan Abdollah

Doctor of Philosophy in Electrical Engineering (Communication Theory and Systems)

University of California San Diego, 2021

Professor Behrouz Touri, Chair

In this thesis, we study distributed averaging dynamics and its main application, i.e., distributed optimization. More specifically, the results of this thesis can be divided into two main parts: **(i)** Ergodicity of distributed averaging dynamics, and **(ii)** Distributed optimization over dependent random networks.

On Topic (i), we study both discrete-time and continuous-time time-varying distributed averaging dynamics. First, we show a necessary and a sufficient condition for ergodicity of discrete-time time-varying distributed averaging dynamics. We extend a well-known result in ergodicity of time-homogeneous (time-invariant) averaging dynamics and we show that ergodicity of a dynamics necessitates that its (directed) infinite flow graph has a spanning rooted tree.

Then, we show that if groups of agents are connected using a rooted tree and the averaging dynamics restricted to each group is $\mathcal{P}^*$ and ergodic, then the dynamics over the whole networks is ergodic. In particular, this provides a general condition for convergence of consensus dynamics where groups of agents, which are capable of reaching consensus in the absence of other agents, follow each other on a time-varying network. Then, we extent this condition for convergence of consensus dynamics to continuous-time distributed averaging dynamics.

On Topic (ii), we study distributed optimization solvers over random networks for both convex and strongly convex functions. We show a general result on the convergence of such schemes for a broad class of dependent weight-matrix sequences. In addition to implying many of the previously known results on this domain, our work shows the robustness of distributed optimization results to link-failure. Also, it provides a new tool for synthesizing distributed optimization algorithms. To prove our main theorems on this topic, we establish new results on the rate of convergence analysis of averaging dynamics and non-averaging dynamics over (dependent) random networks. These secondary results, along with the required martingale-type results to establish them, might be of interest to broader research endeavors in distributed computation over random networks.

Chapter 1

Introduction

Distributed averaging dynamics has received increasing attention in recent years due to its applications in distributed control of robotic networks (see e.g., [10]), study of opinion dynamics in social networks (see e.g., [17, 24, 2]), distributed estimation and signal processing (see e.g., [50, 13, 63]), and power networks (see e.g., [19, 18, 16]). However, the main application of distributed averaging dynamics is distributed optimization (see e.g., [76, 28, 46, 45, 66]), which is the main subject of this thesis. Before studying distributed optimization, we study distributed averaging dynamics and prove some results for ergodicity of those dynamics. Therefore, the material of the thesis can be divided into two main topics:

1. Ergodicity of Distributed Averaging Dynamics
2. Distributed Optimization Over Dependent Random Networks

1.1 Ergodicity of Distributed Averaging Dynamics

A discrete-time distributed averaging dynamics is a linear time-varying dynamics $x : \{t_0, t_0+1, \ldots\} \to \mathbb{R}^n$ driven by stochastic matrices $W(t) = [w_{ij}(t)]_{n\times n}$ with some initial condition $x(t_0) \in \mathbb{R}^n$, i.e.,

$$x(t+1) = W(t)x(t), \qquad \text{for } t \geq t_0. \tag{1.1}$$

A continuous-time distributed averaging dynamics is a linear time-varying dynamics[1]

$$\dot{x}(t) = A(t)x(t) \qquad \text{for } t \geq t_0, \tag{1.2}$$

driven by a Laplacian process $A(t) = [a_{ij}(t)]_{n\times n}$ with an initial condition $x(t_0) \in \mathbb{R}^n$. In this thesis, we assume that for all $i, j \in [n]$, $a_{ij}(t)$ is a measurable function of time and $\int_{\tau}^{t} a_{ij}(s)\, ds < \infty$ for all $t_0 \leq \tau \leq t$, implying that (1.2) has a unique continuous solution [62].

We refer to the solution $x(t)$ of (1.2) (resp. (1.1)) as the dynamics driven by $\{A(t)\}$ (resp. $\{W(t)\}$). It can be shown that the transition matrices of a Laplacian process are stochastic matrices [37]. This fact makes the continuous-time and discrete-time distributed averaging dynamics closely related. In this thesis, we study both the discrete-time and continuous-time distributed averaging dynamics.

The main problem that we deal with in this part of the thesis is Ergodicity. We say that $\{W(t)\}$ is ***ergodic*** if $\lim_{t\to\infty} x(t) = \bar{x}e$ for some $\bar{x} \in \mathbb{R}$, and for all initial time $t_0 \in \mathbb{Z}^+$ and all choices of initial condition $x(t_0) \in \mathbb{R}^n$ in the dynamics (1.1). Similarly, $\{A(t)\}$ is ***ergodic*** if $\lim_{t\to\infty} x(t) = \bar{x}e$ for some $\bar{x} \in \mathbb{R}$, and for all initial time $t_0 \in \mathbb{R}^+$ and all choices of initial condition $x(t_0) \in \mathbb{R}^n$ for the dynamics (1.2).

For time-invariant continuous-time distributed averaging dynamics, having a spanning directed rooted tree in the associated graph is shown to be the necessary and sufficient condition for ergodicity [56]. Also, the counterpart of this condition for time-invariant discrete-time distributed averaging dynamics besides having a root with a self-loop is the weakest sufficient

[1] In this thesis, for consistency of notation, we use t to denote time variable for both discrete-time and continuous-time settings.

condition for ergodicity in terms of associated graph [80].

We study time-varying distributed averaging dynamics. We can divide the literature of time-varying distributed averaging dynamics into two parts, based on the assumption on whether there is a uniform lower bound on the non-zero entries of the underlying matrices or not:

1. *With a uniform lower bound on the non-zero entries:* Most of the existing works on distributed averaging assume that over every given fixed time window, there is guaranteed connectivity between the nodes in the time-varying dynamics. In this regard, it is proved that a dynamics is ergodic if the union of the associated graphs over every window of a given fixed length has a spanning directed rooted tree [76, 77, 27, 11, 12, 55, 40].

2. *Without a uniform lower bound on the non-zero entries:* In this case, which has been investigated more recently, instead of focusing on non-zero entries, the focus is on the study of the unbounded interactions graph of the underlying process. This approach was introduced in [69] by introducing the infinite flow graph of an underlying chain, which is a generalization of the associated graph in time-invariant dynamics to time-varying dynamics, and a class of chains, named $\mathcal{P}^*$, which is a generalization of non-negative matrices with positive Perron eigenvectors to non-negative matrix chains. Later this approach was extended to the continuous-time dynamics in [6, 8, 25, 37]. In [25], it was shown that the connectivity of the infinite flow graph and instantaneous cut-balancedness conditions are sufficient conditions for ergodicity. This result was further extended for average cut-balanced processes in [37]. In [8] it was shown that the instantaneously cut-balanced processes are in fact in class $\mathcal{P}^*$.

In the first part of the thesis, we take the modern approach, described above, and do not assume a uniform lower bound on the non-zero entries of the underlying chains. For distributed optimization (i.e., the second part of the thesis), to ensure fast-enough mixing, we assume a condition similar to uniform lower bound condition on the entries of the underlying matrix sequence.

1.2 Distributed Optimization Over Dependent Random Networks

In distributed optimization, we are often interested in finding an optimizer of a decomposable function

$$F(z) = \sum_{i=1}^{n} f_i(z)$$

such that $f_i(\cdot)$s are distributed through a network of n agents, i.e., agent i only knows $f_i(\cdot)$, and we are seeking to solve this problem without sharing the local objective functions. Therefore, the goal is to find distributed dynamics over (possibly time-varying) communication networks that, asymptotically, all the nodes agree on an optimizer of $F(\cdot)$.

The most well-know algorithm that achieves this is, what we refer to as, the averaging-based distributed optimization solver, is introduced in [45] which is the dynamics of the form

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \eta(t)\mathbf{g}(t), \tag{1.3}$$

where $W(t) = [w_{ij}(t)]$ is a weight matrix, $\mathbf{x}_i(t)$ is agent i's estimate of an optimizer of $F(\cdot)$, $\mathbf{g}_i(t) \in \mathbb{R}^m$ is a subgradient of $f_i(z)$ at $z = \mathbf{x}_i(t)$ for $i \in [n]$, and $\{\eta(t)\}$ is a step-size sequence[2]. In this algorithm, each node maintains an estimate of an optimal point, and at each time step, each node computes the average of the estimates of its neighbors and performs (sub-)gradient descent on its local objective function. To show the convergence of such an algorithm, the corresponding weight matrices are often assumed to be doubly stochastic. Doubly stochasticity of weight matrices causes that the average of estimates of all agents behaves similar to the gradient descent dynamics for $F(\cdot)$

While making a row-stochastic or a column-stochastic matrix is easy, this is not necessarily

[2]The constant step-size variation of this dynamics was studied in [45].

the case for doubly stochastic matrices. By taking average of the neighbors' estimates that are received by all agents, we can construct a row-stochastic weight matrix. Also, if every agent sends out its estimate after dividing it by its out-degree, we can construct an update mechanism involving column-stochastic weight matrices. However, we cannot do both of these procedures simultaneously to construct a doubly stochastic matrix sequence and constructing such a sequence often requires extra efforts and assumptions. For example, to construct a doubly stochastic sequence, one can use Metropolis algorithm. In this algorithm, it is assumed that all the nodes are connected through an undirected graph sequence, and every node knows the degree of all of its neighbors (in addition to its own degree). In this case, we can set

$$w_{ij}(t) = \begin{cases} \frac{1}{\max\{d_i(t), d_j(t)\}}, & \text{if } i \neq j \\ 1 - \sum_{\ell \neq i} w_{i\ell}(t), & \text{if } i = j \end{cases},$$

where $d_i(t)$ is the degree of node i at time t. It can be verified that such a matrix is doubly stochastic for all t.

In vast majority of the distributed optimization literature, it is assumed that, through such mechanisms, a doubly stochastic matrix sequence is established and is given. A solution to avoid such an assumption is to establish more complicated distributed algorithms that effectively *reconstruct* the average-state distributively. The first algorithm in this category was proposed in [73, 75, 74], which is called push-sum or subgradient-push, and later was extended for time-varying networks [43]. In this scheme, the weight matrices are assumed to be column-stochastic, and through the use of auxiliary state variables the approximate average state is reconstructed. Another scheme in this category that works with row-stochastic matrices, but does not need the column-stochastic assumption, is proposed in [36, 79]. However, to use this scheme, every node needs to be assigned and know its unique label. Assigning those labels distributively is also another challenge in this respect. In addition, both these schemes invoke division operation which results in theoretical challenges in establishing their stability in random networks [58, 57].

An alternative way to avoid using doubly stochastic weight matrices is to use gossip-based algorithms over random networks [42, 31]. In gossip algorithms, which were originally studied in [9, 3], at each round, a node randomly wakes up and shares its value with all or some of its neighbors. The weight matrices of gossip-based algorithms are row-stochastic and in-expectation column-stochastic. This fact was generalized in [4, 41], where it is proven that it is sufficient to have row-stochastic weight matrices, that are column-stochastic in-expectation and satisfy certain connectivity assumptions. In all the above works on distributed optimization over random networks, all weight matrices are assumed to be independent and identically distributed (i.i.d.) or independent. In [34], a broader class of random networks, i.e., Markovian networks, was studied for distributed optimization; however, weight matrices were assumed to be doubly stochastic almost surely. As mentioned, distributed optimization is an application of distributed averaging dynamics. Therefore, our work in this part of the thesis is also closely related to the existing works on distributed averaging on random networks [29, 64, 70, 65].

If the objective function is β-smooth and α-strongly convex, the centralized variation of (1.3), which is the gradient descent algorithm, geometrically converges to the minimizer with a constant step-size. However, the dynamics (1.3) cannot converge to the minimizer with a constant step-size. Even worse, Theorem 6 in [49] proves that any dynamics similar to (1.3) cannot converge to the global minimizer geometrically fast. To remedy this, in [82, 49, 44], the following dynamics is proposed

$$\mathbf{x}(t+1) = W^x(t+1)\mathbf{x}(t) - \eta\mathbf{s}(t), \tag{1.4}$$

$$\mathbf{s}(t+1) = W^s(t+1)\mathbf{s}(t) + \mathbf{g}(t+1) - \mathbf{g}(t), \tag{1.5}$$

where both $\{W^x(t)\}$ and $\{W^s(t)\}$ are doubly stochastic sequences and its geometric convergence rate is established there. Since, in this dynamics, $\mathbf{s}(t)$ tracks the gradient of $F(\cdot)$, this dynamics is referred to as a distributed gradient-tracking algorithm. Later, in [81] (time-invariant setting) and [60] (time-varying setting), it was shown that having a row-stochastic sequence of $\{W^x(t)\}$ and a column-stochastic sequence of $\{W^s(t)\}$ is enough for the convergence of the dynamics (1.4)-(1.5) to the minimizer.

In this thesis, we consider communication networks that have link-failure. Hence, each agent receives perfect information or no information. However, links in the communication networks can also be noisy where each agent receives imperfect information. This problem was studied in [52, 78, 53] for time-invariant networks and in [54] for time-varying networks.

1.3 Dissertation Overview

The rest of this dissertation is organized as follows.

In Chapter 2, we consider discrete-time time-varying distributed averaging dynamics. We show a necessary and a sufficient condition for ergodicity of such dynamics. First, we extend a well-known result in ergodicity of time-homogeneous (time-invariant) averaging dynamics and we show that ergodicity of a dynamics necessitates that its (directed) infinite flow graph has a spanning rooted tree. Then, we show that cut-balanced and even class $\mathcal{P}^*$ assumption on the underlying processes is a restrictive assumption and there are ergodic processes that do not have these conditions and the focus of this study is on extending these conditions for ergodicity of time-varying dynamics. More specificity, we show that if groups of agents are connected using a rooted tree and the averaging dynamics restricted to each group is $\mathcal{P}^*$ and ergodic, then the dynamics over the whole network is ergodic.

In Chapter 3, we consider continuous-time time-varying distributed averaging dynamics and extend the results of Chapter 2 to continuous-time dynamics. More specificity, motivated by a necessary condition on the ergodicity, we provide a sufficient condition for the ergodicity of such dynamics. We show that if groups of agents are connected using a directed acyclic graph containing a spanning directed rooted tree and the averaging dynamics restricted to each group is $\mathcal{P}^*$, then the dynamics over the whole network is ergodic. Also, we discuss how much our sufficient condition generalizes the known result in the literature in this chapter.

In Chapter 4, we study distributed optimization over random networks, where the randomness is not only time-varying but also, possibly, dependent on the past. Under the standard assumptions on the local objective functions and step-size sequences for the gradient descent algorithm, we show that the averaging-based distributed optimization solver at each node converges to a global optimizer almost surely if the weight matrices are row-stochastic almost surely, column-stochastic in-expectation, and satisfy certain connectivity assumptions. It is worth mentioning that to prove the main result in this chapter, we establish new results on the rate of convergence analysis of averaging dynamics over (dependent) random networks.

In Chapter 5, we study the gradient-tracking distributed optimization solvers over random dependent networks. In this chapter, we consider a probabilistic model for the underlying random networks similar to the one introduced in Chapter 4. We show that as in the deterministic setting, for strongly convex and smooth functions, the gradient-tracking algorithm finds the minimizer geometrically fast almost surely in this probabilistic setting. In this algorithm, we use two weight matrices, one for tracking the minimizer and one for tracking the gradient. While the gradient-tracking algorithm still works if the weight matrix corresponding to the minimizer is just row-stochastic, we show that the weight matrix corresponding to the gradient needs to be at least column stochastic. To prove our main theorem, we study the limiting behavior of products of random matrices and establish a sufficient condition for the convergence of the limit to zero.

1.4 Notation and Basic Terminology

The following notation and terminologies are used throughout this thesis.

Sets: Let $[n] \triangleq \{1, \ldots, n\}$, and for $S \subset [n]$, $\bar{S}$ be the complement of S, i.e., $\bar{S} = [n] \setminus S$. We denote the space of real numbers by $\mathbb{R}$ and natural numbers by $\mathbb{N}$. We denote the space of n-dimensional real-valued vectors by $\mathbb{R}^n$.

Vectors: In this thesis, all vectors are assumed to be column vectors. The transpose of a vector $x \in \mathbb{R}^n$ is denoted by x^T. For a vector $x \in \mathbb{R}^n$, x_i represents the ith coordinate of x, except, where for notational convenience, we denote $e_1^n, \ldots, e_n^n$ as the standard basis vectors of $\mathbb{R}^n$. We denote the all-one vector in $\mathbb{R}^n$ by $e^n = [1, 1, \ldots, 1]^T$. We drop the superscript n in e^n and $e_1^n, \ldots, e_n^n$ whenever the dimension of the space is understandable from the context. For convenience and due to the frequent use of ℓ_2 norm in the paper, we use $\|\cdot\|$ to denote the ℓ_2 norm $\|x\| = \sqrt{\sum_{i=1}^m x_i^2}$. Also, we denote ℓ_∞ norm with $\|x\|_\infty \triangleq \max_{i \in [m]} |x_i|$.

Matrices: For two $m \times n$ matrices $A = [a_{ij}]$ and $B = [b_{ij}]$, we use $A \geq B$ when $a_{ij} \geq b_{ij}$ for all $i \in [m]$ and $j \in [n]$. A non-negative vector x is a stochastic vector if $x^T e = 1$ and a non-negative matrix A is a stochastic (or row-stochastic) matrix if $Ae = e$. A non-negative matrix A is a sub-stochastic matrix if $Ae \leq e$, and it is a column-stochastic matrix if $e^T A = e^T$. A non-negative matrix $A = [a_{ij}]$ is a Laplacian matrix if $Ae = 0e$ and it is a sub-Laplacian matrix if $Ae \leq 0e$. We refer to a continuous-time indexed sequence of (sub-)Laplacian matrices $\{A(t)\}$ as a (sub-)Laplacian process. A matrix A is cut-balanced with parameter $K \geq 1$ if for all non-empty proper subsets $S \subset [n]$, there holds

$$\sum_{i \in S, j \notin S} a_{ij} \leq K \sum_{i \notin S, j \in S} a_{ji}.$$

For a matrix A, we write $\|A\|_\infty$ to denote the induced matrix norm induced by the vector norm $\|\cdot\|_\infty$. We use block matrix notation $B = [A_{ij}]_{m \times n}$ to represent a matrix B that is composed of matrices A_{ij} for all $i \in [m]$ and $j \in [n]$, where the matrix A_{ij} is the ijth sub-matrix of B. In particular, $[C, D]$ is a block matrix composed of C and D.

Graphs: In this thesis, unless otherwise stated, all graphs are directed graphs. A directed graph $\mathcal{G} = ([n], \mathcal{E})$ (on n vertices) is defined by a vertex set (identified by) $[n]$ and an edge set $\mathcal{E} \subseteq [n] \times [n]$. A graph $\mathcal{G} = ([n], \mathcal{E})$ has a spanning directed rooted tree if it has a vertex $r \in [n]$ as a root such that there exists a (directed) path from r to every other vertex $\hat{r} \in [n]$, i.e., there exists a sequence of vertices $r_0 = r, r_1, \ldots, r_k = \hat{r} \in [n]$, such that $(r_i, r_{i+1}) \in \mathcal{E}$ for $i = 0, \ldots, k-1$. A directed acyclic graph containing a spanning directed rooted tree is a graph $\mathcal{G} = ([n], \mathcal{E})$ having a spanning directed rooted tree such that $\mathcal{G}$ does not contain any directed cycle. For a matrix $A = [a_{ij}]_{n \times n}$, the associated directed graph with parameter $\gamma > 0$ is the graph $\mathcal{G}^{\gamma}(A) = ([n], \mathcal{E}^{\gamma}(A))$ with the edge set $\mathcal{E}^{\gamma}(A) = \{(j,i) \mid i, j \in [n], a_{ij} > \gamma\}$. Later, we fix the value $0 < \gamma < 1$ when it is clear from the context, and hence, unless otherwise stated, for notational convenience, we use $\mathcal{G}(A)$ and $\mathcal{E}(A)$ instead of $\mathcal{G}^{\gamma}(A)$ and $\mathcal{E}^{\gamma}(A)$. Finally, for a matrix $A = [a_{ij}]_{n \times n}$, the associated or underlying directed graph is the graph $\mathcal{G}^0(A) = ([n], \mathcal{E}^0(A))$.

Probability: Let $(\Omega, \mathcal{F}, \Pr)$ be a probability space and let $\{W(t)\}$ be a chain of random matrices, i.e., for all $t \geq 0$ and $i, j \in [n]$, $w_{ij}(t) : \Omega \to \mathbb{R}$ is a Borel-measurable function. For random vectors (variables) $x(0), \ldots, x(t)$, we denote the σ-algebra generated by these random vectors by $\sigma(x(0), \ldots, x(t))$. We say that $\{\mathcal{F}(t)\}$ is a filtration for $(\Omega, \mathcal{F})$ if $\mathcal{F}(0) \subseteq \mathcal{F}(1) \subseteq \cdots \subseteq \mathcal{F}$. Further, we say that a random process $\{V(t)\}$ (of random variables, vectors, or matrices) is adapted to $\{\mathcal{F}(t)\}$ if $V(t)$ is measurable with respect to $\mathcal{F}(t)$.

Functions: The function $f : \mathbb{R}^m \to \mathbb{R}$ is convex if for all $x, y \in \mathbb{R}^m$ and all $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta) y) \leq \theta f(x) + (1-\theta) f(y).$$

We say that $g \in \mathbb{R}^m$ is a subgradient of the function $f(\cdot)$ at $\hat{x}$ if for all $x \in \mathbb{R}^m$, $f(x) - f(\hat{x}) \geq \langle g, x - \hat{x} \rangle$, where $\langle u_1, u_2 \rangle = u_1^T u_2$ is the standard inner product in $\mathbb{R}^m$. The set of all subgradients of $f(\cdot)$ at x is denoted by $\nabla f(x)$. For a convex function $f(\cdot)$, $\nabla f(x)$ is not empty for all $x \in \mathbb{R}^m$ (see e.g., Theorem 3.1.15 in [48]). If the function $f(\cdot)$ is differentiable, $\nabla f(x)$ denotes the gradient of $f(\cdot)$. The function $f : \mathbb{R}^m \to \mathbb{R}$ is β-smooth for some $\beta > 0$ if it is differentiable, and its gradient is β-Lipschitz, i.e., for all $x, y \in \mathbb{R}^m$,

$$\|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\|.$$

For $\alpha > 0$, a function f is α-strongly convex if for all $x, y \in \mathbb{R}^m$

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2}\|y - x\|^2.$$

Miscellaneous: We say that a sequence $\{f(t)\}$ is summable if $\sum_\tau f(\tau) < \infty$. We say that a function $f(t)$ is integrable if $\int_{t_0}^{\infty} f(\tau)\, d\tau < \infty$.

Part I

Ergodicity of Distributed Averaging Dynamics

Chapter 2

Discrete-Time Distributed Averaging Dynamics

In this chapter, we study discrete-time distributed average dynamics and prove some results about the ergodicity of those dynamics

The structure of this chapter is as follows: In Section 2.1, we state the main definitions related to this chapter and formulate the problem of interest. In Section 2.2, we state the main results which are a necessary condition and a sufficient condition for ergodicity. In this section, we also provide our motivation to prove those results. In Section 2.3, we prove the necessary condition, and in Section 2.4, we prove the sufficient condition.

2.1 Main Definitions and Problem Statement

In this section, we discuss the main problem and definitions of this chapter.

We study linear time-varying dynamics $x : \{t_0, t_0 + 1, \ldots\} \to \mathbb{R}^n$ driven by stochastic matrices $A(t) = [a_{ij}(t)]_{n\times n}$ with some initial condition $x(t_0) \in \mathbb{R}^n$, i.e.,

$$x(t+1) = A(t)x(t), \qquad \text{for } t \geq t_0. \tag{2.1}$$

For convenience, we refer to $x(t)$ as a dynamics driven by $\{A(t)\}$.

Definition 1. *Let $x(t)$ be the dynamics driven by $\{A(t)\}$. We say that a chain $\{A(t)\}$ is ergodic if $\lim_{t\to\infty} x(t) = \bar{x}e$ for some $\bar{x} \in \mathbb{R}$, for all initial time $t_0 \in \mathbb{N}$ and all choices of initial condition $x(t_0) \in \mathbb{R}^n$.*

From (2.1), we have $x(t) = \Phi(t,\tau)x(\tau)$ where $\Phi(t,\tau) = A(t-1)\cdots A(\tau)$ is the transition matrix for (2.1) for all $t \geq \tau \geq t_0$. Since the product of any two stochastic matrices is also a stochastic matrix, the transition matrices $\Phi(t,\tau)$ are all stochastic matrices. An equivalent condition for ergodicity based on the transition matrices is that (2.1) is ergodic if and only if $\lim_{t\to\infty} \Phi(t,\tau) = e\pi^T(\tau)$ for a stochastic vector $\pi(\tau) \in \mathbb{R}^n$ and all $\tau \geq t_0$. We say that $\{A(t)\}$ is ergodic if this condition holds.

When considering time-invariant dynamics, i.e., when $A(t)$ is constant ($A(t) = A$ for all t and a stochastic matrix A), interestingly, the limiting behavior of the dynamics (2.1) becomes closely related to the graph theoretic properties of the associated graph of A. Indeed, it was proved that a time-invariant chain $\{A\}$ is ergodic if its associated graph has a spanning rooted tree with a root having a self loop [55, 80]. Such results significantly reduce the complexity of dealing with the dynamics (2.1). Similarity, in [69], the *infinite flow graph* of a discrete-time chain (process) is defined and its graph theoretic properties were used to analyze the ergodicity of such chains.

Definition 2. *(Directed Infinite Flow Property) For a stochastic chain $\{A(t)\}$, we define its* *directed infinite flow graph $G^{\infty} = ([n], E^{\infty})$ to be the graph with the edge set*

$$E^{\infty} = \left\{ (j,i) \middle| i,j \in [n], \sum_{\tau=t_0}^{\infty} a_{ij}(\tau) = \infty \right\}.$$

Note that the associated graph and infinite flow graph of a time-invariant chain are the same. As a result, from the discussion above, for time-invariant chains the ergodicity of (2.1) and the existence of spanning rooted tree in G^{∞} is equivalent.

Another important notion to study ergodicity is a class of stochastic chains, namely the class $\mathcal{P}^*$ which was first introduced in [71] for discrete-time distributed averaging dynamics.

Definition 3. *(Class $\mathcal{P}^*$) A stochastic chain $\{A(t)\}$ is in the class $\mathcal{P}^*$ if $e^T \Phi(t,\tau) \geq pe^T$ for some $p > 0$ and all $t \geq \tau \geq t_0$.*

2.2 Motivations and Main Results

The main results of this chapter are a necessary condition and a sufficient condition for ergodicity. In this section, we provide the main theorems and what motivates us to investigate those results. The proofs of those results will be presented in the subsequent sections.

2.2.1 Necessary Condition

As mentioned, for time-invariant chains the ergodicity of (2.1) and the existence of spanning rooted tree in G^{∞} is equivalent. One might be tempted to generalize this result to time-varying chains, i.e., $A(t)$ is ergodic if and only if the directed infinite flow graph of $A(t)$ has directed spanning rooted tree. However, such a result does not hold in general as discussed in the following example.

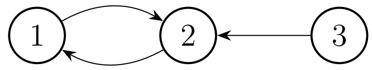

Figure 2.1: The infinite flow graph of the chain in Example 1

Example 1. *Consider the stochastic matrices*

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix},$$

and the stochastic chain

$$A(t) = \begin{cases} A_1, & \sigma_k \leq t < \sigma_k + \tau_k \\ A_2, & \sigma_k + \tau_k \leq t < \sigma_k + 2\tau_k \\ A_3, & \sigma_k + 2\tau_k \leq t < \sigma_k + 3\tau_k = \sigma_{k+1} \end{cases},$$

where $\tau_k = \left\lceil \log_2 \frac{x_1(\sigma_k) - x_2(\sigma_k)}{x_1(\sigma_k)} \right\rceil$*, and* $\sigma_k = 1 + 3\sum_{\ell=1}^{k-1} \tau_\ell$ *for* $k > 1$ *with* $\sigma_1 = 1$*. Let* $t_0 = 1$ *and* $x(1) = [-2, 1, 2]^T$*. For* $\sigma_k < t \leq \sigma_k + \tau_k$*, we have*

$$x_2(t) = (x_2(\sigma_k) - x_1(\sigma_k))2^{-t+\sigma_k} + x_1(\sigma_k). \tag{2.2}$$

Note that $x_3(t) = x_3(1) = 2$ *for all* $t \geq 1$*. Initially,* $x_2(\cdot)$ *moves toward* $x_1(\cdot)$ *and due to* (2.2)*, stops somewhere less than or equal to* 0 *at time* $\tau_1 + 1$*, and then* $x_1(\cdot)$ *moves toward* $x_2(\cdot)$*. Therefore, we have* $x_1(2\tau_1 + 1) < x_2(2\tau_1 + 1) \leq 0$*. The same argument holds for the behavior of the dynamics in the subsequent intervals* $[\sigma_k, \sigma_{k+1})$ *for* $k = 2, 3, \ldots$*. Therefore, this chain is not ergodic as* $x_1(t) \leq 0$ *and* $x_3(t) = 2$ *for all* $t \geq 1$*. Since the dynamics ends when* $x_1(t) = x_2(t)$ *for some* $t \geq 1$*, it does not end in finite time. Thus,* $\sum_{\tau=1}^{\infty} a_{12}(\tau) = \sum_{\tau=1}^{\infty} a_{21}(\tau) = \sum_{\tau=1}^{\infty} a_{32}(\tau) = \frac{\sigma_\infty}{3} = \infty$*, and the infinite flow graph has a spanning rooted tree as shown in Fig. 2.1.*

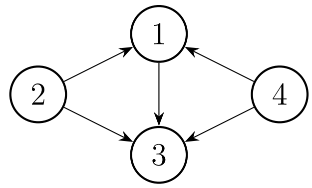

Figure 2.2: A graph that is connected as an undirected graph but does not have a spanning rooted tree

Although, the existence of a spanning rooted tree in the infinite flow graph is not equivalent to ergodicity, in [8], for continuous processes, it is shown that it is still necessary for ergodicity of time-varying processes, which is implied from Theorem 3 and 4 of [8]. In the following theorem, we extend this result for discrete chains.

Theorem 1. *(Necessary Condition for Ergodicity) If $\{A(t)\}$ is ergodic then, its infinite flow graph G^∞ has a spanning rooted tree.*

This result strengthens the existing necessary condition for ergodicity as discussed in [69, 72]. There, it is shown that the connectivity of the *undirected* infinite flow graph is necessary for ergodicity, where the undirected infinite flow graph is obtained by removing the directions on the edges of the directed infinite flow graph. Note that graphs such as the graph shown in Fig. 2.2 do not have a spanning rooted tree but the associated undirected graph is connected.

2.2.2 Sufficient Condition

In [71] and [6], it was shown that for processes or chains that are in class $\mathcal{P}^*$ having a connected undirected infinite flow graph plus some mild additional conditions (namely, weakly periodic for discrete-time chains and boundedness for continuous-time processes) is sufficient for ergodicity for discrete-time and continuous-time dynamics, respectively. However, there are many chains that are not in class $\mathcal{P}^*$ and they are ergodic as discussed in the following example.

Example 2. *Consider the stochastic chain*

$$A(t) = \begin{bmatrix} 1-a(t) & a(t) \\ 0 & 1 \end{bmatrix},$$

where $1 \geq a(t) \geq 0$ *such that* $\sum_{s=t_0}^{\infty} a(s) = \infty$. *Then, we have*

$$\Phi(t,t_0) = \begin{bmatrix} \phi_{11}(t,t_0) & 1-\phi_{11}(t,t_0) \\ 0 & 1 \end{bmatrix},$$

where $\phi_{11}(t,t_0) = \prod_{s=t_0}^{t}(1-a(s))$. *Therefore, we have*

$$\lim_{t\to\infty} \Phi(t,\tau) = e \begin{bmatrix} 0 & 1 \end{bmatrix},$$

and hence, $A(t)$ *is ergodic but not in* $\mathcal{P}^*$.

Intuitively, for class $\mathcal{P}^*$ chains, the initial condition of each agent (entry) contributes to the steady state value of the dynamics (2.1). In this chapter, we deal with the chains that some of their coordinates does not contribute to the steady state, such as the process discussed in Example 2. We consider a general class of chains that are driven by class $\mathcal{P}^*$ chains that are interacting with each other through an underlying directed acyclic graph containing a spanning directed

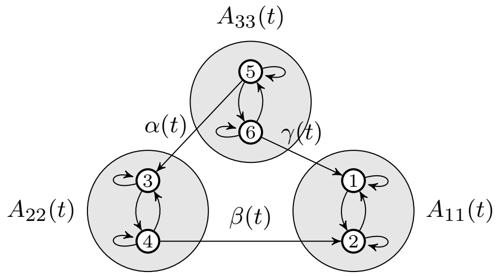

Figure 2.3: A chain with 6 agents which can be viewed as a chain with 3 sub-chains $A_{11}(t)$, $A_{22}(t)$, and $A_{33}(t)$

rooted tree, such as the chain

$$\mathcal{A}(t) = \left[\begin{array}{cc|cc|cc} \multicolumn{2}{c|}{\multirow{2}{*}{$A_{11}(t)$}} & 0 & 0 & 0 & \gamma(t) \\ & & 0 & \beta(t) & 0 & 0 \\ \hline 0 & 0 & \multicolumn{2}{c|}{\multirow{2}{*}{$A_{22}(t)$}} & \alpha(t) & 0 \\ 0 & 0 & & & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \multicolumn{2}{c}{\multirow{2}{*}{$A_{33}(t)$}} \\ 0 & 0 & 0 & 0 & & \end{array}\right],$$

which is shown in Fig. 2.3. As we can see, there are three *clusters/sub-graphs* of agents, each of them contains two agents, and the underlying graph connecting each sub-graph (inside gray circles) is a directed acyclic graph containing a spanning directed rooted tree, while sub-graphs are in class $\mathcal{P}^*$.

Definition 4. *For a matrix $A = [a_{ij}]_{n\times n}$ with non-negative elements and $\sum_{j=1}^{n} a_{ij} \le 1$ for $i \in [n]$, we define the stochastic matrix $\mathcal{S}[A] \triangleq [b_{ij}]_{n\times n}$ as follows:*

$$b_{ij} \triangleq \begin{cases} a_{ij}, & i \ne j \\ 1 - \sum_{j\in[n]\setminus\{i\}} a_{ij}, & i = j \end{cases}.$$

Theorem 2. *(Sufficient Condition for Ergodicity) Consider a stochastic chain $B(t) = [A_{ij}(t)]_{m\times m}$,* *where for each $i \in [m]$, $\mathcal{S}[A_{ii}(t)]$ is in $\mathcal{P}^*$ and ergodic, and for $i > j$, $\|A_{ij}(t)\|_\infty$ is summable. The stochastic chain $\{B(t)\}$ is ergodic, if its infinite flow graph has a spanning rooted tree.*

2.3 Necessary Condition

In this section, we prove the necessary condition outlined in Theorem 1. To do so, we need to remove some unnecessary structures of a chain and approximate it with a simpler chain. In this regard, we use the ℓ_1-approximation concept and the lemma about it, which were introduced and proved in [68].

Definition 5. *We define $B(t)$ as an ℓ_1-approximation of $A(t)$ if $\|B(t) - A(t)\|_\infty$ is summable. Also, we say $B(t) = [b_{ij}(t)]_{n\times n}$ is minimal ℓ_1-approximation of $A(t)$ with the infinite flow graph $([n], E^\infty)$ if $B(t)$ is stochastic for $t \geq t_0$ and*

$$b_{ij}(t) = \begin{cases} a_{ij}(t), & \text{if } (j,i) \in E^\infty \\ 0, & \text{if } i \neq j \in [n], (j,i) \notin E^\infty \end{cases}.$$

Lemma 1 ([68])**.** *(Approximation Lemma) Let stochastic chains $\{A(t)\}$ and $\{B(t)\}$ be an ℓ_1-approximation of each other. Then, $\{A(t)\}$ is ergodic if and only if $\{B(t)\}$ is ergodic.*

Proof of Theorem 1: We prove that if the infinite flow graph of $\{A(t)\}$ does not have a spanning rooted tree, then it is not ergodic. Let $B(t) = [b_{ij}(t)]$ be the minimal ℓ_1-approximation of $\{A(t)\}$. Note that the infinite flow graphs of $\{A(t)\}$ and $\{B(t)\}$ are the same. Hence, it is enough to prove that $\{B(t)\}$ is not ergodic because by Lemma 1, this would imply that $\{A(t)\}$ is not ergodic.

So, assume that $\{B(t)\}$ does not have spanning rooted tree. Let $z(t)$ be a dynamics driven by $B(t)$. Consider a rooted tree $\mathcal{T}$ with the maximum number of nodes. Suppose that the vertex set of this tree is $S = \{i_1, i_2, \ldots, i_m\}$ with the root i_1. Because this tree is not spanning, we have

$m < n$, and there are nodes $\bar{S} = \{i_{m+1}, \ldots, i_n\}$ that are not in the tree. Because of the maximal assumption on the size of this tree, we cannot expand this tree and hence, there is no edge from S to $\bar{S}$, and we can write

$$z_i(t+1) = \sum_{j \in \bar{S}} b_{ij}(t) z_j(t),$$

for $i \in \bar{S}$ and $t > t_0$. This itself implies that if we set $z_i(t_0) = 0$ for $i \in \bar{S}$, we will have $z_i(t) = 0$ for all $i \in \bar{S}$ and $t \geq t_0$. Now let $D \subseteq [n]$ be all nodes such that there is a path from them to the root i_1. First notice that $D \cap \bar{S} = \varnothing$. This is because if $i \in D \cap \bar{S}$, then there is a path from i to every node in S. So, we have a rooted tree with root i and $m+1$ nodes which is a contradiction with the assumption that $\mathcal{T}$ was maximal. By the same argument, there is no edge from $\bar{S}$ to D. Also, there is no edge from $S \setminus D$ to D, otherwise we can expand D. Therefore, there is no edge from outside of D to D. Hence, similar to the nodes $i \in S$, if we set $z_j(t_0) = 1$ for $j \in D$, we have $z_j(t) = 1$ for $j \in D$ and $t \geq t_0$, implying that $\{B(t)\}$ (and hence, $\{A(t)\}$) is not ergodic. ■

2.4 Sufficient Condition

In this section, we prove the sufficient condition discussed in Theorem 2. To do so, we need to work on a more general form of the dynamics (2.1).

A dynamics $x : \{t_0, t_0+1, \ldots\} \to \mathbb{R}^n$ driven by a sub-stochastic chain $\{A(t)\}$ with the perturbation $\lambda(t)$ and the initial condition $x(t_0)$ is a dynamics satisfying

$$x(t+1) = A(t)x(t) + \lambda(t), \qquad \text{for } t \geq t_0. \tag{2.3}$$

From (2.3), it follows that

$$x(t) = \Phi(t,t_0)x(t_0) + \sum_{s=t_0+1}^{t} \Phi(t,s)\lambda(s), \tag{2.4}$$

where $\Phi(t,\tau) = A(t-1)\cdots A(\tau)$ is the transition matrix from time τ to t. In the following, i.e., Lemma 2 and Remark 1 and 2, we study (2.3) where $\lambda(t) = B(t)y(t)$ and $[A(t), B(t)]$ is stochastic.

Lemma 2. *Let $x(t)$ be defined by*

$$x(t+1) = A(t)x(t) + B(t)y(t),$$

with the initial condition $x(t_0) \in \mathbb{R}^n$ where $\{[A(t), B(t)]\}$ is a stochastic chain. Then, we have

$$\max_i x_i(t) \leq \max\left\{\max_i x_i(t_0), \max_{t>\tau\geq t_0} \max_i y_i(\tau)\right\},$$
$$\min_i x_i(t) \geq \min\left\{\min_i x_i(t_0), \min_{t>\tau\geq t_0} \min_i y_i(\tau)\right\}.$$

Proof: We prove by induction on t. The lemma is true for $t = t_0$; hence, suppose that it is true for t. Suppose that $B(t)$ is $n \times m$ for $m \geq 1$. Since, $A(t)$ and $B(t)$ are non-negative, we have:

$$\begin{aligned} x_i(t+1) &= \sum_{j=1}^{n} a_{ij}(t)x_j(t) + \sum_{j=1}^{m} b_{ij}(t)y_j(t) \\ &\leq \max\left\{\max_{i\in[n]} x_i(t), \max_{i\in[m]} y_i(t)\right\}\left(\sum_{j=1}^{n} a_{ij}(t) + \sum_{j=1}^{m} b_{ij}(t)\right) \\ &= \max\left\{\max_{i\in[n]} x_i(t), \max_{i\in[m]} y_i(t)\right\}, \end{aligned}$$

where the last equality follows from $[A(t), B(t)]$ being stochastic. Therefore,

$$\begin{aligned}\max_i x_i(t+1) &\leq \max\left\{\max_{i\in[n]} x_i(t), \max_{i\in[m]} y_i(t)\right\}\\ &\overset{(a)}{\leq} \max\left\{\max_i x_i(t_0), \max_{t>\tau\geq t_0}\max_i y_i(\tau), \max_i y_i(t)\right\}\\ &= \max\left\{\max_i x_i(t_0), \max_{t+1>\tau\geq t_0}\max_i y_i(\tau)\right\},\end{aligned}$$

where (a) follows from the induction hypothesis. The proof of the lower bound follows from a similar argument. ■

Remark 1. *Note that if $x^{(i)}(t)$ is driven by $\{A(t)\}$ with the perturbation $B(t)y^{(i)}(t)$ and the initial condition $x^{(i)}(t_0)$, then, $\sum_i \mu_i x^{(i)}(t)$ is driven by $\{A(t)\}$ with the perturbation $B(t)\sum_i \mu_i y^{(i)}(t)$ and the initial condition $\sum_i \mu_i x^{(i)}(t_0)$.*

Remark 2. *Let*

$$\Psi(t,\tau) \triangleq \left[\Phi(t,\tau), \sum_{s=t_0+1}^{t} \Phi(t,s)B(s)\right], \textit{ and } z(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}.$$

Assume $y(t)$ is constant, i.e., $y(t) = y_0$ for some $y_0 \in \mathbb{R}^m$. Then, we have

1. *$x(t) = \Psi(t,\tau)z(\tau)$ which is implied by* (2.4)*.*

2. *$\Psi(t,\tau)$ is stochastic as $[A(t), B(t)]$ is stochastic.*

From classical results in the study of inhomogeneous Markov chains [61], we know that the stochastic chain $\{A(t)\}$ (with dynamics (2.1)) is ergodic if and only if there exists $q > 0$ such that for all $\tau \geq t_0$

$$\liminf_{t\to\infty} \Phi(t,\tau)e_{k(\tau)} \geq qe \tag{2.5}$$

for some $k(\tau) \in [n]$. Motivated by this we have the following definition.

Definition 6. *For a stochastic chain $\{A(t)\}$ with the transition matrix $\Phi(t,\tau)$, we say that the kth coordinate is q^*, if there exists $q > 0$ such that $\liminf_{t\to\infty} \Phi(t,\tau)e_k \geq qe$ for some $t_q \geq t_0$, and all $\tau \geq t_q$.*

Note that by (2.5), if one of the coordinates of $\{A(t)\}$ is q^*, then $\{A(t)\}$ is ergodic. Also, by definition $\mathcal{P}^*$, if $A(t)$ is in $\mathcal{P}^*$ and ergodic, then all its coordinates are q^*. The following lemma plays a key role in the proof of Theorem 2.

Lemma 3. *Let $x(t)$ be defined by*

$$x(t+1) = A(t)x(t) + B(t)\bar{z}e,$$

where $\{[A(t), B(t)]\}$ is a stochastic chain. Suppose that the kth coordinate of $\mathcal{S}[A(t)]$ is q^* and

$$\sum_{\tau=t_0}^{\infty} e_k^T B(t)e = \infty.$$

Then, we have $\lim_{t\to\infty} x(t) = \bar{z}e$ for all $x(t_0) \in \mathbb{R}^n$.

Proof: To prove the lemma, according to Remark 2, it is enough to prove that if $\bar{z} = 1$, then $\lim_{t\to\infty} x(t) = e$ for all initial conditions $x^{(\ell)}(t_0)$ for $0 \leq \ell \leq n$ where $x^{(0)}(t_0) = 0e$ and $x^{(\ell)}(t_0) = \sum_{i=\ell}^n e_i$ for $\ell \in [n]$. Let $\alpha(t) = B(t)e$ and $u_i(t) = \alpha_i(t)(1 - x_i(t))$ for $i \in [n]$. Note that from Lemma 2, $x(t_0) \leq e$ and $\alpha_1(t) \leq 1$ imply that $x(t) \leq e$ for $t \geq t_0$, and hence, $u_i(t) \geq 0$ for $i \in [n]$ and $t \geq t_0$. Without loss of generality suppose that $k = 1$.

For now, assume that $b_{ij}(t) = 0$ for $i > 1$ and all $t \geq t_0$. Let $\Phi(t,\tau)$ be the transition matrix associated with the chain $\mathcal{S}[A(t)]$. We can write the dynamics as:

$$x(t+1) = \mathcal{S}[A(t)]x(t) + u_1(t)e_1. \tag{2.6}$$

Therefore, from (2.4), any solution $x(t)$ to (2.6), satisfies

$$\begin{aligned} x(t) &= \Phi(t,t_0)x(t_0) + \sum_{\tau=t_0+1}^{t} \Phi(t,\tau)u_1(\tau)e_1 \\ &= \Phi(t,t_0)x(t_0) + \sum_{\tau=t_0+1}^{\infty} 1_{(t_0,t]}u_1(\tau)\Phi(t,\tau)e_1. \end{aligned} \tag{2.7}$$

Thus, we can write

$$\begin{aligned} \liminf_{t\to\infty} x(t) &\geq \liminf_{t\to\infty} \Phi(t,t_0)x(t_0) + \liminf_{t\to\infty} \sum_{\tau=t_0+1}^{\infty} 1_{(t_0,t]}u_1(\tau)\Phi(t,\tau)e_1 \\ &\stackrel{(a)}{\geq} \liminf_{t\to\infty} \sum_{\tau=t_0+1}^{\infty} 1_{(t_0,t]}u_1(\tau)\Phi(t,\tau)e_1 \\ &\stackrel{(b)}{\geq} \sum_{\tau=t_0+1}^{\infty} \liminf_{t\to\infty} 1_{(t_0,t]}u_1(\tau)\Phi(t,\tau)e_1 \\ &\stackrel{(c)}{\geq} \sum_{\tau=t_q}^{\infty} u_1(\tau)qe, \end{aligned} \tag{2.8}$$

where (a) follows from the fact that $\Phi(t,t_0)$ is a non-negative matrix and $x(t_0) \geq 0e$, (b) follows from Fatou's Lemma (Theorem 2.18 [21]) and the fact that $1_{(t_0,t]}u_1(\tau)\Phi(t,\tau)e_1$ is non-negative, and (c) holds as the first coordinate is q^*, and hence, there exists $q > 0$ such that $\lim_{t\to\infty} \Phi(t,\tau)e_1 \geq qe$ for all $\tau \geq t_q$. Since, $\liminf_{t\to\infty} x(t)$ is bounded, and $q > 0$, (2.8) implies that

$$\sum_{\tau=t_0+1}^{\infty} u_1(\tau) < \infty. \tag{2.9}$$

On the other hand, since $\Phi(t,\tau) \leq ee^T$, we have

$$\sum_{\tau=t_0+1}^{\infty} u_1(\tau)e \geq \sum_{\tau=t_0+1}^{\infty} 1_{(t_0,t]}u_1(\tau)\Phi(t,\tau)e_1.$$

Hence, using the Dominated Convergence Theorem (Theorem 2.24 [21]), we have

$$\begin{aligned}\lim_{t\to\infty}\sum_{\tau=t_0+1}^{\infty} 1_{(t_0,t]}u_1(\tau)\Phi(t,\tau)e_1 &= \sum_{\tau=t_0+1}^{\infty}\lim_{t\to\infty} 1_{(t_0,t]}u_1(\tau)\Phi(t,\tau)e_1\\ &= \sum_{\tau=t_0+1}^{\infty} u_1(\tau)\pi_1(\tau)e.\end{aligned}$$

This and (2.7) imply that there exists $\bar{x}\in[0,1]$ such that $\lim_{t\to\infty}x(t)=\bar{x}e$. Note that $\bar{x}<1$ contradicts (2.9) and hence, $\bar{x}=1$ which concludes the proof for the case $b_{ij}(t)=0$ for $i>1$ and all $t\geq t_0$. In other words, so far, we have proved that

$$\lim_{t\to\infty}\Psi(t,t_0)\begin{bmatrix}x(t_0)\\1\end{bmatrix}=e, \tag{2.10}$$

where $\Psi(t,\tau)$ is the transition matrix of the dynamics

$$x(t+1)=\mathcal{S}[[A(t),\alpha_1(t)e_1]]\begin{bmatrix}x(t)\\1\end{bmatrix}.$$

In the following, we prove the lemma for the general case. We can write the dynamics as follows

$$x(t+1)=\mathcal{S}[[A(t),\alpha_1(t)e_1]]\begin{bmatrix}x(t)\\1\end{bmatrix}+\lambda(t),$$

where $\lambda(t) = \sum_{i=2}^{n} u_i(t)e_i$. From (2.4), we have

$$\lim_{t\to\infty} x(t) = \lim_{t\to\infty} \Psi(t,t_0) \begin{bmatrix} x(t_0) \\ 1 \end{bmatrix} + \sum_{\tau=t_0+1}^{t} \Phi(t,\tau)\lambda(\tau)$$
$$\geq \lim_{t\to\infty} \Psi(t,t_0) \begin{bmatrix} x(t_0) \\ 1 \end{bmatrix},$$

which follows as $\Phi(t,\tau)$ and $\lambda(t)$ are non-negative. Finally, $x(t) \leq e$ and (2.10) complete the proof. ■

The main theorem is proved using the following lemma.

Lemma 4. *Consider a stochastic chain*

$$C(t) = \begin{bmatrix} A(t) & B(t) \\ 0ee^T & \hat{A}(t) \end{bmatrix}.$$

where $\mathcal{S}[A(t)]$ is in $\mathcal{P}^*$ and ergodic, and $\hat{A}(t)$ is ergodic. The stochastic chain $C(t)$ is ergodic if its infinite flow graph has a spanning rooted tree.

Proof: Let $z(t)$ be driven by $\{C(t)\}$ with the initial condition $z(t_0)$, and size $A(t)$ and $C(t)$ be $n \times n$ and $N \times N$, respectively. Let $x(t) \triangleq [z_1(t), \ldots, z_n(t)]$ and $y(t) \triangleq [z_{N-n+1}, \ldots, z_N(t)]$. Because $x(t)$ does not contribute to $y(t)$, and $\hat{A}(t)$ is ergodic, we have $\lim_{t\to\infty} y(t) = \bar{z}e$ for some $\bar{z}$. Set $s_\epsilon \geq t_0$ such that

$$\sup_{\tau\in[s_\epsilon,\infty)} \|y(\tau) - \bar{z}e\|_\infty < \frac{\epsilon}{2}. \tag{2.11}$$

Let $w(t)$ be defined by

$$w(t+1) = A(t)w(t) + B(t)\bar{z}e, \tag{2.12}$$

with the initial time s_ϵ and the initial condition $w(s_\epsilon) = x(s_\epsilon)$. Because $\mathcal{S}[A(t)]$ is in $\mathcal{P}^*$ and ergodic, and hence, all its coordinates are q^*, Lemma 3 implies $\lim_{t\to\infty} w(t) = \bar{z}e$. Set $t_\epsilon \geq s_\epsilon$ such that

$$\sup_{\tau\in[t_\epsilon,\infty)} \|w(t) - \bar{z}e\|_\infty < \frac{\epsilon}{2}. \tag{2.13}$$

Let $v(t)$ be driven by

$$v(t+1) = A(t)v(t) + B(t)(y(t) - \bar{z}e) \tag{2.14}$$

with the initial condition $v(s_\epsilon) = 0e$. Because the block $[A(t), B(t)]$ is stochastic, Lemma 2 implies

$$\|v(t)\|_\infty \leq \max_{s_\epsilon\leq\tau<t} \|y(\tau) - \bar{z}e\|_\infty < \frac{\epsilon}{2}.$$

Adding (2.12) and (2.14), we have

$$w(t+1) + v(t+1) = A(t)\left[w(t) + v(t)\right] + B(t)y(t).$$

Also, we have $x(s_\epsilon) = w(s_\epsilon) + v(s_\epsilon)$. Therefore, we have $x(t) = w(t) + v(t)$ for $s_\epsilon \leq t$, and hence, for $t_\epsilon \leq t$

$$\begin{aligned}\|x(t)-\bar{z}e\|_\infty &= \|w(t)-\bar{z}e+v(t)\|_\infty \\ &\leq \|w(t)-\bar{z}e\|_\infty + \|v(t)\|_\infty \\ &\leq \frac{\epsilon}{2}+\frac{\epsilon}{2} = \epsilon,\end{aligned}$$

which follows from (2.11) and (2.13). This is true for every $\epsilon > 0$, and hence, we have $\lim_{t\to\infty} x(t) = \bar{z}$, and the proof is complete. ■

Proof of Theorem 2: From Lemma 1, without loss of generality, we can assume that for $i > j$, $A_{ij}(t)$ is a zero matrix for $t \geq t_0$. We use induction on m to prove the theorem. The result is true for $m = 1$. Assume theorem is true for $m-1$. Therefore, $\hat{B}(t) \triangleq [\hat{A}_{ij}(t)]_{(m-1)\times(m-1)}$ where $\hat{A}_{ij} = A_{i+1j+1}(t)$ is ergodic. Now, we have

$$B(t) = \begin{bmatrix} A_{11}(t) & C(t) \\ 0ee^T & \hat{B}(t) \end{bmatrix},$$

where $C(t) = [A_{1j}(t)]_{1\times m}$. Note that $\hat{B}(t)$ is ergodic and $\mathcal{S}[A_{11}(t)]$ is in $\mathcal{P}^*$ and ergodic, and hence, by Lemma 2.4, we conclude the proof. ■

Chapter 2, in full, is a reprint of the material as it appears in A. Aghajan and B. Touri, On ergodicity of time-varying distributed averaging dynamics, *2020 American Control Conference (ACC)*, IEEE, 2020, pp. 4417–4422. The dissertation author was the primary investigator and author of this paper.

Chapter 3

Continuous-Time Distributed Averaging Dynamics

In this chapter, we extend the results for discrete-time distributed averaging dynamics, which are derived in Chapter 2, to continuous-time distributed averaging dynamics. However, The logic behind the proof of most lemmas in this chapter are similar to the proof of the corresponding lemma in discrete-time setting, we prove them separately in this chapter for completeness.

The structure of this chapter is as follows: In Section 3.1, we formulate the problem of interest and state the main results of this work. In Section 3.2, we discuss the broad class of processes that satisfy the proposed sufficient condition. In Section 3.3, we prove the main results from Section 3.1, and finally, in Section 3.4, we provide the proof of the lemmas related from Section 3.2 and some other lemmas from Section 3.3.

3.1 Problem Statement and Main result

In this section, we discuss the problem statement and the main result of this work. The proof of the result will be presented in the subsequent sections.

We study the linear time-varying dynamics

$$x(t) = x(t_0) + \int_{t_0}^{t} A(\tau)x(\tau)\, d\tau, \qquad \text{for } t \geq t_0, \tag{3.1}$$

driven by a Laplacian process $\{A(t) = [a_{ij}(t)]_{n\times n}\}$ with an initial condition $x(t_0) \in \mathbb{R}^n$. We refer to the solution $x(t)$ of (3.1) as the dynamics driven by $\{A(t)\}$. Because $A(t)$ is a Laplacian matrix, we can write the above system as

$$x_i(t) = x(t_0) + \int_{t_0}^{t} \sum_{j=1}^{n} a_{ij}(\tau)(x_j(\tau) - x_i(\tau))\, d\tau \tag{3.2}$$

for $i \in [n]$. In this chapter, we assume that for all $i, j \in [n]$, $a_{ij}(t)$ is a measurable function of time and $\int_{\tau}^{t} a_{ij}(s)\, ds < \infty$ for all $t_0 \leq \tau \leq t$, implying that (3.1) has a unique continuous solution [62]. Note that we can write $x(t) = \Phi(t,\tau)x(\tau)$ where $\Phi(t,\tau)$ is the transition matrix for (3.1) for all $t_0 \leq \tau \leq t$. For Laplacian processes, from Lemma 6 in [37], we know that $\Phi(t,\tau)$ is a stochastic matrix for all $t_0 \leq \tau \leq t$.

Definition 7 (Ergodicity)**.** *Let $x(t)$ be the dynamics driven by $\{A(t)\}$. We say that a process $\{A(t)\}$ is ergodic if for all initial time $t_0 \in \mathbb{R}$ and all choices of initial condition $x(t_0) \in \mathbb{R}^n$, $\lim_{t\to\infty} x(t) = \bar{x}e$ for some $\bar{x} \in \mathbb{R}$ which depends on t_0 and $x(t_0)$.*

Since $x(t) = \Phi(t,\tau)x(\tau)$, if $\{A(t)\}$ is ergodic, then for all $i \in [n]$, we have

$$\lim_{t\to\infty} \Phi(t,\tau)e_i = e\pi_i(\tau),$$

for some $\pi_i(\tau) \in \mathbb{R}$, and hence $\lim_{t\to\infty} \Phi(t,\tau) = e\pi^T(\tau)$. On the other hand, if $\lim_{t\to\infty} \Phi(t,\tau) = e\pi^T(\tau)$, then

$$\lim_{t\to\infty} \Phi(t,\tau)x(\tau) = e\pi^T(\tau)x(\tau) = \bar{x}e,$$

where $\bar{x} = \pi^T(\tau)x(\tau)$. Therefore, an equivalent characterization of ergodicity using the transition matrices $\Phi(t,\tau)$ is that (3.1) is ergodic if and only if $\lim_{t\to\infty}\Phi(t,\tau)=e\pi^T(\tau)$ for a stochastic vector $\pi(\tau)\in\mathbb{R}^n$ and all $\tau\geq t_0$.

As mentioned in Chapter 2, in [69], the *infinite flow graph* of a discrete-time chain is defined and its graph theoretic properties were used to analyze the ergodicity of discrete-time chains. Similar structures and results were developed for the continuous-time setting [25, 6].

Definition 8. *(Directed Infinite Flow Graph) For a Laplacian process $\{A(t)\}$, we define its directed infinite flow graph $G^\infty=([n],E^\infty)$ to be the graph with the edge set*

$$E^\infty=\left\{(j,i)\middle| i,j\in[n],\int_s^\infty a_{ij}(\tau)\,d\tau=\infty \quad \forall s\geq t_0\right\}.$$

Unless stated otherwise, we simply refer to this graph as the infinite flow graph of $\{A(t)\}$. Again, similar to the discrete-time distributed averaging dynamics, One might think $A(t)$ is ergodic if and only if the directed infinite flow graph of $A(t)$ has a spanning directed rooted tree. However, such a result does not hold in general as discussed in the following example, which is the extension of Example 1 to the continuous-time setting.

Example 3. *Consider the Laplacian matrices*

$$A_1=\begin{bmatrix}0&0&0\\1&-1&0\\0&0&0\end{bmatrix},A_2=\begin{bmatrix}-1&1&0\\0&0&0\\0&0&0\end{bmatrix},A_3=\begin{bmatrix}0&0&0\\0&-1&1\\0&0&0\end{bmatrix},$$

and the Laplacian process

$$A(t)=\begin{cases}A_1, & \sigma_k\leq t<\sigma_k+\tau_k\\ A_2, & \sigma_k+\tau_k\leq t<\sigma_k+2\tau_k\\ A_3, & \sigma_k+2\tau_k\leq t<\sigma_k+3\tau_k=\sigma_{k+1}\end{cases},$$

where

$$\tau_k = \max\left\{\ln\frac{x_1(\sigma_k)-x_2(\sigma_k)}{x_1(\sigma_k)}, 1\right\},$$

and $\sigma_k = 3\sum_{\ell=1}^{k-1}\tau_\ell$ *for* $k > 1$ *with* $\sigma_1 = 0$. *Let* $t_0 = 0$ *and* $x(0) = [-2,1,2]^T$. *For* $\sigma_k \le t < \sigma_k + \tau_k$, *we have*

$$x_2(t) = (x_2(\sigma_k) - x_1(\sigma_k))\exp(-t+\sigma_k) + x_1(\sigma_k). \tag{3.3}$$

Note that $x_3(t) = x_3(0) = 2$ *for all* $t \ge 0$. *Initially,* $x_2(\cdot)$ *moves toward* $x_1(\cdot)$ *and due to* (3.3), *stops somewhere less than or equal to* 0 *at time* τ_1, *and then* $x_1(\cdot)$ *moves toward* $x_2(\cdot)$. *Therefore, we have* $x_1(2\tau_1) < x_2(2\tau_1) \le 0$. *The same argument holds for the behavior of the dynamics in the subsequent intervals* $[\sigma_k, \sigma_{k+1})$ *for* $k = 2,3,\ldots$. *Therefore, this process is not ergodic as* $x_1(t) \le 0$ *and* $x_3(t) = 2$ *for all* $t \ge 0$. *Since the dynamics ends when* $x_1(t) = x_2(t)$ *for some* $t \ge 0$, *it does not end in finite time. Thus,*

$$\int_{\sigma_k}^{\infty} a_{12}(\tau)d\tau = \int_{\sigma_k}^{\infty} a_{21}(\tau)d\tau = \int_{\sigma_k}^{\infty} a_{32}(\tau)d\tau = \frac{\sigma_\infty - \sigma_k}{3} = \infty,$$

for $k \in \mathbb{N}$, *and the infinite flow graph has a spanning directed rooted tree as shown in Fig. 2.1.*

Although, the existence of a spanning directed rooted tree in the infinite flow graph is not equivalent to ergodicity, in [8], it is shown that it is still necessary for ergodicity of time-varying processes. More specifically, from Theorems 3 and 4 of [8], we can imply if $\{A(t)\}$ is ergodic, then its infinite flow graph G^∞ has a spanning directed rooted tree.

To present the main result of this chapter, we define an important class of Laplacian processes, namely the class $\mathcal{P}^*$, which is first introduced in [71] for discrete-time chains and were studied in [6] for continuous-time dynamics.

Definition 9 (*Class* $\mathcal{P}^*$)**.** *A Laplacian process* $\{A(t)\}$ *is said to be in class* $\mathcal{P}^*$ *if for some* $p > 0$, *we have* $e^T \Phi(t,\tau) \geq pe^T$ *for all* $t_0 \leq \tau \leq t$.

Despite the fact that for processes that are in class $\mathcal{P}^*$, having a connected undirected infinite flow graphs plus a mild additional condition is sufficient for ergodicity for continuous-time dynamics [6], there are many processes that are not in class $\mathcal{P}^*$ and they are ergodic as discussed in the following example, which is extension of Example 2 to the continuous-time setting.

Example 4. *Consider the Laplacian process*

$$A(t) = \begin{bmatrix} -a(t) & a(t) \\ 0 & 0 \end{bmatrix},$$

where $a(t) \geq 0$ *such that* $\int_{t_0}^{\infty} a(s)\, ds = \infty$. *Then, we have*

$$\Phi(t,\tau) = \begin{bmatrix} \phi_{11}(t,\tau) & 1 - \phi_{11}(t,\tau) \\ 0 & 1 \end{bmatrix},$$

where $\phi_{11}(t,\tau) = \exp\left(-\int_{\tau}^{t} a(s)\, ds\right)$. *Therefore, we have* $\lim_{t\to\infty} \Phi(t,\tau) = e\begin{bmatrix} 0 & 1 \end{bmatrix}$ *for all* τ, *and hence,* $A(t)$ *is ergodic but not in class* $\mathcal{P}^*$.

Similar to the discrete-time setting, in this chapter, we deal with the processes that some of their coordinates does not contribute to the steady state, such as the process discussed in Example 4. We consider a general class of processes that are driven by class $\mathcal{P}^*$ processes that are interacting with each other through an underlying directed acyclic graph containing a spanning directed rooted tree, such as the processes that is shown in Fig. 2.3.

Definition 10. *For a matrix* $A = [a_{ij}]_{m\times n}$ *with* $a_{ij} \geq 0$ *for all* $i \neq j$, *we define the Laplacian* *matrix* $\mathcal{L}[A] \triangleq [c_{ij}]_{m\times n}$ *by:*

$$c_{ij} \triangleq \begin{cases} a_{ij}, & i \neq j \\ -\sum_{j\in[n]\setminus\{i\}} a_{ij}, & i = j \end{cases}.$$

Theorem 3. *(Sufficient Condition for Ergodicity) Consider a Laplacian process*

$$\mathcal{A}(t) = [A_{ij}(t)]_{m\times m},$$

where for each $i \in [m]$*,* $\mathcal{L}[A_{ii}(t)]$ *is in class* $\mathcal{P}^*$*, and for* $i > j$*,* $\|A_{ij}(t)\|_\infty$ *is integrable. The Laplacian process* $\{\mathcal{A}(t)\}$ *is ergodic, if its infinite flow graph has a spanning directed rooted tree.*

3.2 Implications

Theorem 3 studies ergodicity (and hence, consensus) of time-varying processes that are specific combinations of processes that are in class $\mathcal{P}^*$. It is worth discussing what processes are in class $\mathcal{P}^*$ and the implications of Theorem 3 to these processes. We show that a very broad class of processes, i.e., the *average cut-balanced* processes (that were studied in [37]), are subset of class $\mathcal{P}^*$.

Definition 11 (Average Cut-Balanced Processes)**.** *We say that* $\{A(t)\}$ *satisfies average cut-balanced property if there exists a sequence* $\{t_s\}_{s\in\mathbb{N}}$ *of increasing times with* $\lim_{s\to\infty} t_s = \infty$ *such that for* $R(s) \triangleq \int_{t_s}^{t_{s+1}} A(\tau)\, d\tau$*:*

(a) There exists a uniform upper bound M *such that for all* $s \in \mathbb{N}$ *we have* $R(s) \leq Mee^T$*.*

(b) $R(s)$ *is cut-balanced for all* $s \in \mathbb{N}$ *with some constant parameter* $K \geq 1$ *(independent of* s*).*

In [37], it is proved that the average cut-balanced processes that satisfy the infinite flow property (i.e., its directed infinite flow graph is connected), admit consensus. Furthermore, this class includes instantaneously cut-balanced processes (i.e., processes $\{A(t)\}$ that $A(t)$ is cut-balanced for all t) that were introduced and their ergodic behaviors were studied in [25]. We show that class $\mathcal{P}^*$ includes average cut-balanced processes.

Lemma 5. *Processes $\{A(t)\}$ satisfying Definition 11 are in class $\mathcal{P}^*$.*

The proof of this result is provided in Appendix. Note that if in the statement of Theorem 3, for each $i \in [m]$, the diagonal block $A_{ii}(t)$ is average cut-balanced (Definition 11), then by Lemma 5, $A_{ii}(t)$ is in class $\mathcal{P}^*$. Therefore, in this case, by Theorem 3 the process $\{\mathcal{A}(t)\}$ will be ergodic. However, $\{\mathcal{A}(t)\}$ itself does not satisfy Definition 11, and hence, its ergodicity cannot be deduced using the previously known results (e.g., [37]).

Note that the class of cut-balanced processes itself contains many of the existing classes of processes in the literature:

- *M-subsymmetric processes*: where $a_{ij} \leq M a_{ji}$ for all $i,j \in [n]$. The discrete-time variation of such processes have been studied in [7] and [35] and they have important implications in the study of Hegselmann-Krause opinion dynamics [5, 39, 67].
- *Weight-balanced processes*: where $\sum_{i\neq j} a_{ij} = \sum_{i\neq j} a_{ji}$ for all $i \in [n]$. These processes have been studied extensively in the past and has application in the study of distributed optimization algorithms over networks [22, 30, 32].

In addition to the above broad class of $\mathcal{P}^*$-processes, in the following we show that this class is large enough to be invariant under ℓ_1-*approximations* as defined below.

Definition 12. *We say $Z(t)$ as an ℓ_1-approximation of $A(t)$ if $\|Z(t)-A(t)\|_\infty$ is integrable. Also, we say $Z(t) = [\zeta_{ij}(t)]_{n\times n}$ is minimal ℓ_1-approximation of $A(t)$ with the directed infinite flow graph $([n], E^\infty)$ if $Z(t)$ is Laplacian for $t \geq t_0$ and*

$$\zeta_{ij}(t) = \begin{cases} a_{ij}(t), & \textit{if } (j,i) \in E^\infty \\ 0, & \textit{if } i \neq j \in [n], (j,i) \notin E^\infty \end{cases}.$$

If $Z(t)$ is an ℓ_1-approximation of $A(t)$, in [8] (for discrete chain in [68]), it is proved that if the process (chain) $\{A(t)\}$ is ergodic, then the process (chain) $\{Z(t)\}$ is ergodic. In the following lemma, we show that this mutual relation is also the case for the class $\mathcal{P}^*$.

Lemma 6. *(Approximation Lemma for class $\mathcal{P}^*$) Let $\{Z(t)\}$ be an ℓ_1-approximation of $\{A(t)\}$. Then, if $\{A(t)\}$ is in class $\mathcal{P}^*$, $\{Z(t)\}$ is in class $\mathcal{P}^*$.*

We prove this result in Appendix. Lemma 6 and Lemma 5 imply that not only the average cut-balanced processes are in class $\mathcal{P}^*$, but also any of their ℓ_1-approximations belong to class $\mathcal{P}^*$, and Theorem 3 applies to any group of such processes that are interconnected by a spanning directed rooted tree.

3.3 Proof of the Main Result

In this section, we provide the proof of the main result of this chapter, i.e., Theorem 3.

We say that a dynamics $x : [t_0, +\infty) \to \mathbb{R}^n$ is *driven* by a sub-Laplacian process $\{A(t)\}$ with a *perturbation* $\lambda : [t_0, +\infty) \to \mathbb{R}^n$ (where $\int_\tau^t \lambda(s)ds < \infty$ for all $t_0 \leq \tau \leq t$) and the initial condition $x(t_0)$ is a dynamics satisfying

$$x(t) = x(t_0) + \int_{t_0}^{t} (A(\tau)x(\tau) + \lambda(\tau))\, d\tau. \tag{3.4}$$

Here, $\lambda(t)$ can be viewed as the input to the averaging dynamics and since it is assumed $\int_\tau^t \lambda(s)ds < \infty$ for all $t_0 \leq \tau \leq t$, the solution $x(t)$ is a continuous function. From (3.4), it follows that

$$x(t) = \Phi(t, t_0)x(t_0) + \int_{t_0}^{t} \Phi(t, \tau)\lambda(\tau)\; d\tau, \tag{3.5}$$

where $\Phi(t, \tau)$ is the transition matrix from time τ to t (for the unforced dynamics $\dot{x}(t) = A(t)x(t)$).

In the following discussion, we study (3.4) where $\lambda(t) = B(t)y(t)$ and $\{[A(t), B(t)]\}$ is a Laplacian process.

Remark 3. *Note that if $x^{(i)}(t)$ is driven by $\{A(t)\}$ with the perturbation $\lambda^{(i)}(t)$ and the initial condition $x^{(i)}(t_0)$, then, $\sum_i \mu_i x^{(i)}(t)$ is driven by $\{A(t)\}$ with the perturbation $\sum_i \mu_i \lambda^{(i)}(t)$ and the initial condition $\sum_i \mu_i x^{(i)}(t_0)$.*

Lemma 7. *Let $x(t) \in \mathbb{R}^n$ be defined by*

$$x(t) = x(t_0) + \int_{t_0}^{t} A(\tau)x(\tau) + B(\tau)y(\tau)\, d\tau.$$

where $[A(t), B(t)]$ is a Laplacian matrix, $x(t)$ is a continuous function, and $B(t)$ is some $n \times m$ matrix, and $y(t) \in \mathbb{R}^m$ for some $m \geq 1$ and for all $t \geq t_0$. Then, we have

$$\max_i x_i(t) \leq \max\left\{\max_i x_i(t_0), \sup_{t \geq \tau \geq t_0} \max_i y_i(\tau)\right\},$$
$$\min_i x_i(t) \geq \min\left\{\min_i x_i(t_0), \inf_{t \geq \tau \geq t_0} \min_i y_i(\tau)\right\}.$$

Proof: Let $M(t) = \max_i x_i(t)$ and $p(t) \in \arg\max_i x_i(t)$ for $t \geq t_0$. Since $x_i(t)$ is a continuous function for all $i \in [n]$, $M(t)$ is also a continuous function of t. Fix $t \geq t_0$, and let $\mu \triangleq \sup_{t \geq \tau \geq t_0} \max_i y_i(\tau)$. If $M(t) \leq \mu$, there is nothing to prove; hence, assume that $\mu < M(t)$. Let $s \triangleq \inf\{s' \geq t_0 | M(\tau) \geq \mu, \forall \tau \in [s', t]\}$. Since $M(t)$ is continuous, if $s > t_0$, then $M(s) = \mu$. Therefore, $M(s) \leq \max\{M(t_0), \mu\}$. From Proposition 2 in [26], we have

$$\begin{aligned}
M(t) &= M(s) + \int_s^t \sum_{j=1}^{n} a_{p(\tau)j}(\tau)(x_j(\tau) - M(\tau))\, d\tau + \int_s^t \sum_{j=1}^{m} b_{p(\tau)j}(\tau)(y_j(\tau) - M(\tau))\, d\tau \\
&\overset{(a)}{\leq} M(s) \\
&\leq \max\{M(t_0), \mu\}
\end{aligned}$$

where (a) follows from the facts that $M(\tau)$ is the maximum of $x(\tau)$, $y(\tau) \leq \mu e \leq M(\tau)e$ for $\tau \in (s,t)$, and $a_{p(\tau)j}(\tau)$ and $b_{p(\tau)j}(\tau)$ are non-negative for all τ. The proof of the lower bound follows from a similar argument. ∎

Remark 4. *Suppose that $\{[A(t),B(t)]\}$ is a Laplacian process, and $\Phi(t,\tau)$ is the transition matrix of the process $\{A(t)\}$. Let*

$$\Psi(t,\tau) \triangleq \left[\Phi(t,\tau), \int_{t_0}^{t} \Phi(t,\tau)B(\tau)d\tau\right], \text{ and } z(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}.$$

Assume $y(t)$ is constant, i.e., $y(t) = y_0$ for some $y_0 \in \mathbb{R}^m$. Then, we have

1. *$x(t) = \Psi(t,\tau)z(\tau)$, which is implied by* (3.5).
2. *$\Psi(t,\tau)$ is a stochastic matrix. This is because from Lemma 7, $\min_\ell z_\ell(\tau) \leq z_i(t) \leq \max_\ell z_\ell(\tau)$, and hence, the elements of $\Psi(t,\tau)$ are all non-negative. Moreover, because $\{[A(t),B(t)]\}$ is a Laplacian process, from* (3.4)*, if $z(\tau) = ae$ for some $a \in \mathbb{R}$, then $z(t) = ae$. Therefore, the summation of each row of $\Psi(t,\tau)$ is equal to one.*

Inspired by classical results in the study of inhomogeneous Markov chains [61], we have the following lemma whose proof is provided in Appendix.

Lemma 8. *The Laplacian process $\{A(t)\}$ is ergodic if and only if there exists $q > 0$ such that for all $\tau \geq t_0$*

$$\liminf_{t\to\infty} \Phi(t,\tau)e_{k(\tau)} \geq qe \tag{3.6}$$

for some $k(\tau) \in [n]$.

Motivated by this result we have the following definition.

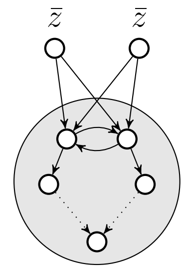

Figure 3.1: Lemma 9 considers this special case of the main theorem

Definition 13. *For a Laplacian process* $\{A(t)\}$ *with the transition matrix* $\Phi(t,\tau)$*, we say that the* k*th coordinate is* q^**, if there exists* $q > 0$ *such that* $\liminf_{t\to\infty} \Phi(t,\tau)e_k \geq qe$ *for some* $t_q \geq t_0$*, and all* $\tau \geq t_q$*.*

Note that by Lemma 8, if one of the coordinates of $\{A(t)\}$ is q^*, then $\{A(t)\}$ is ergodic. Also, by the definition of class $\mathcal{P}^*$, if $A(t)$ is in class $\mathcal{P}^*$ and ergodic, then all its coordinates are q^*.

The following lemma plays a key role in the proof of Theorem 3. This lemma is a special case of the main theorem, in which we have only one $\mathcal{P}^*$ process, the outside nodes' values are fixed to $\bar{z}$, and they are only connected to the nodes in the $\mathcal{P}^*$ process. This situation is similar to Lemma 3 in Chapter 2, which is depicted in Fig. (3.1) where the gray area is representing the nodes in the $\mathcal{P}^*$ process.

Lemma 9. *Let* $x(t)$ *be defined by*

$$x(t) = x(t_0) + \int_{t_0}^{t} A(\tau)x(\tau) + B(\tau)\bar{z}e \, d\tau, \tag{3.7}$$

where $\{[A(t), B(t)]\}$ *is a Laplacian process. Suppose that the* k*th coordinate of* $\mathcal{L}[A(t)]$ *is* q^* *and* $\int_s^\infty e_k^T B(\tau)e \, d\tau = \infty$ *for all* $s \geq t_0$*. Then, we have* $\lim_{t\to\infty} x(t) = \bar{z}e$ *for all* $x(t_0) \in \mathbb{R}^n$*.*

Proof: Without loss of generality, we assume that $k = 1$. First, we show that instead of considering all initial conditions $x(t_0) \in \mathbb{R}^n$, it suffices to prove the lemma for a finite number of initial conditions. For this, let $v^0 = 0e$ and for $\ell \in [n]$ define $v^\ell = \sum_{i=\ell}^n e_i$. Then the vectors $\begin{bmatrix} v^0 \\ e \end{bmatrix}, \ldots, \begin{bmatrix} v^n \\ e \end{bmatrix}$ span

$$\left\{ \begin{bmatrix} v \\ \bar{z}e \end{bmatrix} \middle| v \in \mathbb{R}^n, \bar{z} \in \mathbb{R} \right\}.$$

Therefore, for any initial condition $x(t_0) \in \mathbb{R}^n$, we can write

$$\begin{bmatrix} x(t_0) \\ \bar{z}e \end{bmatrix} = \sum_{\ell=0}^{n} \mu_\ell \begin{bmatrix} v^\ell \\ e \end{bmatrix},$$

where $\sum_{\ell=0}^n \mu_\ell = \bar{z}$. But by Remark 4, we have

$$\lim_{t\to\infty} x(t) = \lim_{t\to\infty} \sum_{\ell=0}^{n} \mu_\ell \Psi(t,t_0) \begin{bmatrix} v^\ell \\ e \end{bmatrix},$$

therefore, it is sufficient to show that $\lim_{t\to\infty} x(t) = e$ for initial conditions $x(t_0) = v^\ell$ for all $\ell \in [n] \cup \{0\}$, and $\bar{z} = 1$, i.e.,

$$\lim_{t\to\infty} \Psi(t,t_0) \begin{bmatrix} v^\ell \\ e \end{bmatrix} = e,$$

as then, for all $x(t_0) \in \mathbb{R}^n$ and $\bar{z} \in \mathbb{R}$, and the above discussion, we would have

$$\lim_{t\to\infty} x(t) = \lim_{t\to\infty} \sum_{\ell=0}^{n} \mu_\ell \Psi(t,t_0) \begin{bmatrix} v^\ell \\ e \end{bmatrix} = \sum_{\ell=0}^{n} \mu_\ell e = \bar{z}e.$$

Let $x(t_0) = v^\ell$ and $\bar{z} = 1$. Let $\alpha(t) = B(t)e$ and $u_i(t) = \alpha_i(t)(1 - x_i(t))$ for $i \in [n]$. Since $x(t_0) \le e$ and $\bar{z} = 1$, Lemma 7 implies that $x(t) \le e$ for $t \ge t_0$, and hence, $u_i(t) \ge 0$ for all $i \in [n]$ and $t \ge t_0$.

For now assume that, except the first row, all other rows of $B(t)$ are zero, i.e., $b_{ij}(t) = 0$ (and hence, $\alpha_i(t) = 0$) for $i \neq 1$, for all $t \geq t_0$. Let $\Phi(t,\tau)$ be the transition matrix associated with the process $\mathcal{L}[A(t)]$. We can write (3.7) as

$$x(t) = x(t_0) + \int_{t_0}^{t} \mathcal{L}[A(\tau)]x(\tau) + u_1(\tau)e_1 \, d\tau. \tag{3.8}$$

Therefore, from (3.5), any solution $x(t)$ to (3.8), satisfies

$$\begin{aligned} x(t) &= \Phi(t,t_0)x(t_0) + \int_{t_0}^{t} \Phi(t,\tau)u_1(\tau)e_1 \, d\tau \\ &= \Phi(t,t_0)x(t_0) + \int_{t_0}^{\infty} 1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1 \, d\tau, \end{aligned} \tag{3.9}$$

where

$$1_{(a,b]}(\tau) = \begin{cases} 1 & \tau \in (a,b] \\ 0 & \text{otherwise} \end{cases}.$$

Now, we show that we can take the limit into the integral, i.e.,

$$\lim_{t\to\infty} \int_{t_q}^{\infty} 1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1 \, d\tau = \int_{t_q}^{\infty} \lim_{t\to\infty} 1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1 \, d\tau$$

for some $t_q \geq t_0$, which will result in $\lim_{t\to\infty} x(t) = \bar{x}e$ for some $\bar{x} \in [0,1]$. From (3.9), we can write

$$\begin{aligned} \liminf_{t\to\infty} x(t) &\geq \liminf_{t\to\infty} \Phi(t,t_0)x(t_0) + \liminf_{t\to\infty} \int_{t_0}^{\infty} 1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1 \, d\tau \\ &\overset{(a)}{\geq} \liminf_{t\to\infty} \int_{t_0}^{\infty} 1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1 \, d\tau \\ &\overset{(b)}{\geq} \int_{t_0}^{\infty} \liminf_{t\to\infty} 1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1 \, d\tau \\ &\overset{(c)}{\geq} \int_{t_q}^{\infty} u_1(\tau)qe \, d\tau, \end{aligned} \tag{3.10}$$

where (a) follows from the fact that $\Phi(t,t_0)$ is a non-negative matrix and $x(t_0) \geq 0e$, (b) follows from Fatou's Lemma (cf. Theorem 2.18 in [21]) and the fact that $1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1$ is non-negative, and (c) holds as the first coordinate is q^*, and hence, there exists $q > 0$ such that $\liminf_{t\to\infty}\Phi(t,\tau)e_1 \geq qe$ for all $\tau \geq t_q$. Since, $\liminf_{t\to\infty} x(t)$ is bounded, and $q > 0$, (3.10) implies that

$$\infty > \frac{1}{q}\liminf_{t\to\infty} x(t) \geq \int_{t_q}^{\infty} u_1(\tau)e\, d\tau. \tag{3.11}$$

On the other hand, since $\Phi(t,\tau) \leq ee^T$, we have

$$\int_{t_q}^{\infty} u_1(\tau)e\, d\tau \geq \int_{t_q}^{\infty} 1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1\, d\tau.$$

Hence, we have

$$\begin{aligned}
&\lim_{t\to\infty}\int_{t_0}^{\infty} 1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1\, d\tau \\
&\qquad = \lim_{t\to\infty}\int_{t_0}^{t_q} u_1(\tau)\Phi(t,\tau)e_1\, d\tau + \lim_{t\to\infty}\int_{t_q}^{\infty} 1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1\, d\tau \\
&\qquad \overset{(a)}{=} \left(\lim_{t\to\infty}\Phi(t,t_q)\right)\int_{t_0}^{t_q} u_1(\tau)\Phi(t_q,\tau)e_1\, d\tau + \int_{t_q}^{\infty}\lim_{t\to\infty}\left(1_{(t_0,t]}(\tau)u_1(\tau)\Phi(t,\tau)e_1\right)\, d\tau \\
&\qquad \overset{(b)}{=} e\pi^T(t_q)\left(x(t_q) - \Phi(t_q,t_0)x(t_0)\right) + \int_{t_q}^{\infty} u_1(\tau)\pi_1(\tau)e\, d\tau,
\end{aligned}$$

where (a) follows from the Dominated Convergence Theorem (cf. Theorem 2.24 [21]), and (b) follows from (3.9) for $t = t_q$ and Lemma 8. This and (3.9) imply that there exists $\bar{x} \in [0,1]$ such that $\lim_{t\to\infty} x(t) = \bar{x}e$. If $\bar{x} < 1$, then let $\epsilon = \frac{1-\bar{x}}{2}$ and $t_\epsilon \geq t_q$ be such that $x_1(t) \leq \bar{x} + \epsilon$ for all

$t \geq t_\epsilon$. Then, from (3.11), we have

$$\begin{aligned}\infty &> \int_{t_q}^{\infty} \alpha_1(\tau)(1-x_1(\tau))\, d\tau \\ &\geq \int_{t_\epsilon}^{\infty} \alpha_1(\tau)(1-x_1(\tau))\, d\tau \\ &\geq \int_{t_\epsilon}^{\infty} \alpha_1(\tau)(1-\bar{x}-\epsilon)\, d\tau \\ &= \frac{1-\bar{x}}{2}\int_{t_\epsilon}^{\infty} \alpha_1(\tau)\, d\tau.\end{aligned}$$

which contradicts the assumption $\int_s^\infty \alpha(\tau) d\tau = \infty$ for all $s \geq t_0$. Therefore, $\bar{x} = 1$ which concludes the proof for the case $b_{ij}(t) = 0$ for $i > 1$, all j, and all $t \geq t_0$. In other words, so far, we have proved that

$$\lim_{t\to\infty} \Psi(t,t_0) \begin{bmatrix} x(t_0) \\ 1 \end{bmatrix} = e, \tag{3.12}$$

where $\Psi(t,\tau)$ is the transition matrix of the dynamics

$$x(t) = x(t_0) + \int_{t_0}^{t} \mathcal{L}[[A(\tau), \alpha_1(\tau)e_1]] \begin{bmatrix} x(\tau) \\ 1 \end{bmatrix} d\tau.$$

For the general case $b_{ij}(t) \geq 0$, we can write the dynamics as

$$x(t) = x(t_0) + \int_{t_0}^{t} \mathcal{L}[[A(\tau), \alpha_1(\tau)e_1]] \begin{bmatrix} x(\tau) \\ 1 \end{bmatrix} + \lambda(\tau)\, d\tau,$$

where $\lambda(t) = \sum_{i=2}^{n} u_i(t)e_i$. From (3.5), we have

$$\lim_{t\to\infty} x(t) = \lim_{t\to\infty} \Psi(t,t_0) \begin{bmatrix} x(t_0) \\ 1 \end{bmatrix} + \int_{t_0}^{t} \Phi(t,\tau)\lambda(\tau)\, d\tau$$
$$\geq \lim_{t\to\infty} \Psi(t,t_0) \begin{bmatrix} x(t_0) \\ 1 \end{bmatrix},$$

which is implied by the fact that $\Phi(t,\tau)$ and $\lambda(t)$ are non-negative. Finally, $x(t) \leq e$ and (3.12) complete the proof. ∎

Using an induction on m and the following result, we will show the main result.

Lemma 10. *Consider a Laplacian process*

$$\mathcal{A}(t) = \begin{bmatrix} A(t) & B(t) \\ 0ee^T & \hat{A}(t) \end{bmatrix}.$$

where all the coordinates of $\mathcal{L}[A(t)]$ are q^*, and $\hat{A}(t)$ is ergodic. The Laplacian process $\{\mathcal{A}(t)\}$ is ergodic if its directed infinite flow graph has a spanning directed rooted tree.

Proof: Let $z(t)$ be driven by $\{\mathcal{A}(t)\}$ with an initial condition $z(t_0)$, and suppose that $A(t)$ and $\mathcal{A}(t)$ are $n \times n$ and $N \times N$, respectively. Let $x(t) \triangleq [z_1(t), \ldots, z_n(t)]$ and $y(t) \triangleq [z_{N-n+1}, \ldots, z_N(t)]$, respectively. Because $x(t)$ does not contribute to the dynamics of $y(t)$, and $\hat{A}(t)$ is ergodic, we have $\lim_{t\to\infty} y(t) = \bar{z}e$ for some $\bar{z}$. Set $s_\epsilon \geq t_0$ such that

$$\sup_{\tau\in[s_\epsilon,\infty)} \|y(\tau) - \bar{z}e\|_\infty < \frac{\epsilon}{2}. \tag{3.13}$$

Let $w(t)$ be defined by

$$w(t) = w(s_\epsilon) + \int_{s_\epsilon}^{t} A(\tau)w(\tau) + B(\tau)\bar{z}e \, d\tau, \tag{3.14}$$

with the initial time s_ϵ and the initial condition $w(s_\epsilon) = x(s_\epsilon)$. Since the directed infinite flow graph has a spanning directed rooted tree, we have $\int_s^\infty e_k^T B(\tau)e \, d\tau = \infty$ for all $s \geq t_0$. Because of this and the fact that the coordinates of $\mathcal{L}[A(t)]$ are q^*, Lemma 9 implies $\lim_{t\to\infty} w(t) = \bar{z}e$. Set $t_\epsilon \geq s_\epsilon$ such that

$$\sup_{\tau \in [t_\epsilon, \infty)} \|w(t) - \bar{z}e\|_\infty < \frac{\epsilon}{2}. \tag{3.15}$$

Let $v(t)$ be driven by

$$v(t) = v(s_\epsilon) + \int_{s_\epsilon}^{t} A(\tau)v(\tau) + B(\tau)(y(\tau) - \bar{z}e) \, d\tau \tag{3.16}$$

with the initial condition $v(s_\epsilon) = 0e$. Because the block $[A(t), B(t)]$ is a Laplacian matrix, Lemma 7 implies

$$\|v(t)\|_\infty \leq \max_{s_\epsilon \leq \tau < t} \|y(\tau) - \bar{z}e\|_\infty < \frac{\epsilon}{2}. \tag{3.17}$$

Adding (3.14) and (3.16), we have

$$w(t) + v(t) = w(s_\epsilon) + v(s_\epsilon) + \int_{s_\epsilon}^{t} A(\tau)\left[w(\tau) + v(\tau)\right] + B(\tau)y(\tau) \, d\tau.$$

Also, we have $x(s_\epsilon) = w(s_\epsilon) + v(s_\epsilon)$. Therefore, we have $x(t) = w(t) + v(t)$ for $s_\epsilon \leq t$, and hence, for $t_\epsilon \leq t$

$$\begin{aligned}\|x(t)-\bar{z}e\|_\infty &= \|w(t)-\bar{z}e+v(t)\|_\infty \\ &\leq \|w(t)-\bar{z}e\|_\infty + \|v(t)\|_\infty \\ &\leq \frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon,\end{aligned}$$

which follows from (3.15) and (3.17). This is true for every $\epsilon > 0$, and hence, we have $\lim_{t\to\infty} x(t) = \bar{z}e$. ∎

To use Lemma 10 in the proof of the main theorem, we need to identify q^* coordinates in $\mathcal{P}^*$-processes, which is established in the following result.

Lemma 11. *Let $\{\mathcal{Q}(t)\}$ be an $\ell \times \ell$ Laplacian process in the class $\mathcal{P}^*$. Then, there is a permutation of indices in $[\ell]$ such that the permuted process $\{\hat{\mathcal{Q}}(t)\}$ can be written as the block matrix form $\hat{\mathcal{Q}}(t) = [\hat{Q}_{ij}(t)]_{m\times m}$ such that all coordinates of $\mathcal{L}[\hat{Q}_{ii}(t)]$ are q^* for $i \in [m]$, and $\|\hat{Q}_{ij}(t)\|_\infty$ is integrable for $i \neq j \in [m]$.*

The proof of this result is provided in Appendix.

Proof of Theorem 3: Without loss of generality we can assume that the diagonal matrices $\mathcal{L}[A_{ii}(t)]$ are q^* for all $i \in [m]$, otherwise, by Lemma 11, we can rename the indices of $[A_{ij}(t)]_{m\times m}$ such that the resulting process has a block-matrix form $[\tilde{A}_{ij}(t)]_{\tilde{m}\times\tilde{m}}$ with q^* diagonal blocks. Since renaming the indices that are associated with a block (shuffling rows and columns associated with that block, accordingly) does not affect the sum of the lower diagonal elements, the lower diagonal blocks $\|\tilde{A}_{ij}(t)\|_\infty$ are integrable for $i > j$. Also, having a spanning directed rooted tree is invariant under permuting the vertices and hence, we can work with the matrix $[\tilde{A}_{ij}(t)]_{\tilde{m}\times\tilde{m}}$ and we can as well assume that the original process $[A_{ij}(t)]_{m\times m}$ has this property.

Also, since $\|A_{ij}(t)\|_\infty$ is integrable for $i > j$, from Proposition 1 in [8], setting $A_{ij}(t) = 0$

for $i > j$ and $t \geq t_0$ will not affect the ergodic properties of the process. Therefore, again, without loss of generality, we assume that $A_{ij}(t) = 0$ for all blocks $i > j$ and $t \geq t_0$. We use induction on m to prove the theorem. The result is true for $m = 1$ as the process has a q^*-coordinate and hence, by Lemma 8, it is ergodic. Now assume that $\hat{\mathcal{A}}(t) \triangleq [\hat{A}_{ij}(t)]_{(m-1)\times(m-1)}$, where $\hat{A}_{ij} = A_{i+1j+1}(t)$, is ergodic for $m > 1$. Then

$$\mathcal{A}(t) = \begin{bmatrix} A_{11}(t) & C(t) \\ 0ee^T & \hat{\mathcal{A}}(t) \end{bmatrix},$$

where $C(t) = [A_{1j}(t)]_{1\times m}$. Note that $\hat{\mathcal{A}}(t)$ is ergodic and all the coordinates in $\mathcal{L}[A_{11}(t)]$ are q^*, and hence, by Lemma 3.5, we conclude the proof. ■

3.4 Proof of the Minor Results

In this section, we provide the proof of the minor result of this chapter, i.e., the lemmas related to the implications of the main result and some other lemmas in the proof of the main result. The following lemma is used in the proof of Lemma 5.

Lemma 12 ([37]). *For a Laplacian process $\{A(t)\}$ with the transition matrix $\Phi(t,\tau) = [\phi_{ij}(t,\tau)]$,*

(a) if $\int_\tau^t A(s)\, ds \leq M$, then $\phi_{ii}(t,\tau) \geq \beta(M)$ for some $\beta(M) > 0$, and

(b) if in addition $\int_\tau^t A(s)\, ds$ is cut-balanced with parameter K, then $\Phi(t,\tau)$ is cut-balanced with parameter $\alpha(K,M)$ for some $\alpha(K,M) > 0$.

For the proof of this lemma, see the proof of Lemma 8 in [37], noting that cut-balanced assumption is not used for the proof of Lemma 8-(a) in [37].

Lemma 13. *Let $\{A(t)\}$ and $\{Z(t)\}$ with transition matrices $\Phi_A(t,\tau)$ and $\Phi_Z(t,\tau)$, respectively, be an ℓ_1-approximation of each other. Then, we have*

$$\lim_{\tau\to\infty}\sup_{t\in[\tau,\infty)}\|\Phi_A(t,\tau)-\Phi_Z(t,\tau)\|_\infty=0.$$

Proof: Let

$$r(\tau)\triangleq\int_\tau^\infty\|A(s)-Z(s)\|_\infty\,ds.$$

From Lemma 6 in [8], we have

$$\|\Phi_A(t,\tau)-\Phi_Z(t,\tau)\|_\infty\le\int_\tau^t\|A(s)-Z(s)\|_\infty\,ds.$$

Thus, for all t, we have $\|\Phi_A(t,\tau)-\Phi_Z(t,\tau)\|_\infty\le r(\tau)$, and hence, $\sup_{t\in[\tau,\infty)}\|\Phi_A(t,\tau)-\Phi_Z(t,\tau)\|_\infty\le r(\tau)$. Since, $\{A(t)\}$ and $\{Z(t)\}$ are an ℓ_1-approximation of each other, we have $\lim_{\tau\to\infty}r(\tau)=0$, which completes the proof. ■

Proof of Lemma 5: Let $\Phi(t,\tau)$ be the transition matrix of $\{A(t)\}$ and let $H(s)=[h_{ij}(s)]=\Phi(t_s,t_{s-1})$ where $\{t_s\}$ is the sequence defined in Definition 11. Since $A(t)$ satisfies Definition 11 for all $s\in\mathbb{N}$, Lemma 12-(b) implies that $H(s)$ is cut-balanced with parameter α for some $\alpha>0$. Thus, from Lemma 9 in [71], we know that there exists a stochastic vector sequence $\{\theta(t)\}$ such that

$$\theta^T(s+1)H(s+1)=\theta^T(s) \tag{3.18}$$

and $\theta(s)\ge pe$ for all $s\ge0$ and for some $p>0$. From (3.18), we have

$$\theta^T(s+\kappa)\prod_{\iota=s+1}^{s+\kappa}H(\iota)=\theta^T(s).$$

Since $e \geq \theta(s+\kappa), \theta(s) \geq pe$, we can conclude

$$e^T \prod_{\iota=s+1}^{s+\kappa} H(\iota) \geq \theta^T(s+k) \prod_{\iota=s+1}^{s+\kappa} H(\iota) \geq pe^T.$$

Suppose that $t \geq t_{s+\kappa} \geq t_s \geq \tau$, where $s \triangleq \min\{\hat{s} | t_{\hat{s}} \geq \tau\}$ and $\kappa \triangleq \max\{\hat{\kappa} | t \geq t_{s+\hat{\kappa}}\}$. Thus, we can write

$$\Phi(t,\tau) = \Phi(t, t_{s+\kappa}) \left[\prod_{\iota=s+1}^{s+\kappa} H(\iota) \right] \Phi(t_s, \tau). \tag{3.19}$$

By Definition 11, we have $\int_{t_{s+\kappa}}^{t} A(s)\, ds, \int_{\tau}^{t_s} A(s)\, ds \leq M$. Hence, Lemma 12-(a) implies

$$\phi_{ii}(t, t_{s+\kappa}), \phi_{ii}(t_s, \tau) \geq \beta(M) > 0.$$

Therefore, from (3.19), we have $e^T \Phi(t,\tau) \geq \beta^2(M) pe^T$. ■

Proof of Lemma 6: Let $A(t)$ and $Z(t)$ have transition matrices $\Phi_A(t,\tau)$ and $\Phi_Z(t,\tau)$, respectively. Since $\{A(t)\}$ is a class $\mathcal{P}^*$ process, there exists $p > 0$ such that $e^T \Phi_Z(t,\tau) \geq pe^T$ for all $t_0 \leq \tau \leq t$. Since, $\{A(t)\}$ and $\{Z(t)\}$ are an ℓ_1-approximation of each other, from Lemma 13, there exists τ_p such that for all $\tau_p \leq \tau \leq t$

$$e^T \Phi_Z(t,\tau) \geq \frac{p}{2} e^T.$$

Also, from Lemma 12-(a) $\phi_{ii}(t,\tau) \geq p'$ for some $p' > 0$ and all $t_0 \leq \tau \leq t \leq \tau_p$, where $\Phi_Z(t,\tau) = [\phi_{ij}(t,\tau)]$. Thus, we have

$$e^T \Phi_Z(t,\tau) \geq \frac{pp'}{2} e^T$$

for all $t_0 \leq \tau \leq t$. ■

Proof of Lemma 8: If $\{A(t)\}$ is an ergodic Laplacian process, we have $\liminf_{t\to\infty} \Phi(t,\tau)e_i = p_i(\tau)e$. Moreover, because $\Phi(t,\tau)$ is stochastic, we have $\sum_i p_i(\tau) = 1$, which implies $\max_i p_i(\tau) \geq \frac{1}{n}$. Therefore, letting $k(\tau) \in \arg\max_i p_i(\tau)$ completes the proof of the necessary condition.

For the sufficient condition, consider a sequence $(t_s)_{s\in\mathbb{N}}$ of increasing times with $t_1 = \tau$ and $\lim_{s\to\infty} t_s = \infty$ such that the condition

$$\Phi(t_{s+1}, t_s)e_{k(t_s)} \geq \frac{q}{2}e$$

holds. Finally, classical results in the study of inhomogeneous Markov chains, such as Theorem 4 in [15], imply that

$$\lim_{t\to\infty} \Phi(t,\tau) = \lim_{s\to\infty} \Phi(t_{s+1}, t_s)\cdots\Phi(t_2, t_1) = e\pi^T(\tau)$$

for some $\pi(\tau)$, which means that the process $\{A(t)\}$ is ergodic and the proof is complete. ■

Proof of Lemma 11: Let $z(t)$ be driven by $\{\mathcal{Q}(t)\}$. Suppose the undirected infinite flow graph of $\mathcal{Q}(t)$ has m connected subgraphs with agent sets $V_1, \ldots, V_m$, where $V_1 \cup \cdots \cup V_m = [\ell]$. From Theorem 6 in [6][1], if $\mathcal{Q}(t)$ is in class $\mathcal{P}^*$, then

$$\lim_{t\to\infty} z_i(t) = \bar{z}_j, \text{for some } \bar{z}_j \in \mathbb{R}, \forall i \in V_j, \forall j \in [m]. \tag{3.20}$$

Consider a permutation $r : [\ell] \to [\ell]$ such that

$$r(V_j) = \left\{ \sum_{k=1}^{j-1} |V_k| + 1, \sum_{k=1}^{j-1} |V_k| + 2, \ldots, \sum_{k=1}^{j} |V_k| \right\},$$

[1]The theorem in [6] is not numbered, but it appears between Theorem 5 and Theorem 7.

where $r(S) = \{r(i) \mid i \in S\}$. Apply this permutation to convert $\mathcal{Q}(t)$ into $\hat{\mathcal{Q}}(t) = [\hat{Q}_{ij}(t)]_{m\times m}$ where size of $\hat{Q}_{ij}(t)$ is $|V_i| \times |V_j|$, for $i, j \in [m]$. From (3.20), we have $\lim_{t\to\infty} \Phi_{\hat{\mathcal{Q}}}(t, \tau) = [\Gamma_{ij}(\tau)]_{m\times m}$ where $\Gamma_{ij}(\tau) = e\pi_{ij}^T(\tau)$ for some vector $\pi_{ij}(\tau) \in \mathbb{R}^{|V_j|}$ and $i, j \in [m]$. Let $C(t)$ be the minimal ℓ_1-approximation of $\hat{\mathcal{Q}}(t)$ with the transition matrix $\Phi_C(t, \tau) = [\Lambda_{ij}(\tau)]_{m\times m}$. For i and j that are not connected in $C(t)$, $z_i(t)$ has no contribution in $z_j(t)$, and we have $\Lambda_{ij}(\tau) = 0ee^T$ for $i \neq j \in [m]$. Since, $\{\hat{\mathcal{Q}}(t)\}$ and $\{C(t)\}$ are an ℓ_1-approximation of each other, from Lemma 13, there exists τ_p such that for all $\tau_p \leq \tau$, $\Gamma_{ij}(\tau) \leq \frac{p}{2\ell} ee^T$ for $i \neq j \in [m]$. Because $\hat{\mathcal{Q}}(t)$ is in class $\mathcal{P}^*$, we have

$$\lim_{t\to\infty} e^T \Phi_{\hat{\mathcal{Q}}}(t, \tau) \geq pe^T$$

for some $p > 0$. Hence, for $\tau \geq \tau_p$, we conclude $\Gamma_{ii}(\tau) \geq \frac{p}{2|V_i|} ee^T$, for $i \in [m]$, and the proof is complete. ∎

Chapter 3, in full, is a reprint of the material as it appears in A. Aghajan and B. Touri, Ergodicity of continuous-time distributed averaging dynamics: A spanning directed rooted tree approach, being accepted for publication in IEEE Transactions on Automatic Control. The dissertation author was the primary investigator and author of this paper.

Part II

Distributed Optimization Over Dependent Random Networks

Chapter 4

Averaging-Based Distributed Optimization Solvers

In this chapter, we study distributed optimization over random networks, where the randomness is not only time-varying but also, possibly, dependent on the past. Under the standard assumptions on the local objective functions and step-size sequences for the gradient descent algorithm, we show that the averaging-based distributed optimization solver at each node converges to a global optimizer almost surely if the weight matrices are row-stochastic almost surely, column-stochastic in-expectation, and satisfy certain connectivity assumptions.

The structure of this chapter is as follows: In Section 4.1, we formulate the problem of interest, and in Section 4.2 we state the main result of this chapter, which is Theorem 4. In Section 4.3, we discuss some immediate consequences of the main result. To assist the readability of this chapter, we provide a sketch of the proof of Theorem 4, in Section 4.4. To prove the main result, first we study the behavior of the distributed averaging dynamics over random networks in Section 4.5. Then, in Section 4.6, we extent this analysis to the dynamics with arbitrary control inputs. Finally, the main result, i.e. Theorem 4, is proved in Section 4.7. The lemmas that can be considered roughly general and stand-alone results that are not tied to the specific assumptions related to Theorem 4, are proved in Section 4.8.

4.1 Problem Formulation

Consider a communication network with n nodes or agents such that node i has the cost function $f_i : \mathbb{R}^m \to \mathbb{R}$. Let $F(z) \triangleq \sum_{i=1}^{n} f_i(z)$. The goal of this chapter is to solve

$$\arg\min_{z\in\mathbb{R}^m} F(z) \tag{4.1}$$

distributively with the following assumption on the objective function.

Assumption 1 (Assumption on the Objective Function)**.** *We assume that:*

(a) f_i *is a convex function over* $\mathbb{R}^m$ *for all* $i \in [n]$.

(b) The optimizer set $\mathcal{Z} \triangleq \arg\min_{z\in\mathbb{R}^m} F(z)$ *is non-empty.*

(c) The subgradients of f_i *are uniformly upper bounded, i.e., for all* $g \in \nabla f_i(z)$, $\|g\| \leq L_i$ *for all* $z \in \mathbb{R}^m$ *and all* $i \in [n]$. *We let* $L \triangleq \sum_{i=1}^{n} L_i$.

In this chapter, we are dealing with the dynamics of the n agents estimates of an optimizer $z^* \in \mathcal{Z}$ which we denote them by $\mathbf{x}_i(t)$ for all $i \in [n]$. Therefore, we view $\mathbf{x}(t)$ as a **vector** of n elements in the vector space $\mathbb{R}^m$. One can think of $\mathbf{x}(t)$ as an $n \times m$ matrix.

A distributed solution of (4.1) was first proposed in [45] using the following deterministic dynamics

$$\mathbf{x}_i(t+1) = \sum_{j=1}^{n} w_{ij}(t+1)\mathbf{x}_j(t) - \alpha(t)\mathbf{g}_i(t)$$

for[1] $t \geq 0$, initial conditions $\mathbf{x}_i(0) \in \mathbb{R}^m$ for all $i \in [n]$, where $\mathbf{g}_i(t) \in \mathbb{R}^m$ is a subgradient of

[1]The dynamics work for any initial time t_0, but since it does not make any difference, in this chapter, we set the initial to be zero.

$f_i(z)$ at $z = \mathbf{x}_i(t)$ for $i \in [n]$, and $\{\alpha(t)\}$ is a step-size sequence (in [45] the constant step-sizes variation of this dynamics was studied). We simply refer to this dynamics as the averaging-based distributed optimization solver. We can compactly write the above dynamics as

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\mathbf{g}(t), \tag{4.2}$$

where, $\mathbf{g}(t) = [\mathbf{g}_1(t), \ldots, \mathbf{g}_n(t)]^T$ is the vector of the sub-gradient vectors and matrix multiplication should be understood over the vector-field $\mathbb{R}^m$, i.e.,

$$[W(t+1)\mathbf{x}(t)]_i \triangleq \sum_{j=1}^{n} w_{ij}(t+1)\mathbf{x}_j(t).$$

In distributed optimization, the goal is to find distributed dynamics $\mathbf{x}_i(t)$s such that $\lim_{t\to\infty} \mathbf{x}_i(t) = z$ where $z \in \mathcal{Z}$ for all $i \in [n]$.

4.2 Main Result

In this section, we discuss the main result of this chapter. The proof of the result is provided in the subsequent sections. In this chapter, we consider the random variation of (4.2), i.e., when $\{W(t)\}$ is a chain of random matrices. This random variation was first studied in [33] where to ensure the convergence, it was assumed that this sequence is doubly stochastic almost surely and i.i.d.. This was generalized to random networks that is Markovian in [34]. The dynamics (4.2) with i.i.d. weight matrices that are row-stochastic almost surely and column-stochastic *in-expectation* was studied in [41]. A special case of [41] is the asynchronous gossip algorithm that was introduced in [42]. In this chapter, we provide an overarching framework for the study of (4.2) with possibly dependent random weight matrices that are row-stochastic almost surely and column-stochastic in-expectation. The following assumption highlights the technical requirements for the random weight matrix sequences.

Assumption 2 (Stochastic Assumption)**.** *We assume that the weight matrix sequence* $\{W(t)\}$*, adapted to a filtration* $\{\mathcal{F}(t)\}$*, satisfies*

(a) For all $t \geq 0$*,* $W(t)$ *is row-stochastic almost surely.*

(b) For every $t > 0$*,* $\mathbb{E}[W(t) \mid \mathcal{F}(t-1)]$ *is column-stochastic (and hence, doubly stochastic) almost surely.*

Similar to other works in this domain, our goal is to ensure that $\lim_{t\to\infty} \mathbf{x}_i(t) = z$ *almost surely* for some optimal $z \in \mathcal{Z}$ for all $i \in [n]$. To reach such a consensus value, we need to ensure enough flow of information between the agents, i.e., the associated graph sequence of $\{W(t)\}$ satisfies some form of connectivity over time. More precisely, we assume the following connectivity conditions.

Assumption 3 (Conditional B-Connectivity Assumption)**.** *We assume that for all* $t \geq 0$

(a) Every node in $\mathcal{G}(W(t))$ *has a self-loop, almost surely.*

(b) There exists an integer $B > 0$ *such that the random graph* $\mathcal{G}_B(t) = ([n], \mathcal{E}_B(t))$ *where*

$$\mathcal{E}_B(t) = \bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}(\mathbb{E}[W(\tau)|\mathcal{F}(tB)])$$

has a spanning rooted tree almost surely.

In deterministic distributed optimization, the connectivity condition for time-invariant networks is that $\mathcal{E}(W)$ has a spanning rooted tree, which is generalized to $\bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}(W(\tau))$ having a spanning rooted tree[2] for time-varying networks. In random setting, the connectivity condition for i.i.d. random networks is that $\mathcal{E}(\mathbb{E}[W])$ has a spanning rooted tree. A natural generalization of this condition to dependent random networks is that $\bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}(\mathbb{E}[W(\tau)|\mathcal{F}(\tau -$

[2]Note that $W(t)$s need to doubly stochastic too, which means $W(t)$s are strongly connected.

1)]) has a spanning rooted tree almost surely. We further generalize this condition to Assumption 3-(b). This is due to the following lemma which is proved in Appendix.

Lemma 14. *If the random graph with the vertex set $[n]$ and the edge set*

$$\bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}^{\gamma}(\mathbb{E}[W(\tau)|\mathcal{F}(\tau-1)]),$$

has a spanning rooted tree almost surely, then from some $\gamma \geq \tilde{\gamma} > 0$, the random graph with the vertex set $[n]$ and the edge set

$$\bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}^{\tilde{\gamma}}(\mathbb{E}[W(\tau)|\mathcal{F}(tB)]),$$

has a spanning rooted tree almost surely.

Finally, we assume the following standard condition on the step-size sequence $\{\alpha(t)\}$.

Assumption 4 (Assumption on Step-size)**.** *For the step-size sequence $\{\alpha(t)\}$, we assume that $0 < \alpha(t) \leq Kt^{-\beta}$ for some $K, \beta > 0$ and all $t \geq 0$, $\lim_{t\to\infty} \frac{\alpha(t)}{\alpha(t+1)} = 1$, and*

$$\sum_{t=0}^{\infty} \alpha(t) = \infty \quad \textit{and} \quad \sum_{t=0}^{\infty} \alpha^2(t) < \infty. \tag{4.3}$$

The main result of this chapter is the following theorem.

Theorem 4. *Under the Assumptions 1-4 on the model and the dynamics* (4.2)*, $\lim_{t\to\infty} \mathbf{x}_i(t) = z^*$ almost surely for all $i \in [n]$ and all initial conditions $\mathbf{x}_i(0) \in \mathbb{R}^m$, where z^* is a random vector that is supported on the optimal set $\mathcal{Z}$.*

4.3 Implications

Before continuing with the technical details of the proof, let us first discuss some of the higher-level implications of this result:

4.3.1 Gossip-based sequential solvers

Gossip algorithms, which were originally studied in [9, 3], have been used in solving distributed optimization problems [42, 31]. In gossip algorithms, at each round, a node randomly wakes up and shares its value with all or some of its neighbors. However, it is possible to leverage Theorem 4 to synthesize algorithms that do not require choosing a node independently and uniformly at random or use other coordination methods to update information at every round. An example of such a scheme is as follows:

Example 5. *Consider a connected undirected network*[3] $\mathcal{G} = ([n], E)$*. Consider a token that is handed sequentially in the network and initially it is handed to an arbitrary agent* $\ell(0) \in [n]$ *in the network. If at time* $t \geq 0$*, agent* $\ell(t) \in [n]$ *is in the possession of the token, it chooses one of its neighbors* $s(t+1) \in [n]$ *randomly and by flipping a coin, i.e., with probability* $\frac{1}{2}$ *shares its information to* $s(t+1)$ *and passes the token and with probability* $\frac{1}{2}$ *keeps the token and asks for information from* $s(t+1)$*. It means*

$$\ell(t+1) = \begin{cases} \ell(t), & \textit{with probability } \frac{1}{2} \\ s(t+1), & \textit{with probability } \frac{1}{2} \end{cases}.$$

Finally, the agent $\ell(t+1)$*, who has the token at time* $t+1$ *and is receiving the information, does*

$$\mathbf{x}_{\ell(t+1)}(t+1) = \frac{1}{2}(\mathbf{x}_{s(t+1)}(t) + \mathbf{x}_{\ell(t)}(t)) - \alpha(t)\mathbf{g}_{\ell(t+1)}(t).$$

[3] The graphs do not need to be time-invariant, and this example can be extended to processes over underlying time-varying graphs.

For the other agents $i \neq \ell(t+1)$, we set

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) - \alpha(t)\mathbf{g}_i(t).$$

Let $\mathcal{F}(t) = \sigma(\mathbf{x}(0), \ldots, \mathbf{x}(t), \ell(t))$, and the weight matrix $W(t) = [w_{ij}(t)]$ be

$$w_{ij}(t) = \begin{cases} \frac{1}{2}, & i = j = \ell(t) \\ \frac{1}{2}, & i = \ell(t), j \in \{s(t), \ell(t-1)\} \setminus \{\ell(t)\} \\ 1, & i = j \neq \ell(t) \\ 0, & \textit{otherwise} \end{cases},$$

which is the weight matrix of this scheme. Note that $\mathbb{E}[W(t)|\mathcal{F}(t-1)] = V(\ell(t-1))$ where $R(h) = [r_{ij}(h)]$ with

$$r_{ij}(h) = \begin{cases} \frac{3}{4}, & i = j = h \\ \frac{1}{4\delta_i}, & i = h, (i,j) \in E \\ \frac{1}{4\delta_i}, & j = h, (i,j) \in E \\ 1, & i = j \neq h \\ 0, & \textit{otherwise} \end{cases},$$

where δ_i is the degree of the node i. Note that the matrix $\mathbb{E}[W(t)|\mathcal{F}(t-1)]$ is doubly stochastic, satisfies Assumption 3-(a), and only depends on $\ell(t-1)$. Now, we need to check whether $\{W(t)\}$

satisfies Assumption 3-(b). We have

$$\begin{aligned}
\mathbb{E}[W(t+n)|\mathcal{F}(t)] &= \mathbb{E}[\mathbb{E}[W(t+n) \mid \mathcal{F}(t+n-1)] \mid \mathcal{F}(t)] \\
&= \mathbb{E}[R(\ell(t+n-1)) \mid \mathcal{F}(t)] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} R(\ell(t+n-1))1_{\{\ell(t+n-1)=i\}} \middle| \mathcal{F}(t)\right] \\
&= \sum_{i=1}^{n} \mathbb{E}[R(i)1_{\{\ell(t+n-1)=i\}} \mid \mathcal{F}(t)] \\
&= \sum_{i=1}^{n} R(i)\mathbb{E}[1_{\{\ell(t+n-1)=i\}} \mid \mathcal{F}(t)].
\end{aligned}$$

If the network is connected, starting from any vertex, after $n-1$ steps, the probability of reaching any other vertex is at least $(2\Delta)^{-(n-1)} > 0$, where $\Delta \triangleq \max_{i\in[n]} \delta_i$. Therefore, we have $\mathbb{E}[1_{\{\ell(t+n-1)=i\}}|\mathcal{F}(t)] > 0$ for all $i \in [n]$ and t, and hence, Assumption 3-(b) is satisfied with $B = n$.

4.3.2 Robustness to link-failure

Our result shows that (4.2) is robust to random link-failures. Note that the results such as [33] will not imply the robustness of the algorithms to link failure as it assumes that the resulting weight matrices remain doubly stochastic. To show the robustness of averaging-based solvers, suppose that we have a deterministic doubly stochastic sequence $\{A(t)\}$, and suppose that each link at any time t fails with some probability $p(t) > 0$. More precisely, let $B(t)$ be a failure matrix where $b_{ij}(t) = 0$ if a failure on link (i, j) occurs at time t and otherwise $b_{ij}(t) = 1$ and we have

$$\mathbb{E}[b_{ij}(t)|\mathcal{F}(t-1)] = 1 - p(t), \tag{4.4}$$

for $i, j \in [n]$. For example, if $B(t)$ is independent and identically distributed, i.e.,

$$b_{ij}(t) = \begin{cases} 0, & \text{with probability } p \\ 1, & \text{with probability } 1-p \end{cases},$$

then $B(t)$ satisfies (4.4). Define $W(t) = [w_{ij}(t)]$ as follows

$$w_{ij}(t) \triangleq \begin{cases} a_{ij}(t)b_{ij}(t), & i \neq j \\ 1 - \sum_{\ell \neq i} a_{i\ell}(t)b_{i\ell(t)}, & i = j \end{cases}.$$

Note that $W(t)$ is row-stochastic, and since $A(t)$ is column-stochastic, $\mathbb{E}[W(t)|\mathcal{F}(t-1)]$ is column-stochastic. Thus, Theorem 4, using $W(t)$, translates to a theorem on robustness of the distributed dynamics (4.2): as long as the connectivity conditions of Theorem 4 holds, the dynamics will reach a minimizer of the distributed problem almost surely. For example, if the link failure probability satisfies $p(t) \leq \bar{p}$ for all t and some $\bar{p} < 1$, our result implies that the result of Proposition 4 in [47] (for unconstrained case) would still hold under the above link-failure model. It is worth mentioning that if $\{A(t)\}$ is time-varying, then $\mathbb{E}[W(t)]$ would be time-varying and hence, the previous results on distributed optimization using i.i.d. row-stochastic weight matrices that are column-stochastic in-expectation [41] would not imply such a robustness result.

4.4 Theorem 4: Sketch of Proof

Here, we provide the sketch of the proof of the main result (Theorem 4) to assist with its readability. We can divide the proof into two main steps:

I. We show that $\lim_{t\to\infty} \bar{\mathbf{x}}(t) = z^*$ almost surely, where $\bar{\mathbf{x}} \triangleq \frac{1}{n}e^T\mathbf{x}$ is the average of $\mathbf{x}(t)$ and z^* is a random vector whose support lies on the optimizer set $\mathcal{Z}$.

II. We show that almost surely $\lim_{t\to\infty}\|\mathbf{x}_i(t)-\bar{\mathbf{x}}(t)\|=0$ for all $i\in[n]$.

To show Step I, we fix a $z\in\mathcal{Z}$ and use the Lyapunov (like) function $V(t)=\|\bar{\mathbf{x}}(t)-z\|^2$. In Lemma 26, we show that this Lyapunov function satisfies

$$\mathbb{E}[V(t+1)|\mathcal{F}(t)]\le V(t)-b(t)+c(t), \tag{4.5}$$

where

$$\begin{aligned} b(t) &\triangleq -\frac{2\alpha(t)}{n}(F(\bar{\mathbf{x}}(t))-F(z)), \text{ and} \\ c(t) &\triangleq \alpha^2(t)\frac{L^2}{n^2}+\sum_{i=1}^{n}\|\mathbf{x}_i(t)-\bar{\mathbf{x}}(t)\|^2+\frac{4\alpha(t)}{n}\sum_{i=1}^{n}L_i\|\mathbf{x}_i(t)-\bar{\mathbf{x}}(t)\|. \end{aligned} \tag{4.6}$$

To analyze (4.5), we apply Robbins-Siegmund Theorem [59], which plays a key role in the proof of the above Step I.

Theorem 5 (Robbins-Siegmund Theorem [59]). *Suppose that a non-negative random process $\{\tilde{V}(t)\}$ (adapted to a filtration $\{\tilde{\mathcal{F}}(t)\}$) satisfies*

$$\mathbb{E}\left[\tilde{V}(t+1)\middle|\tilde{\mathcal{F}}(t)\right]\le(1+\tilde{a}(t))\tilde{V}(t)-\tilde{b}(t)+\tilde{c}(t), \tag{4.7}$$

where $\tilde{a}(t),\tilde{b}(t),\tilde{c}(t)\ge 0$ almost surely for all t. Then if $\sum_{t=0}^{\infty}\tilde{a}(t)<\infty$ and $\sum_{t=0}^{\infty}\tilde{c}(t)<\infty$ almost surely, $\lim_{t\to\infty}\tilde{V}(t)$ exists and $\sum_{t=0}^{\infty}\tilde{b}(t)<\infty$ almost surely.

Note that in (4.6), $b(t),c(t)\ge 0$ for all t. To apply the Robbins-Siegmund result for (4.5), we need to prove that $\sum_{t=0}^{\infty}c(t)<\infty$, almost surely. Since $\sum_{t=0}^{\infty}\alpha^2(t)<\infty$, we need to establish

$$\sum_{t=0}^{\infty}\|\mathbf{x}_i(t)-\bar{\mathbf{x}}(t)\|^2<\infty$$

and

$$\sum_{t=0}^{\infty} \alpha(t) \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\| < \infty,$$

almost surely, for all $i \in [n]$. To establish this, in Lemma 22, we find an upper bound on the diameter of $\mathbf{x}(t)$, i.e. $\mathrm{d}(\mathbf{x}(t))$, which is defined in (4.15). Combining $\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\| \leq \mathrm{d}(\mathbf{x})$ (Lemma 16-(e)) and Lemma 22, we arrive at

$$\mathbb{E}\left[\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|^2\right] \leq M\alpha^2(t), \tag{4.8}$$

for some constant $M > 0$.

While in (4.8), we show that $\mathbf{x}_i(t)$ converges to $\bar{\mathbf{x}}(t)$ (in second moment) with the convergence rate $\alpha^2(t)$, to prove Step II, we need to show that it converges to $\bar{\mathbf{x}}(t)$ almost surely. However, we provide a stronger result and in Lemma 24, we show that

$$\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\| \lessapprox \hat{M}\alpha(t), \tag{4.9}$$

for some constant $\hat{M} > 0$.

To show (4.8) and (4.9), we study the conditional expectation of the diameter of $\mathbf{x}(t)$. To do so, we derive

$$\begin{aligned} \mathbb{E}[\mathrm{d}(\mathbf{x}(t))|\mathcal{F}(\tau)] \leq \mathbb{E}[\mathrm{diam}(\Phi(t,\tau))|\mathcal{F}(\tau)]\mathrm{d}(\mathbf{x}(\tau)) \\ + \tilde{L}\sum_{s=\tau}^{t-1} \mathbb{E}[\mathrm{diam}(\Phi(t,s+1))|\mathcal{F}(\tau)]\alpha(s), \end{aligned} \tag{4.10}$$

from the main dynamics (4.2) for some constant $\tilde{L} > 0$ where $\mathrm{diam}(A)$ is the diameter of the matrix A and defined in (4.13). Therefore, we need to investigate $\mathbb{E}[\mathrm{diam}(\Phi(t,\tau))|\mathcal{F}(\tau)]$ for all

$t \geq \tau$. In Lemma 19, we show that $\mathbb{E}[\mathrm{diam}(\Phi(t,\tau))|\mathcal{F}(\tau)]$ exponentially goes to zero, i.e.,

$$\mathbb{E}[\mathrm{diam}(\Phi(t,\tau))|\mathcal{F}(\tau)] \leq C\lambda^{t-\tau}. \tag{4.11}$$

To prove (4.11), in Lemma 17, we show that for a large enough T, $\mathbb{E}[\mathrm{diam}(\Phi(T,\tau))|\mathcal{F}(\tau)] \leq 1-\theta$ for some $\theta > 0$ (Lemma 17 is based on $\Lambda(\cdot)$ which is defined in (4.14). Note that from Lemma 16-(c), we have $\mathrm{diam} = 1-\Lambda$). The main challenge to prove Lemma 17 is to show that the probability of the event $\{\mathrm{diam}(\Phi(T,\tau)) \leq 1-\tilde{\theta}\}$ for some $\tilde{\theta} > \theta$ is away form zero, which is possible due to Assumption 3 (and Assumption 2-(a)). Using (4.11), in Lemma 22, we prove (4.8) and complete the proof of Step I.

To prove Step II, using Assumption 4 ($\alpha(t) \leq Kt^{-\beta}$), first we simplify (4.10) to

$$\mathbb{E}[\mathrm{d}(\mathbf{x}(t))|\mathcal{F}(\tau)] \leq \mathbb{E}[\mathrm{diam}(\Phi(t,\tau))|\mathcal{F}(\tau)]\mathrm{d}(\mathbf{x}(\tau)) + \tilde{K}\tau^{-\beta},$$

for some $\tilde{K} > 0$. However, since $\sum_{\tau=0}^{\infty} \tau^{-\beta}$ is not necessarily summable, we cannot use the standard Robbins-Siegmund Theorem [59] to argue $\mathrm{d}(\mathbf{x}(t)) \to 0$ based on this inequality. We will use the facts that $\mathbb{E}[\mathrm{diam}(\Phi(t,\tau))|\mathcal{F}(\tau)] < 1$ if $t-\tau$ is large enough, and $\mathrm{diam}(\Phi(t,\tau)) \leq 1$ for all $t \geq \tau \geq t$. This leads us to prove a martingale-type result in Lemma 23, which helps us to prove (4.9) (in Lemma 24). This step completes the proof of Step II.

4.5 Autonomous Averaging Dynamics

To prove Theorem 4, we need to study the time-varying distributed averaging dynamics with a particular control input (gradient-like dynamics). To do this, first we study the autonomous averaging dynamics (i.e., without any input) and then, we use the established results to study the controlled dynamics.

For this, consider the time-varying distributed averaging dynamics

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t), \tag{4.12}$$

where $\{W(t)\}$ satisfying Assumption 3. Defining transition matrix

$$\Phi(t,\tau) \triangleq W(t)\cdots W(\tau+1),$$

and $\Phi(\tau,\tau) = I$, we have $\mathbf{x}(t) = \Phi(t,\tau)\mathbf{x}(\tau)$. Note that since $W(t)$s are row-stochastic matrices (a.s.) and the set of row-stochastic matrices is a semi-group (with respect to multiplication), the transition matrices $\Phi(t,\tau)$ are all row-stochastic matrices (a.s.).

We say that a chain $\{W(t)\}$ achieves *consensus* for the initial time 0 if for all i

$$\lim_{t\to\infty} \|\mathbf{x}_i(t) - \tilde{x}\| = 0,$$

almost surely, for all choices of initial condition $\mathbf{x}(0) \in (\mathbb{R}^m)^n$ in (4.12) and some random vector $\tilde{x} = \tilde{x}_{\mathbf{x}(0)}$. It can be shown that an equivalent condition for consensus is to have $\lim_{t\to\infty}\Phi(t,0) = e\pi^T(0)$ for a random stochastic vector $\pi(0) \in \mathbb{R}^n$, almost surely.

For a matrix $A = [a_{ij}]$, let

$$\operatorname{diam}(A) = \max_{i,j\in[n]} \frac{1}{2}\sum_{\ell=1}^{n} |a_{i\ell} - a_{j\ell}|, \tag{4.13}$$

and the mixing parameter

$$\Lambda(A) = \min_{i,j\in[n]} \sum_{\ell=1}^{n} \min\{a_{i\ell}, a_{j\ell}\}. \tag{4.14}$$

Note that for a row-stochastic matrix A, $\operatorname{diam}(A) \in [0,1]$. For a vector $\mathbf{x} = [\mathbf{x}_i]$ where $\mathbf{x}_i \in \mathbb{R}^m$

for all i, let

$$\mathrm{d}(\mathbf{x}) = \max_{i,j\in[n]} \|\mathbf{x}_i - \mathbf{x}_j\|_\infty, \tag{4.15}$$

where $\|\cdot\|$ is ℓ_∞ norm. For convenience and due to the frequent use of ℓ_∞ norm in this chapter, we use $\|\cdot\|$ to denote the ℓ_∞ norm $\|x\| \triangleq \max_{i\in[m]} |x_i|$. Note that $\mathrm{d}(\mathbf{x}) \le 2\max_{i\in[n]} \|\mathbf{x}_i\|$. Also, if we have consensus, then $\lim_{t\to\infty} \mathrm{d}(\mathbf{x}(t)) = 0$ and $\lim_{t\to\infty} \mathrm{diam}(\Phi(t,0)) = 0$ and in fact, the reverse implications are true [15], i.e., a chain achieves consensus if and only if $\lim_{t\to\infty} \mathrm{d}(\mathbf{x}(t)) = 0$ for all $\mathbf{x}(0) \in (\mathbb{R}^m)^n$ or $\lim_{t\to\infty} \mathrm{diam}(\Phi(t,0)) = 0$.

The following results relating the above quantities are useful for our future discussions.

Lemma 15 ([23, 61])**.** *For $n\times n$ row-stochastic matrices A, B, we have*

$$\mathrm{diam}(AB) \le (1-\Lambda(A))\mathrm{diam}(B).$$

Lemma 16. *For any $n\times n$ row-stochastic matrices A, B, we have*

(a) $\mathrm{d}(A\mathbf{x}) \le \mathrm{diam}(A)\mathrm{d}(\mathbf{x})$ *for all* $\mathbf{x} \in (\mathbb{R}^m)^n$*,*

(b) $\mathrm{d}(\mathbf{x}+\mathbf{y}) \le \mathrm{d}(\mathbf{x}) + \mathrm{d}(\mathbf{y})$ *for all* $\mathbf{x}, \mathbf{y} \in (\mathbb{R}^m)^n$*,*

(c) $\mathrm{diam}(A) = 1 - \Lambda(A)$*,*

(d) $\mathrm{diam}(AB) \le \mathrm{diam}(A)\mathrm{diam}(B)$*, and*

(e) $\left\|\mathbf{x}_i - \sum_{j=1}^n \pi_j \mathbf{x}_j\right\| \le \sqrt{n}\mathrm{d}(\mathbf{x})$ *for all* $i \in [n]$*,* $\mathbf{x} \in (\mathbb{R}^m)^n$*, and any stochastic vector* $\pi \in [0,1]^n$ *(i.e.,* $\sum_{i=1}^n \pi_i = 1$*).*

Proof: The proof is provided in Appendix. ■

It is worth mentioning that similar results to Lemma 16 were established in [14] for $m = 1$, and Lemma 16 extends those results to $m > 1$. In [14], it was shown that the semi-norm of a matrix A induced by $\mathrm{d}(\cdot)$ is equal to $1 - \Lambda(A)$ for $m = 1$.

The main goal of this section is to obtain an exponentially decreasing upper bound (in terms of $t_1 - \tau_1$ and $t_2 - \tau_2$) on $\mathbb{E}[\mathrm{diam}(\Phi(t_2, \tau_2))\mathrm{diam}(\Phi(t_1, \tau_1)) \mid \mathcal{F}(\tau_1)]$.

Using this result and a proper connectivity assumption (Assumption 3), we can show that the transition matrices $\Phi(t, s)$ become mixing *in-expectation* for large enough $t > s$.

Lemma 17. *Under Assumption 2-(a) and 3, there exists a parameter $\theta > 0$ such that for every $s \geq 0$, we have almost surely*

$$\mathbb{E}[\Lambda(\Phi((n^2 + s)B, sB)) \mid \mathcal{F}(sB)] \geq \theta.$$

Proof: Fix $s \geq 0$. Let $\mathbb{T}$ be the set of all collection of edges E such that the graph $([n], E)$ has a spanning rooted tree, and for $k \in [n^2]$,

$$\mathcal{E}_B(k) \triangleq \bigcup_{\tau=(s+k-1)B+1}^{(s+k)B} \mathcal{E}(\mathbb{E}[W(\tau)|\mathcal{F}((s+k-1)B)]).$$

For notational simplicity, denote $\mathcal{F}(sB)$ by $\mathcal{F}$ and $\mathcal{F}((s+k)B)$ by $\mathcal{F}_k$ for $k \in [n^2]$. Let $V = \{\omega \mid \forall k \; \mathcal{E}_B(k) \in \mathbb{T}\}$. From Assumption 3, we have $P(V) = 1$. For $\omega \in V$ and $k \geq 1$, define the random graph $([n], \mathcal{T}_k)$ on n vertices by

$$\mathcal{T}_k = \begin{cases} \mathcal{T}_{k-1}, & \text{if } \mathcal{T}_{k-1} \in \mathbb{T} \\ \mathcal{T}_{k-1} \cup \{u_k\}, & \text{if } \mathcal{T}_{k-1} \notin \mathbb{T} \end{cases},$$

with $\mathcal{T}_0 = \emptyset$, where

$$u_k \in \mathcal{E}_B(k) \cap \overline{\mathcal{T}}_{k-1}, \tag{4.16}$$

and $\overline{\mathcal{T}}_k$ is the edge-set of the complement graph of $([n], \mathcal{T}_k)$. Note that since $\mathcal{E}_B(k)$ has a spanning rooted tree, if $\mathcal{T}_{k-1} \notin \mathbb{T}$, then $\mathcal{E}_B(k)$ should contain an edge that does not belong to $\mathcal{T}_{k-1}$, which we identify it as u_k in (4.16). Hence, $\mathcal{T}_k$ is well-defined. Since there are at most $n(n-1)$ potential edges in a graph on n vertices, $\mathcal{T}_{n^2}$ has a spanning rooted tree for $\omega \in V$.

For $k \in [n^2]$, let

$$\mathcal{D}_B(k) \triangleq \bigcup_{\tau=(s+k-1)B+1}^{(s+k)B} \mathcal{E}^{\nu}(W(\tau)),$$

for some fixed $0 < \nu < \gamma$, and

$$\mathcal{H}(k) \triangleq \bigcup_{\tau=1}^{k} \mathcal{D}_B(\tau).$$

Consider the sequences of events $\{U_k\}$ defined by

$$U_k \triangleq \{\omega \in V \mid \mathcal{T}_k \subset \mathcal{H}(k)\},$$

for $k \geq 1$, and $U_0 = V$. Note that if $\mathcal{T}_{k-1} \in \mathbb{T}$, then $\mathcal{T}_{k-1} \subset \mathcal{H}(k-1)$ implies $\mathcal{T}_k \subset \mathcal{H}(k)$, and if $\mathcal{T}_{k-1} \notin \mathbb{T}$, then $\mathcal{T}_{k-1} \subset \mathcal{H}(k-1)$ and $u_k \in \mathcal{D}_B(k)$ imply $\mathcal{T}_k \subset \mathcal{H}(k)$. Hence,

$$1_{\{U_k\}} \geq 1_{\{U_{k-1}\}} 1_{\{\mathcal{T}_{k-1} \notin \mathbb{T}\}} 1_{\{u_k \in \mathcal{D}_B(k)\}} + 1_{\{U_{k-1}\}} 1_{\{\mathcal{T}_{k-1} \in \mathbb{T}\}} \tag{4.17}$$

holds for $k \geq 1$.

On the other hand, from Tower rule (see e.g., Theorem 5.1.6 in [20]), we have

$$\mathbb{E}[1_{\{U_{k-1}\}}1_{\{\mathcal{T}_{k-1}\notin\mathbb{T}\}}1_{\{u_k\in\mathcal{D}_B(k)\}} \mid \mathcal{F}] = \mathbb{E}[1_{\{U_{k-1}\}}1_{\{\mathcal{T}_{k-1}\notin\mathbb{T}\}}\mathbb{E}[1_{\{u_k\in\mathcal{D}_B(k)\}}|\mathcal{F}_{k-1}] \mid \mathcal{F}]. \quad (4.18)$$

Let $u_k(\omega) = (j_k(\omega), i_k(\omega))$. Since $u_k \in \mathcal{E}_B(k)$, there exists $(s+k-1)B < \tau_k \leq (s+k)B$ such that

$$u_k \in \mathcal{E}(\mathbb{E}[W(\tau_k)|\mathcal{F}_{k-1}]),$$

and, we have

$$\mathbb{E}[(1-\nu)1_{\{1-w_{i_kj_k}(\tau_k)\geq 1-\nu\}}|\mathcal{F}_{k-1}] \leq \mathbb{E}[1-w_{i_kj_k}(\tau_k)|\mathcal{F}_{k-1}] \leq 1-\gamma. \quad (4.19)$$

Therefore,

$$\begin{aligned}
\mathbb{E}[1_{\{u_k\in\mathcal{D}_B(k)\}} \mid \mathcal{F}_{k-1}] &\geq \mathbb{E}[1_{\{u_k\in\mathcal{E}^\nu(W(\tau_k))\}} \mid \mathcal{F}_{k-1}] \\
&= \mathbb{E}[1_{\{w_{i_kj_k}(\tau_k)>\nu\}} \mid \mathcal{F}_{k-1}] \\
&= 1-\mathbb{E}[1_{\{w_{i_kj_k}(\tau_k)\leq\nu\}} \mid \mathcal{F}_{k-1}] \\
&= 1-\mathbb{E}[1_{\{1-w_{i_kj_k}(\tau_k)\geq 1-\nu\}} \mid \mathcal{F}_{k-1}] \\
&\geq 1-\frac{1-\gamma}{1-\nu} \triangleq p > 0,
\end{aligned}$$

which holds as $\nu < \gamma$. This inequality and (4.18) imply that

$$\mathbb{E}[1_{\{U_{k-1}\}}1_{\{\mathcal{T}_{k-1}\notin\mathbb{T}\}}1_{\{u_k\in\mathcal{D}_B(k)\}} \mid \mathcal{F}] \geq p\mathbb{E}[1_{\{U_{k-1}\}}1_{\{\mathcal{T}_{k-1}\notin\mathbb{T}\}} \mid \mathcal{F}].$$

Therefore, (4.17) implies

$$\begin{aligned}\mathbb{E}[1_{\{U_k\}}|\mathcal{F}] &\geq p\mathbb{E}[1_{\{U_{k-1}\}}1_{\{\mathcal{T}_{k-1}\notin\mathbb{T}\}}|\mathcal{F}]+\mathbb{E}[1_{\{U_{k-1}\}}1_{\{\mathcal{T}_{k-1}\in\mathbb{T}\}}|\mathcal{F}]\\ &\geq p\left(\mathbb{E}[1_{\{U_{k-1}\}}1_{\{\mathcal{T}_{k-1}\notin\mathbb{T}\}}|\mathcal{F}]+\mathbb{E}[1_{\{U_{k-1}\}}1_{\{\mathcal{T}_{k-1}\in\mathbb{T}\}}|\mathcal{F}]\right)\\ &= p\mathbb{E}[1_{\{U_{k-1}\}}|\mathcal{F}],\end{aligned}$$

and hence $\mathbb{E}[1_{\{U_k\}}|\mathcal{F}]\geq p^k$. Finally, since $\mathcal{T}_{n^2}$ has a spanning rooted tree, from Lemma 1 in [33], we have

$$\Lambda(W(n^2B,\omega)\cdots W(n(n-1)B+n-1,\omega)\cdots W(1,\omega))\geq \nu^{n^2B},$$

for $\omega\in U_{n^2}$. Therefore, we have

$$\mathbb{E}\left[\Lambda(\Phi((n^2+s)B,sB))\middle|\mathcal{F}(sB)\right]\geq \nu^{n^2B}\mathbb{E}[1_{\{U_{n^2}\}}\mid\mathcal{F}]\geq \nu^{n^2B}p^{n^2}\triangleq\theta>0,$$

which completes the proof. ■

Finally, we need the following result, which is proved in Appendix, to prove the main result of this section.

Lemma 18. *For a non-negative random process $\{Y(k)\}$, adapted to a filtration $\{\mathcal{F}(k)\}$, let*

$$\mathbb{E}[Y(k)|\mathcal{F}(k-1)]\leq a(k)$$

for $K_1\leq k\leq K_2$ almost surely, where $K_1\leq K_2$ are arbitrary positive integers and $a(k)$s are (deterministic) scalars. Also, consider the σ-algebra $\tilde{\mathcal{F}}$ such that $\tilde{\mathcal{F}}\subseteq\mathcal{F}(k)$ for all $K_1-1\leq k<$

K_2. *Then,we have almost surely*

$$\mathbb{E}\left[\prod_{k=K_1}^{K_2} Y(k)\middle|\tilde{\mathcal{F}}\right] \le \prod_{k=K_1}^{K_2} a(k).$$

Now, we are ready to prove the main result for the convergence rate of the autonomous random averaging dynamics.

Lemma 19. *Under Assumption 2-(a) and 3, there exist* $0 < C$ *and* $0 \le \lambda < 1$ *such that for every* $0 \le \tau_1 \le t_1$ *and* $0 \le \tau_2 \le t_2$ *with* $\tau_1 \le \tau_2$*, we have almost surely*

$$\mathbb{E}[\operatorname{diam}(\Phi(t_2,\tau_2))\operatorname{diam}(\Phi(t_1,\tau_1))|\mathcal{F}(\tau_1)] \le C\lambda^{t_1-\tau_1}\lambda^{t_2-\tau_2}.$$

Proof: First, we prove

$$\mathbb{E}[\operatorname{diam}(\Phi(t,\tau))|\mathcal{F}(\tau)] \le \tilde{C}\tilde{\lambda}^{t-\tau}, \tag{4.20}$$

for some $0 < \tilde{C}$ and $0 \le \tilde{\lambda} < 1$. Let $s \triangleq \lceil \frac{\tau}{B} \rceil$ and $K \triangleq \lfloor \frac{t-sB}{n^2B} \rfloor$. Note that

$$\begin{aligned}
\operatorname{diam}(\Phi(t,\tau)) &= \operatorname{diam}\left(\Phi(t,sB+Kn^2B)\left[\prod_{k=1}^{K}\Phi(sB+kn^2B,sB+(k-1)n^2B)\right]\Phi(sB,\tau)\right)\\
&\overset{(a)}{\le} \operatorname{diam}(\Phi(t,sB+Kn^2B))\operatorname{diam}(\Phi(sB,\tau))\\
&\quad\left[\prod_{k=1}^{K}\operatorname{diam}(\Phi(sB+kn^2B,sB+(k-1)n^2B))\right]\\
&\overset{(b)}{\le}\left[\prod_{k=1}^{K}(1-\Lambda(\Phi(sB+kn^2B,sB+(k-1)n^2B)))\right],
\end{aligned}$$

where (a) follows from Lemma 16-(d), and (b) follows from the fact that $\operatorname{diam}(A) \le 1$ for all row-stochastic matrices A and Lemma 16-(c). Therefore, we have

$$\begin{aligned}\mathbb{E}[\mathrm{diam}(\Phi(t,\tau))|\mathcal{F}(\tau)] &\leq \mathbb{E}\left[\prod_{k=1}^{K}(1-\Lambda(\Phi(sB+kn^2B, sB+(k-1)n^2B)))\middle|\mathcal{F}(\tau)\right]\\ &\overset{(a)}{\leq} (1-\theta)^K\\ &\leq \tilde{C}(1-\theta)^{\frac{t-\tau}{n^2B}},\end{aligned}$$

where $\tilde{C} = (1-\theta)^{-1-\frac{1}{n^2}}$ and (a) follows from Lemma 17 and 18 with

$$Y(k) = 1-\Lambda(\Phi(sB+kn^2B, sB+(k-1)n^2B)),$$

and $\tilde{\mathcal{F}} = \mathcal{F}(\tau)$. Since $\theta > 0$, we have $\tilde{\lambda} \triangleq (1-\theta)^{\frac{1}{n^2B}} < 1$.

To prove the main statement, we consider two cases:

(i) intervals $(\tau_1, t_1]$ and $(\tau_2, t_2]$ do not have an intersection, and

(ii) $(\tau_1, t_1]$ and $(\tau_2, t_2]$ intersect.

For case (i), since the two intervals do not overlap, we have $t_1 \leq \tau_2$, and hence, Tower rule implies

$$\begin{aligned}&\mathbb{E}[\mathrm{diam}(\Phi(t_2,\tau_2))\mathrm{diam}(\Phi(t_1,\tau_1))|\mathcal{F}(\tau_1)]\\ &\quad= \mathbb{E}\left[\mathbb{E}[\mathrm{diam}(\Phi(t_2,\tau_2))|\mathcal{F}(\tau_2)]\mathrm{diam}(\Phi(t_1,\tau_1))\middle|\mathcal{F}(\tau_1)\right]\\ &\quad\leq \tilde{C}\tilde{\lambda}^{t_1-\tau_1}\tilde{C}\tilde{\lambda}^{t_2-\tau_2},\end{aligned}$$

which follows from (4.20). For case (ii), let us write the union of the intervals $(\tau_1, t_1]$ and $(\tau_2, t_2]$ as disjoint union of three intervals:

$$(\tau_1, t_1] \cup (\tau_2, t_2] = (s_1, s_2] \cup (s_2, s_3] \cup (s_3, s_4],$$

for $s_1 \le s_2 \le s_3$ where $(s_2,s_3] \triangleq (\tau_1,t_1] \cap (\tau_2,t_2]$, $(s_1,s_2] \cup (s_3,s_4] \triangleq (\tau_1,t_1] \triangle (\tau_2,t_2]$. Using this, it can be verified that

$$
\begin{aligned}
&\mathbb{E}[\mathrm{diam}(\Phi(t_2,\tau_2))\mathrm{diam}(\Phi(t_1,\tau_1))|\mathcal{F}(\tau_1)] \\
&\qquad \stackrel{(a)}{\le} \mathbb{E}[\mathrm{diam}(\Phi(s_4,s_3))\mathrm{diam}^2(\Phi(s_3,s_2))\mathrm{diam}(\Phi(s_2,s_1))|\mathcal{F}(\tau_1)] \\
&\qquad \stackrel{(b)}{\le} \mathbb{E}[\mathrm{diam}(\Phi(s_4,s_3))\mathrm{diam}(\Phi(s_3,s_2))\mathrm{diam}(\Phi(s_2,s_1))|\mathcal{F}(\tau_1)] \\
&\qquad \stackrel{(c)}{\le} \tilde{C}\tilde{\lambda}^{s_2-s_1}\tilde{C}\tilde{\lambda}^{s_3-s_2}\tilde{C}\tilde{\lambda}^{s_4-s_3} \\
&\qquad = \tilde{C}\tilde{\lambda}^{s_2-s_1}2\tilde{C}\sqrt{\tilde{\lambda}}^{2(s_3-s_2)}\tilde{C}\tilde{\lambda}^{s_4-s_3} \\
&\qquad \le \tilde{C}^3\sqrt{\tilde{\lambda}}^{t_1-\tau_1}\sqrt{\tilde{\lambda}}^{t_2-\tau_2},
\end{aligned}
$$

where (a) follows from Lemma 16-(d), (b) follows from $\mathrm{diam}(A) \le 1$ for all row-stochastic matrices A, and (c) follows from (4.20) and Lemma 18. Letting $C \triangleq \max\{\tilde{C}^2, \tilde{C}^3\}$ and $\lambda \triangleq \sqrt{\tilde{\lambda}}$, we arrive at the conclusion. ■

4.6 Averaging Dynamics with Gradient-Flow Like Feedback

In this section, we study the controlled linear time-varying dynamics

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) + \mathbf{u}(t). \tag{4.21}$$

Note that the feedback $\mathbf{u}(t) = -\alpha(t)\mathbf{g}(t)$ leads to the dynamics (4.2). The goal of this section is to establish bounds on the convergence-rate of $\mathrm{d}(\mathbf{x})$ (to zero) in-expectation and almost surely for a class of regularized input $\mathbf{u}(t)$.

We start with the following two lemmas.

Lemma 20. *For dynamics* (4.21) *and every* $0 \leq \tau \leq t$*, we have*

$$\mathrm{d}(\mathbf{x}(t)) \leq \mathrm{diam}(\Phi(t,\tau))\mathrm{d}(\mathbf{x}(\tau)) + \sum_{s=\tau}^{t-1} \mathrm{diam}(\Phi(t,s+1))\mathrm{d}(\mathbf{u}(s)). \tag{4.22}$$

Proof: Note that the general solution for the dynamics (4.21) is given by

$$\mathbf{x}(t) = \Phi(t,\tau)\mathbf{x}(\tau) + \sum_{s=\tau}^{t-1} \Phi(t,s+1)\mathbf{u}(s). \tag{4.23}$$

Therefore, using the sub-linearity property of $\mathrm{d}(\cdot)$ (Lemma 16-(b)), we have

$$\begin{aligned}\mathrm{d}(\mathbf{x}(t)) &\leq \mathrm{d}(\Phi(t,\tau)\mathbf{x}(\tau)) + \sum_{s=\tau}^{t-1} \mathrm{d}(\Phi(t,s+1)\mathbf{u}(s)) \\ &\leq \mathrm{diam}(\Phi(t,\tau))\mathrm{d}(\mathbf{x}(\tau)) + \sum_{s=\tau}^{t-1} \mathrm{diam}(\Phi(t,s+1))\mathrm{d}(\mathbf{u}(s)),\end{aligned}$$

where the last inequality follows from Lemma 16-(a). ■

Lemma 21. *Let* $\{\beta(t)\}$ *be a positive (scalar) sequence such that* $\lim_{t\to\infty} \frac{\beta(t)}{\beta(t+1)} = 1$*. Then for any* $\theta \in [0,1)$*, there exists some* $M > 0$ *such that*

$$\sum_{s=\tau}^{t-1} \beta(s)\theta^{t-s} \leq M\beta(t),$$

for all $t \geq \tau \geq 0$*.*

Proof: The proof is provided in Appendix. ■

To prove the main theorem, we need to study how fast $\mathbb{E}[\mathrm{d}(\mathbf{x}(t))]$ and $\mathbb{E}[\mathrm{d}^2(\mathbf{x}(t))]$ approach to zero when the diameter of the control input $\mathrm{d}(\mathbf{u}(t))$ goes to zero. Since

$$\mathbb{E}[\mathrm{d}^2(\mathbf{x}(t))] \geq \mathbb{E}^2[\mathrm{d}(\mathbf{x}(t))],$$

it suffice to study convergence rate of $\mathbb{E}[\mathrm{d}^2(\mathbf{x}(t))]$.

Lemma 22. *Under Assumptions 2-(a), 3, and 4, if almost surely* $\mathrm{d}(\mathbf{u}(t)) < q\alpha(t)$ *for some* $q > 0$*, then we have,*

$$\frac{\mathbb{E}[\mathrm{d}^2(\mathbf{x}(t))]}{\alpha^2(t)} \leq \hat{M}$$

for some $\hat{M} > 0$ *and all* $t \geq 0$.

Proof: Taking the square of both sides of (4.22), for $t > \tau \geq 0$, we have

$$\begin{aligned}\mathrm{d}^2(\mathbf{x}(t)) \leq \mathrm{diam}^2(\Phi(t,\tau))\mathrm{d}^2(\mathbf{x}(\tau)) + 2\mathrm{diam}(\Phi(t,\tau))\mathrm{d}(\mathbf{x}(\tau))\sum_{s=\tau}^{t-1}\mathrm{diam}(\Phi(t,s+1))\mathrm{d}(\mathbf{u}(s)) \\ + \sum_{s=\tau}^{t-1}\sum_{\ell=\tau}^{t-1}\mathrm{diam}(\Phi(t,s+1))\mathrm{d}(\mathbf{u}(s))\mathrm{diam}(\Phi(t,\ell+1))\mathrm{d}(\mathbf{u}(\ell)).\end{aligned}$$

Taking the expectation of both sides of the above inequality, and using $\mathrm{d}(\mathbf{u}(t)) < q\alpha(t)$ almost surely, we have

$$\begin{aligned}
\mathbb{E}[\mathrm{d}^2(\mathbf{x}(t))] &\leq \mathbb{E}[\mathrm{diam}^2(\Phi(t,\tau))\mathrm{d}^2(\mathbf{x}(\tau))] \\
&\quad + 2\sum_{s=\tau}^{t-1}\mathbb{E}\big[\mathrm{diam}(\Phi(t,\tau))\mathrm{d}(\mathbf{x}(\tau))\mathrm{diam}(\Phi(t,s+1))\mathrm{d}(\mathbf{u}(s))\big] \\
&\quad + \sum_{s=\tau}^{t-1}\sum_{\ell=\tau}^{t-1}\mathbb{E}\big[\mathrm{diam}(\Phi(t,s+1))\mathrm{d}(\mathbf{u}(s))\mathrm{diam}(\Phi(t,\ell+1))\mathrm{d}(\mathbf{u}(\ell))\big] \\
&\leq \mathbb{E}\Big[\mathbb{E}[\mathrm{diam}^2(\Phi(t,\tau))|\mathcal{F}(\tau)]\mathrm{d}^2(\mathbf{x}(\tau))\Big] \\
&\quad + 2\sum_{s=\tau}^{t-1}\mathbb{E}\Big[\mathbb{E}[\mathrm{diam}(\Phi(t,\tau))\mathrm{diam}(\Phi(t,s+1))|\mathcal{F}(\tau)]\mathrm{d}(\mathbf{x}(\tau))\Big]\alpha(s)q \\
&\quad + \sum_{s=\tau}^{t-1}\sum_{\ell=\tau}^{t-1}\mathbb{E}[\mathrm{diam}(\Phi(t,s+1))\mathrm{diam}(\Phi(t,\ell+1))]\alpha(s)\alpha(\ell)q^2.
\end{aligned}$$

Therefore, from Lemma 19, we have

$$\begin{aligned}
\mathbb{E}[\mathrm{d}^2(\mathbf{x}(t))] &\leq C\lambda^{2(t-\tau)}\mathbb{E}[\mathrm{d}^2(\mathbf{x}(\tau))] + \frac{2Cq}{\lambda}\lambda^{t-\tau}\mathbb{E}[\mathrm{d}(\mathbf{x}(\tau))]\sum_{s=\tau}^{t-1}\lambda^{t-s}\alpha(s) \\
&\quad + \frac{Cq^2}{\lambda^2}\sum_{s=\tau}^{t-1}\sum_{\ell=\tau}^{t-1}\alpha(s)\alpha(\ell)\lambda^{t-s}\lambda^{t-\ell} \\
&\leq C\lambda^{2(t-\tau)}\mathbb{E}[\mathrm{d}^2(\mathbf{x}(\tau))] + \frac{2CqM}{\lambda}\lambda^{t-\tau}\mathbb{E}[\mathrm{d}(\mathbf{x}(\tau))]\alpha(t) + \frac{Cq^2M^2}{\lambda^2}\alpha^2(t),
\end{aligned}$$

where the last inequality follows from Lemma 21 and the fact that

$$\sum_{s=\tau}^{t-1}\sum_{\ell=\tau}^{t-1}\alpha(s)\alpha(\ell)\lambda^{t-s}\lambda^{t-\ell} = \left(\sum_{s=\tau}^{t-1}\alpha(s)\lambda^{t-s}\right)^2.$$

Dividing both sides of the above inequality by $\alpha^2(t)$ and noting

$$\frac{\alpha(\tau)}{\alpha(t)}\lambda^{t-\tau} = \prod_{\kappa=\tau}^{t-1}\frac{\alpha(\kappa)}{\alpha(\kappa+1)}\lambda,$$

we have

$$\frac{\mathbb{E}[\mathrm{d}^2(\mathbf{x}(t))]}{\alpha^2(t)} \le C\frac{\mathbb{E}[\mathrm{d}^2(\mathbf{x}(\tau))]}{\alpha^2(\tau)}\left(\prod_{\kappa=\tau}^{t-1}\frac{\alpha(\kappa)}{\alpha(\kappa+1)}\lambda\right)^2 + 2\frac{CqM}{\lambda}\frac{\mathbb{E}[\mathrm{d}(\mathbf{x}(\tau))]}{\alpha(\tau)}\left(\prod_{\kappa=\tau}^{t-1}\frac{\alpha(\kappa)}{\alpha(\kappa+1)}\lambda\right) + \frac{Cq^2M^2}{\lambda^2}.$$

Since $\lim_{\tau\to\infty}\frac{\alpha(\tau)}{\alpha(\tau+1)} = 1$, for any $\hat{\lambda} \in (\lambda, 1)$, there exists $\hat{\tau}$ such that for $\tau \ge \hat{\tau}$, we have

$$\frac{\mu(\tau)}{\mu(\tau+1)}\lambda \le \hat{\lambda}.$$

Therefore,

$$\frac{\mathbb{E}[\mathrm{d}^2(\mathbf{x}(t))]}{\alpha^2(t)} \le C\frac{\mathbb{E}[\mathrm{d}^2(\mathbf{x}(\tau))]}{\alpha^2(\tau)}\hat{\lambda}^{2(t-\tau)} + \frac{2CqM}{\lambda}\frac{\mathbb{E}[\mathrm{d}(\mathbf{x}(\tau))]}{\alpha(\tau)}\hat{\lambda}^{t-\tau} + \frac{Cq^2M^2}{\lambda^2}.$$

Taking the limit of the above inequality, we get

$$\begin{aligned}\limsup_{t\to\infty}\frac{\mathbb{E}[\mathrm{d}^2(\mathbf{x}(t))]}{\alpha^2(t)} &\le \lim_{t\to\infty} C\frac{\mathbb{E}[\mathrm{d}^2(\mathbf{x}(\tau))]}{\alpha^2(\tau)}\hat{\lambda}^{2(t-\tau)} + \lim_{t\to\infty}\frac{2CqM}{\lambda}\frac{\mathbb{E}[\mathrm{d}(\mathbf{x}(\tau))]}{\alpha(\tau)}\hat{\lambda}^{t-\tau} + \frac{Cq^2M^2}{\lambda^2} \\ &= \frac{Cq^2M^2}{\lambda^2}.\end{aligned}$$

As a result, there exists an $\hat{M} > 0$ such that $\frac{\mathbb{E}[\mathrm{d}^2(\mathbf{x}(t))]}{\alpha^2(t)} \le \hat{M}$. ■

To prove the main theorem, we also need to show that $\mathrm{d}(\mathbf{x}(t))$ converges to zero almost surely (as will be proved in Lemma 24). To do so, we apply the following result, which is proved in Appendix.

Lemma 23. *Suppose that $\{D(t)\}$ is a non-negative random (scalar) process such that*

$$D(t+1) \le a(t+1)D(t) + b(t), \quad \textit{almost surely} \tag{4.24}$$

where $\{b(t)\}$ is a deterministic sequence and $\{a(t)\}$ is an adapted process (to $\{\mathcal{F}(t)\}$), such that $a(t) \in [0,1]$ and

$$\mathbb{E}[a(t+1) \mid \mathcal{F}(t)] \leq \tilde{\lambda},$$

almost surely for some $\tilde{\lambda} < 1$ and all $t \geq 0$. Then, if

$$0 \leq b(t) \leq Kt^{-\tilde{\beta}}$$

for some $K, \tilde{\beta} > 0$, we have $\lim_{t\to\infty} D(t)t^{\beta} = 0$, almost surely, for all $\beta < \tilde{\beta}$.

Now, we are ready to show the almost sure convergence $\lim_{t\to\infty} \mathrm{d}(\mathbf{x}(t)) = 0$ (and more) under our connectivity assumption and a regularity condition on the input $\mathbf{u}(t)$ for the controlled averaging dynamics (4.21).

Lemma 24. *Suppose that $\{W(t)\}$ satisfies Assumption 2-(a) and 3. Then, if $\mathrm{d}(\mathbf{u}(t)) < qt^{-\tilde{\beta}}$ almost surely for some $q \geq 0$, we have $\lim_{t\to\infty} \mathrm{d}(\mathbf{x}(t))t^{\beta} = 0$, almost surely, for $\beta < \tilde{\beta}$.*

Proof: From inequality (4.22), we have

$$\begin{aligned}
\mathrm{d}(\mathbf{x}(k)) &\leq \mathrm{diam}(\Phi(k,\tau))\mathrm{d}(\mathbf{x}(\tau)) + \sum_{s=\tau}^{k-1} \mathrm{diam}(\Phi(\tau,s+1))\mathrm{d}(\mathbf{u}(s)) \\
&\overset{(a)}{\leq} \mathrm{diam}(k,\tau))\mathrm{d}(\mathbf{x}(\tau)) + \sum_{s=\tau}^{k-1} qs^{-\tilde{\beta}} \\
&\leq \mathrm{diam}(\Phi(k,\tau))\mathrm{d}(\mathbf{x}(\tau)) + (k-\tau)q\tau^{-\tilde{\beta}}, \qquad (4.25)
\end{aligned}$$

where (a) follows from $\mathrm{diam}(\Phi(.,.)) \leq 1$. Let $C > 0$ and $\lambda \in [0,1)$ be the constants satisfying the statement of Lemma 19. Since $\lambda < 1$, for $T = \lceil |\frac{\log C}{\log \lambda}| \rceil + 1$, we have $\tilde{\lambda} \triangleq C\lambda^{T} < 1$. Then, Lemma 19 implies that

$$\mathbb{E}[\mathrm{diam}(\Phi(T(t+1),Tt))|\mathcal{F}(Tt)] \leq C\lambda^{T} = \tilde{\lambda} < 1. \qquad (4.26)$$

Let $D(t) \triangleq \mathrm{d}(\mathbf{x}(Tt))$. From inequality (4.25), for $\tau = Tt$ and $k = T(t+1)$, we have

$$D(t+1) \leq \mathrm{diam}(\Phi(T(t+1), Tt))D(t) + T^{1-\tilde{\beta}} q t^{-\tilde{\beta}}.$$

Taking conditional expectation of both sides of the above inequality given $\mathcal{F}(Tt)$, we have

$$\mathbb{E}[D(t+1)|\mathcal{F}(Tt)] \leq \mathbb{E}[\mathrm{diam}(\Phi(T(t+1), Tt))|\mathcal{F}(Tt)]D(t) + T^{1-\tilde{\beta}} q t^{-\tilde{\beta}}.$$

By letting $a(t+1) \triangleq \mathrm{diam}(\Phi(T(t+1), Tt))$ and $b(t) \triangleq Tqt^{-\tilde{\beta}}$, we are in the setting of Lemma 23. Therefore, by Inequality (4.26) and $\mathrm{diam}(\Phi(T(t+1), Tt)) \leq 1$, the conditions of Lemma 23 hold, and hence,

$$\lim_{t\to\infty} D(t)t^{\beta} = 0 \tag{4.27}$$

almost surely.

On the other hand, letting $\tau = T\left\lfloor \frac{k}{T} \right\rfloor$ in (4.25) we have

$$\mathrm{d}(\mathbf{x}(k)) \leq D\left(\left\lfloor \frac{k}{T} \right\rfloor\right) + T^{1-\tilde{\beta}} q \left\lfloor \frac{k}{T} \right\rfloor^{-\tilde{\beta}}. \tag{4.28}$$

Note that $y - 1 \leq \lfloor y \rfloor \leq y$. Therefore,

$$k^{\beta} = \left(T\left(\frac{k}{T}\right)\right)^{\beta} \leq \left(T\left(\left\lfloor \frac{k}{T} \right\rfloor + 1\right)\right)^{\beta}.$$

Similarly,

$$\left\lfloor \frac{k}{T} \right\rfloor^{-\tilde{\beta}} \leq \left(\frac{k-T}{T}\right)^{-\tilde{\beta}}.$$

Therefore, using these inequalities and (4.28), we get

$$\begin{aligned}\lim_{k\to\infty} \mathrm{d}(\mathbf{x}(k))k^{\beta} &\leq \lim_{k\to\infty} D\left(\left\lfloor \frac{k}{T}\right\rfloor\right)k^{\beta} + T^{1-\tilde{\beta}} q \left\lfloor \frac{k}{T}\right\rfloor^{-\tilde{\beta}} k^{\beta} \\ &\leq \lim_{k\to\infty} D\left(\left\lfloor \frac{k}{T}\right\rfloor\right)\left(T\left(\left\lfloor \frac{k}{T}\right\rfloor + 1\right)\right)^{\beta} + Tq(k-T)^{-\tilde{\beta}}k^{\beta} \\ &= 0,\end{aligned}$$

where the last equality follows from (4.27) and $\tilde{\beta} > \beta$. ■

4.7 Convergence Analysis of the Main Dynamics

Finally, in this section, we will study the main dynamics (4.2), i.e., the dynamics (4.21) with the feedback policy $\mathbf{u}_i(t) = -\alpha(t)\mathbf{g}_i(t)$ where $\mathbf{g}_i(t) \in \nabla f_i(\mathbf{x}_i(t))$. Throughout this section, we let $\bar{\mathbf{x}} \triangleq \frac{1}{n}e^T\mathbf{x}$ for a vector $\mathbf{x} \in (\mathbb{R}^m)^n$,

First, we prove an inequality (Lemma 26) which plays a key role in the proof of Theorem 4 and to do so, we make use of the following result which is proven as a part of the proof of Lemma 8 (Equation (27)) in [43].

Lemma 25 ([43])**.** *Under Assumption 1, for all $v \in \mathbb{R}^m$, we have*

$$n\langle \bar{\mathbf{g}}(t), \bar{\mathbf{x}}(t) - v\rangle \geq F(\bar{\mathbf{x}}(t)) - F(v) - 2\sum_{i=1}^{n} L_i \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|.$$

Lemma 26. *For the dynamics* (4.2)*, under Assumption 1 and 2, for all $v \in \mathbb{R}^m$, we have*

$$\begin{aligned}\mathbb{E}[\|\bar{\mathbf{x}}(t+1) - v\|^2 | \mathcal{F}(t)] \leq \|\bar{\mathbf{x}}(t) - v\|^2 + \alpha^2(t)\frac{L^2}{n^2} + \sum_{i=1}^{n} \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|^2 \\ - \frac{2\alpha(t)}{n}(F(\bar{\mathbf{x}}(t)) - F(v)) + \frac{4\alpha(t)}{n}\sum_{i=1}^{n} L_i \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|.\end{aligned}$$

Proof: Multiplying $\frac{1}{n}e^T$ from left to both sides of (4.2), we have

$$\begin{aligned}\bar{\mathbf{x}}(t+1) &= \overline{W}(t+1)\mathbf{x}(t) - \alpha(t)\bar{\mathbf{g}}(t) \\ &= \bar{\mathbf{x}}(t) - \alpha(t)\bar{\mathbf{g}}(t) + \overline{W}(t+1)\mathbf{x}(t) - \bar{\mathbf{x}}(t),\end{aligned}$$

where $\overline{W}(t) \triangleq \frac{1}{n}e^T W(t)$. Therefore, we can write

$$\begin{aligned}\|\bar{\mathbf{x}}(t+1) - v\|^2 &= \|\bar{\mathbf{x}}(t) - v - \alpha(t)\bar{\mathbf{g}}(t) + \overline{W}(t+1)\mathbf{x}(t) - \bar{\mathbf{x}}(t)\|^2 \\ &= \|\bar{\mathbf{x}}(t) - v\|^2 + \|\alpha(t)\bar{\mathbf{g}}(t)\|^2 + \|\overline{W}(t+1)\mathbf{x}(t) - \bar{\mathbf{x}}(t)\|^2 \\ &\quad - 2\alpha(t)\left\langle \bar{\mathbf{g}}(t), \overline{W}(t+1)\mathbf{x}(t) - v\right\rangle + 2\left\langle \bar{\mathbf{x}}(t) - v, \overline{W}(t+1)\mathbf{x}(t) - \bar{\mathbf{x}}(t)\right\rangle.\end{aligned}$$

Taking conditional expectation of both sides of the above equality given $\mathcal{F}(t)$, we have

$$\begin{aligned}\mathbb{E}[\|\bar{\mathbf{x}}(t+1) - v\|^2|\mathcal{F}(t)] &= \|\bar{\mathbf{x}}(t) - v\|^2 + \|\alpha(t)\bar{\mathbf{g}}(t)\|^2 + \mathbb{E}\left[\|\overline{W}(t+1)\mathbf{x}(t) - \bar{\mathbf{x}}(t)\|^2|\mathcal{F}(t)\right] \\ &\quad - 2\alpha(t)\left\langle \bar{\mathbf{g}}(t), \mathbb{E}\left[\overline{W}(t+1)\mathbf{x}(t) - v|\mathcal{F}(t)\right]\right\rangle \\ &\quad + 2\left\langle (\bar{\mathbf{x}}(t) - v), \mathbb{E}\left[\overline{W}(t+1)\mathbf{x}(t) - \bar{\mathbf{x}}(t)|\mathcal{F}(t)\right]\right\rangle \\ &= \|\bar{\mathbf{x}}(t) - v\|^2 + \|\alpha(t)\bar{\mathbf{g}}(t)\|^2 + \mathbb{E}\left[\|\overline{W}(t+1)\mathbf{x}(t) - \bar{\mathbf{x}}(t)\|^2|\mathcal{F}(t)\right] \\ &\quad - 2\left\langle \alpha(t)\bar{\mathbf{g}}(t), \bar{\mathbf{x}}(t) - v\right\rangle.\end{aligned}$$

The last equality follows from the assumption that, $W(t+1)$ is doubly stochastic in-expectation and hence,

$$\mathbb{E}[\overline{W}(t+1)|\mathcal{F}(t)] = \frac{1}{n}e^T,$$

which implies

$$\left\langle \bar{\mathbf{g}}(t), \mathbb{E}\left[\overline{W}(t+1)\mathbf{x}(t) - v|\mathcal{F}(t)\right]\right\rangle = \left\langle \alpha(t)\bar{\mathbf{g}}(t), \bar{\mathbf{x}}(t) - v\right\rangle,$$

and

$$\mathbb{E}\left[\overline{W}(t+1)\mathbf{x}(t)-\bar{\mathbf{x}}(t)|\mathcal{F}(t)\right]=0.$$

Note that $\overline{W}(t+1)$ is a stochastic vector (almost surely), therefore, due to the convexity of norm-square $\|\cdot\|^2$, we get

$$\begin{aligned}\|\overline{W}(t+1)\mathbf{x}(t)-\bar{\mathbf{x}}(t)\|^2 &\leq \sum_{i=1}^{n}\overline{W}_i(t+1)\|\mathbf{x}_i(t)-\bar{\mathbf{x}}(t)\|^2\\ &\leq \sum_{i=1}^{n}\|\mathbf{x}_i(t)-\bar{\mathbf{x}}(t)\|^2,\end{aligned}$$

as $\overline{W}_i(t+1)\leq 1$ for all $i\in[n]$. Therefore,

$$\begin{aligned}\mathbb{E}[\|\bar{\mathbf{x}}(t+1)-v\|^2|\mathcal{F}(t)]=&\|\bar{\mathbf{x}}(t)-v\|^2+\|\alpha(t)\bar{\mathbf{g}}(t)\|^2+\mathbb{E}\left[\|\overline{W}(t+1)\mathbf{x}(t)-\bar{\mathbf{x}}(t)\|^2|\mathcal{F}(t)\right]\\ &-2\left\langle\alpha(t)\bar{\mathbf{g}}(t),\bar{\mathbf{x}}(t)-v\right\rangle\\ \leq&\|\bar{\mathbf{x}}(t)-v\|^2+\|\alpha(t)\bar{\mathbf{g}}(t)\|^2+\sum_{i=1}^{n}\|\mathbf{x}_i(t)-\bar{\mathbf{x}}(t)\|^2\\ &-2\left\langle\alpha(t)\bar{\mathbf{g}}(t),\bar{\mathbf{x}}(t)-v\right\rangle.\end{aligned}$$

Finally, Lemma 25 and the fact that

$$\|\bar{\mathbf{g}}(t)\|^2=\frac{1}{n^2}\left\|\sum_{i=1}^{n}\mathbf{g}_i(t)\right\|^2\leq\frac{1}{n^2}\left(\sum_{i=1}^{n}\|\mathbf{g}_i(t)\|\right)^2\leq\frac{L^2}{n^2},$$

complete the proof. ■

Proof of Theorem 4: In order to utilize Robbins-Siegmund Theorem [59] and Lemma 26, for all $t\geq 0$, let Lyapunov function $V(t)\triangleq\|\bar{\mathbf{x}}(t)-z\|^2$ where $z\in\mathcal{Z}$ and $a(t)=0$, and consider $b(t)$ and $c(t)$ which are defined in (4.6). First, note that $a(t),b(t),c(t)\geq 0$ for all t. To invoke the Robbins-Siegmund result (4.7), we need to to prove that $\sum_{t=0}^{\infty}c(t)<\infty$, almost surely. Since

$\sum_{t=0}^{\infty} \alpha^2(t) < \infty$, it is enough to show that

$$\sum_{t=0}^{\infty} \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|^2 < \infty$$

and

$$\sum_{t=0}^{\infty} \alpha(t) \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\| < \infty$$

almost surely, for all $i \in [n]$. From Lemma 16-(e) and Lemma 22, we have

$$\mathbb{E}\left[\frac{\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|}{\alpha(t)}\right] \leq \frac{\mathbb{E}\left[\sqrt{n}\mathrm{d}(\mathbf{x}(t))\right]}{\alpha(t)} \leq \sqrt{n\hat{M}} < \infty$$

for some $\hat{M} > 0$. Therefore, we have

$$\begin{aligned}\lim_{T\to\infty} \mathbb{E}\left[\sum_{t=0}^{T} \alpha(t)\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|\right] &= \lim_{T\to\infty} \mathbb{E}\left[\sum_{t=0}^{T} \alpha^2(t)\frac{\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|}{\alpha(t)}\right] \\ &= \lim_{T\to\infty} \sum_{t=0}^{T} \alpha^2(t)\mathbb{E}\left[\frac{\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|}{\alpha(t)}\right] \\ &\leq \sqrt{n\hat{M}} \sum_{t=0}^{\infty} \alpha^2(t) < \infty,\end{aligned}$$

which is followed by Assumption 4. Similarly, using Lemma 22, there exists some $\hat{M} > 0$ such that

$$\frac{\mathbb{E}\left[\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|^2\right]}{\alpha^2(t)} \leq \frac{\mathbb{E}\left[n\mathrm{d}^2(\mathbf{x}(t))\right]}{\alpha^2(t)} \leq n\hat{M},$$

for all $t \geq 0$, where the first inequality follows from Lemma 16 - (e). Therefore,

$$\lim_{T\to\infty} \mathbb{E}\left[\sum_{t=0}^{T} \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|^2\right] = \lim_{T\to\infty} \mathbb{E}\left[\sum_{t=0}^{T} \alpha^2(t) \frac{\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|^2}{\alpha^2(t)}\right]$$
$$\leq n\hat{M} \sum_{t=0}^{\infty} \alpha^2(t) < \infty,$$

which is followed by Assumption 4. Therefore, using Monotone Convergence Theorem (see e.g., Theorem 1.5.5 in [20]), we have

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \alpha(t) \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|\right] < \infty, \text{ and}$$
$$\mathbb{E}\left[\sum_{t=0}^{\infty} \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|^2\right] < \infty,$$

which implies $\sum_{t=0}^{\infty} \alpha(t) \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\| < \infty$ and $\sum_{t=0}^{\infty} \|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|^2 < \infty$, almost surely.

Now that we showed that $c(t)$ is almost surely a summable sequence, Robbins-Siegmund Theorem implies that almost surely

$$\lim_{t\to\infty} V(t) = \lim_{t\to\infty} \|\bar{\mathbf{x}}(t) - z\|^2 \text{ exists,}$$

and

$$\sum_{t=1}^{\infty} \alpha(t)(F(\bar{\mathbf{x}}(t)) - F(z)) < \infty.$$

For $z \in \mathcal{Z}$, let's define

$$\Omega_z \triangleq \left\{ \omega \middle| \begin{array}{c} \lim_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - z\| \text{ exists,} \\ \sum_{t=1}^{\infty} \alpha(t)(F(\bar{\mathbf{x}}(t,\omega)) - F^*) < \infty \end{array} \right\},$$

where $F^* \triangleq \min_{z\in\mathbb{R}^m} F(z)$. Per Sigmund-Robbins result, we know that $P(\Omega_z) = 1$. Now, let

$\mathcal{Z}_d \subset \mathcal{Z}$ be a countable dense subset of $\mathcal{Z}$ and let

$$\Omega_d \triangleq \bigcap_{z \in \mathcal{Z}_d} \Omega_z.$$

Since $\mathcal{Z}_d$ is a countable set, we have $P(\Omega_d) = 1$ and for $\omega \in \Omega_d$, since

$$\sum_{t=1}^{\infty} \alpha(t)(F(\bar{\mathbf{x}}(t,\omega)) - F^*) < \infty$$

and $\alpha(t)$ is not summable, we have

$$\liminf_{t\to\infty} F(\bar{\mathbf{x}}(t)) = F^*.$$

This fact and the fact that $F(\cdot)$ is a continuous function implies that for all $\omega \in \Omega_d$, we have $\liminf_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - z^*(\omega)\| = 0$ for some $z^*(\omega) \in \mathcal{Z}$. To show this, let $\{\bar{\mathbf{x}}(t_k)\}$ be a sub-sequence that $\lim_{k\to\infty} F(\bar{\mathbf{x}}(t_k,\omega)) = F^*$ (such a sub-sequence depends on the sample path ω). Since $\omega \in \Omega_d$ and

$$\lim_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - \hat{z}\| \quad \text{exists}$$

for some $\hat{z} \in \mathcal{Z}_d$, we conclude that $\{\bar{\mathbf{x}}(t,\omega)\}$ is a bounded sequences. Therefore, $\{\bar{\mathbf{x}}(t_k,\omega)\}$ is also bounded and it has an accumulation point $z^* \in \mathbb{R}^m$ and hence, there is a sub-sequence $\{\bar{\mathbf{x}}(t_{k_\tau},\omega)\}_{\tau\geq 0}$ of $\{\bar{\mathbf{x}}(t_k,\omega)\}_{k\geq 0}$ such that

$$\lim_{\tau\to\infty} \bar{\mathbf{x}}(t_{k_\tau},\omega) = z^*.$$

As a result of continuity of $F(\cdot)$, we have

$$\lim_{\tau\to\infty} F(\bar{\mathbf{x}}(t_{k_\tau})) = F(z^*) = F^*$$

and hence, $z^* \in \mathcal{Z}$. Note that the point $z^* = z^*(\omega)$ depends on the sample path ω.

Since $\mathcal{Z}_d \subseteq \mathcal{Z}$ is dense, there is a sequence $\{q^*(s,\omega)\}_{s\geq 0}$ in $\mathcal{Z}_d$ such that

$$\lim_{s\to\infty} \|q^*(s,\omega) - z^*(\omega)\| = 0.$$

Note that since $\omega \in \Omega_d$, $\lim_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - q^*(s,\omega)\|$ exists for all $s \geq 0$ and we have

$$\begin{aligned}
\lim_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - q^*(s,\omega)\| &= \lim_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - z^*(\omega) + z^*(\omega) - q^*(s,\omega)\| \\
&= \liminf_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - z^*(\omega) + z^*(\omega) - q^*(s,\omega)\| \\
&\leq \liminf_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - z^*(\omega)\| + \|q^*(s,\omega) - z^*(\omega)\| \\
&= \|q^*(s,\omega) - z^*(\omega)\|.
\end{aligned}$$

Therefore, we have

$$\lim_{s\to\infty} \lim_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - q^*(s,\omega)\| = 0. \tag{4.29}$$

On the other hand, we have

$$\begin{aligned}
\limsup_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - z^*(\omega)\| &= \limsup_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - q^*(s,\omega) + q^*(s,\omega) - z^*(\omega)\| \\
&\leq \limsup_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - q^*(s,\omega)\| + \|q^*(s,\omega) - z^*(\omega)\| \\
&= \left(\lim_{t\to\infty} \|\bar{\mathbf{x}}(t,\omega) - q^*(s,\omega)\|\right) + \|q^*(s,\omega) - z^*(\omega)\|.
\end{aligned}$$

Therefore,

$$\begin{aligned}\limsup_{t\to\infty}\|\bar{\mathbf{x}}(t,\omega)-z^*(\omega)\| &= \lim_{s\to\infty}\limsup_{t\to\infty}\|\bar{\mathbf{x}}(t,\omega)-z^*(\omega)\| \\ &\le \lim_{s\to\infty}\lim_{t\to\infty}\|\bar{\mathbf{x}}(t,\omega)-q^*(s,\omega)\| + \lim_{s\to\infty}\|q^*(s,\omega)-z^*(\omega)\| \\ &= 0, \end{aligned} \tag{4.30}$$

where the last equality follows by combining (4.29) and $\lim_{s\to\infty}\|q^*(s,\omega)-z^*(\omega)\|=0$. Note that (4.30), implies that almost surely (i.e., for all $\omega\in\Omega_d$), we have

$$\lim_{t\to\infty}\bar{\mathbf{x}}(t) = z^*(\omega)$$

exists and it belongs to $\mathcal{Z}$.

Finally, according to Assumption 1 and 4, we have

$$\mathrm{d}(\alpha(t)\mathbf{g}(t)) \le 2Kt^{-\beta}\max_{i\in[n]} L_i.$$

Therefore, from Lemma 24, we conclude that $\lim_{t\to\infty}\mathrm{d}(\mathbf{x}(t))=0$ almost surely, and hence, Lemma 16-(e) implies

$$\lim_{t\to\infty}\|\bar{\mathbf{x}}(t)-\mathbf{x}_i(t)\| = 0 \quad \text{almost surely.}$$

Since we almost surely have $\lim_{t\to\infty}\bar{\mathbf{x}}(t)=z^*$ for a random vector z^* supported in $\mathcal{Z}$, we have $\lim_{t\to\infty}\mathbf{x}_i(t)=z^*$ for all $i\in[n]$ almost surely and the proof is complete. ■

4.8 Proof of Required Lemmas

Proof of Lemma 14: Let $\mathbb{T}$ be the set of all collection of edges E such that the graph

$([n],E)$ has a spanning rooted tree, and

$$\mathcal{M}_B(t) \triangleq \bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}(\mathbb{E}[W(\tau)|\mathcal{F}(\tau-1)]).$$

Let $A = \{\omega \mid \mathcal{M}_B(t) \in \mathbb{T}\}$. Since, $\mathcal{M}_B(t)$ almost surely has a spanning rooted tee, we have $P(A) = 1$. Since the cardinality of $\mathbb{T}$ is finite, there exists $\mathcal{T}(t) \in \mathbb{T}$ such that

$$P(R(t)) \geq \frac{P(A)}{|\mathbb{T}|} = \frac{1}{|\mathbb{T}|},$$

where $R(t) = \{\omega \mid \mathcal{M}_B(t) = \mathcal{T}(t)\}$. If $(j(\omega), i(\omega)) \in \mathcal{E}(\mathbb{E}[W(\tau) \mid \mathcal{F}(\tau-1)])(\omega)$, then

$$\mathbb{E}[w_{ij}(\tau) \mid \mathcal{F}(\tau-1)](\omega) > \gamma.$$

Therefore, from Tower identity for conditional expectation (e.g. Theorem 5.1.6. [20]), we have

$$\begin{aligned}
\mathbb{E}[w_{ij}(\tau) \mid \mathcal{F}(tB)] &= \mathbb{E}[\mathbb{E}[w_{ij}(\tau) \mid \mathcal{F}(\tau-1)] \mid \mathcal{F}(tB)] \\
&\geq \mathbb{E}[\mathbb{E}[w_{ij}(\tau) \mid \mathcal{F}(\tau-1)] 1_{R(t)} \mid \mathcal{F}(tB)] \\
&\geq \mathbb{E}[\gamma 1_{R(t)} \mid \mathcal{F}(tB)] \\
&= \gamma P(R(t)) \geq \frac{\gamma}{|\mathbb{T}|},
\end{aligned}$$

and hence, $(j(\omega), i(\omega)) \in \mathcal{E}^{\frac{\gamma}{|\mathbb{T}|}}(\mathbb{E}[W(\tau) \mid \mathcal{F}(tB)])(\omega)$ almost surely. ■

Proof of Lemma 16: For the proof of part (a), let $\mathbf{x}_i^{(k)}$ be the kth coordinate of $\mathbf{x}_i$, and define the vector $y = [y^{(1)}, \ldots, y^{(m)}]^T$ where

$$y^{(k)} = \frac{1}{2}(u^{(k)} + U^{(k)}),$$

with $u^{(k)} = \min_{i\in[n]} \mathbf{x}_i^{(k)}$ and $U^{(k)} = \max_{i\in[n]} \mathbf{x}_i^{(k)}$. Therefore, for $\ell \in [n]$, we have

$$\begin{aligned}
\|\mathbf{x}_\ell - y\|_\infty &= \max_{k\in[m]} |\mathbf{x}_\ell^{(k)} - y^{(k)}| \\
&= \max_{k\in[m]} \left| \mathbf{x}_\ell^{(k)} - \frac{1}{2}(u^{(k)} + U^{(k)}) \right| \\
&\overset{(a)}{\leq} \max_{k\in[m]} \frac{1}{2} \left| u^{(k)} - U^{(k)} \right| \\
&= \frac{1}{2}\mathrm{d}(\mathbf{x}), \qquad (4.31)
\end{aligned}$$

where (a) follows from $u^{(k)} \leq \mathbf{x}_\ell^{(k)} \leq U^{(k)}$. Also, we have

$$\begin{aligned}
\mathrm{d}(A\mathbf{x}) &= \max_{i,j\in[n]} \left\| \sum_{\ell=1}^n a_{i\ell}\mathbf{x}_\ell - \sum_{\ell=1}^n a_{j\ell}\mathbf{x}_\ell \right\|_\infty \\
&= \max_{i,j\in[n]} \left\| \sum_{\ell=1}^n a_{i\ell}(\mathbf{x}_\ell - y) - \sum_{\ell=1}^n a_{j\ell}(\mathbf{x}_\ell - y) + \sum_{\ell=1}^n (a_{j\ell} - a_{i\ell})y \right\|_\infty \\
&= \max_{i,j\in[n]} \left\| \sum_{\ell=1}^n (a_{i\ell} - a_{j\ell})(\mathbf{x}_\ell - y) \right\|_\infty
\end{aligned}$$

where the last equality holds as A is a row-stochastic matrix and hence, $\sum_{\ell=1}^n (a_{j\ell} - a_{i\ell}) = 0$. Therefore,

$$\begin{aligned}
\mathrm{d}(A\mathbf{x}) &= \max_{i,j\in[n]} \left\| \sum_{\ell=1}^n (a_{i\ell} - a_{j\ell})(\mathbf{x}_\ell - y) \right\|_\infty \\
&\overset{(a)}{\leq} \max_{i,j\in[n]} \sum_{\ell=1}^n |a_{i\ell} - a_{j\ell}| \|\mathbf{x}_\ell - y\|_\infty \\
&\overset{(b)}{\leq} \max_{i,j\in[n]} \frac{1}{2}\mathrm{d}(\mathbf{x}) \sum_{\ell=1}^n |a_{i\ell} - a_{j\ell}| \\
&\leq \mathrm{diam}(A)\mathrm{d}(\mathbf{x}),
\end{aligned}$$

where (a) follows from the triangle inequality, and (b) follow from (4.31). For the part (b), we have

$$\begin{aligned}
\mathrm{d}(\mathbf{x}+\mathbf{y}) &= \max_{i,j\in[n]} \|(\mathbf{x}+\mathbf{y})_i-(\mathbf{x}+\mathbf{y})_j\|_\infty \\
&= \max_{i,j\in[n]} \|\mathbf{x}_i-\mathbf{x}_j+\mathbf{y}_i-\mathbf{y}_j\|_\infty \\
&\overset{(b)}{\leq} \max_{i,j\in[n]} (\|\mathbf{x}_i-\mathbf{x}_j\|_\infty+\|\mathbf{y}_i-\mathbf{y}_j\|_\infty) \\
&\leq \max_{i,j\in[n]} \|\mathbf{x}_i-\mathbf{x}_j\|_\infty+\max_{i,j\in[n]} \|\mathbf{y}_i-\mathbf{y}_j\|_\infty \\
&= \mathrm{d}(\mathbf{x})+\mathrm{d}(\mathbf{y}),
\end{aligned}$$

where (b) follows from the triangle inequality.

For the proof of part (c), we have

$$\begin{aligned}
\mathrm{diam}(A) &= \max_{i,j\in[n]} \sum_{\ell=1}^{n} \frac{1}{2}|a_{i\ell}-a_{j\ell}| \\
&= \max_{i,j\in[n]} \sum_{\ell=1}^{n} \left(\frac{1}{2}(a_{i\ell}+a_{j\ell})-\min\{a_{i\ell},a_{j\ell}\}\right) \\
&\overset{(a)}{=} \max_{i,j\in[n]} 1-\sum_{\ell=1}^{n} \min\{a_{i\ell},a_{j\ell}\} \\
&= 1-\min_{i,j\in[n]} \sum_{\ell=1}^{n} \min\{a_{i\ell},a_{j\ell}\} \\
&= 1-\Lambda(A),
\end{aligned}$$

where (a) follows from the fact that A is row-stochastic. The proof of part (d) follows from part (c) and Lemma 15.

For the part (e), due to the convexity of $\|\cdot\|$, we have

$$\left\|\mathbf{x}_i-\sum_{j=1}^{n}\pi_j\mathbf{x}_j\right\| \leq \sum_{j=1}^{n}\pi_j\|\mathbf{x}_i-\mathbf{x}_j\| \leq \sum_{j=1}^{n}\pi_j\sqrt{n}\mathrm{d}(\mathbf{x}) = \sqrt{n}\mathrm{d}(\mathbf{x}).$$

■

Proof of Lemma 18: We prove by induction on K_2. By the assumption, the lemma is true for $K_2 = K_1$. For $K_2 > K_1$, from Tower rule, we have

$$\begin{aligned}
\mathbb{E}\left[\prod_{k=K_1}^{K_2+1} Y(k) \middle| \tilde{\mathcal{F}}\right] &= \mathbb{E}\left[\mathbb{E}\left[\prod_{k=K_1}^{K_2+1} Y(k) \middle| \mathcal{F}(K_2)\right] \middle| \tilde{\mathcal{F}}\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y(K_2+1)|\mathcal{F}(K_2)] \prod_{k=K_1}^{K_2} Y(k) \middle| \tilde{\mathcal{F}}\right] \\
&\leq \mathbb{E}\left[a(K_2+1) \prod_{k=K_1}^{K_2} Y(k) \middle| \tilde{\mathcal{F}}\right] \\
&\leq \prod_{k=K_1}^{K_2+1} a(k).
\end{aligned}$$

■

Proof of Lemma 21: Consider $\hat{\tau} \geq 0$ such that $\hat{\theta} \triangleq \sup_{t \geq \hat{\tau}} \frac{\beta(t)}{\beta(t+1)} \theta < 1$, and let

$$D(t) \triangleq \sum_{s=\tau}^{t-1} \beta(s) \theta^{t-s}.$$

Dividing both sides by $\beta(t) > 0$, for $t > \hat{\tau}$, we have

$$\begin{aligned}
\frac{D(t)}{\beta(t)} &= \sum_{s=\tau}^{t-1} \frac{\beta(s)}{\beta(t)} \theta^{t-s} \\
&= \sum_{s=\tau}^{t-1} \prod_{\kappa=s}^{t-1} \frac{\beta(\kappa)}{\beta(\kappa+1)} \theta \\
&\leq \sum_{s=\tau}^{\hat{\tau}-1} \prod_{\kappa=s}^{t-1} \frac{\beta(\kappa)}{\beta(\kappa+1)} \theta + \sum_{s=\hat{\tau}}^{t-1} \hat{\theta}^{t-s} \\
&= \sum_{s=\tau}^{\hat{\tau}-1} \prod_{\kappa=s}^{t-1} \frac{\beta(\kappa)}{\beta(\kappa+1)} \theta + \sum_{k=1}^{t-\hat{\tau}} \hat{\theta}^{k} \\
&\leq \sum_{s=\tau}^{\hat{\tau}-1} \prod_{\kappa=s}^{t-1} \frac{\beta(\kappa)}{\beta(\kappa+1)} \theta + \frac{\hat{\theta}}{1-\hat{\theta}}.
\end{aligned}$$

Let

$$M_1 \triangleq \sup_{t>\hat{\tau}} \sup_{\hat{\tau}\geq\tau\geq 0} \sum_{s=\tau}^{\hat{\tau}-1} \prod_{\kappa=s}^{t-1} \frac{\beta(\kappa)}{\beta(\kappa+1)} \theta + \frac{\hat{\theta}}{1-\hat{\theta}}.$$

Note that

$$\lim_{t\to\infty} \prod_{\kappa=s}^{t-1} \frac{\beta(\kappa)}{\beta(\kappa+1)} \theta \leq \lim_{t\to\infty} \hat{\theta}^{t-\hat{\tau}} \prod_{\kappa=s}^{\hat{\tau}-1} \frac{\beta(\kappa)}{\beta(\kappa+1)} \theta = 0.$$

Therefore

$$\sup_{t\geq\tau} \prod_{\kappa=s}^{t-1} \frac{\beta(\kappa)}{\beta(\kappa+1)} \theta < \infty,$$

and hence, $M_1 < \infty$. Thus, $D(t) \leq \max\{M_1, M_2\}\beta(t)$, where

$$M_2 \triangleq \max_{\hat{\tau}\geq t\geq\tau\geq 0} \sum_{s=\tau}^{t-1} \frac{\beta(s)}{\beta(t)} \theta^{t-s},$$

and the proof is complete. ■

Lemma 27. *Consider a non-negative random process* $a(t)$ *such that*

$$\mathbb{E}[a(t+1) \mid \mathcal{F}(t)] \le \tilde{\lambda} \quad \textit{almost surely,}$$

for some $\tilde{\lambda} < 1$ *and all* $t \ge 0$*. For* λ *satisfying* $\tilde{\lambda} < \lambda < 1$*, define the sequence of stopping-times* $\{t_s\}_{s\ge 0}$ *by*

$$t_s \triangleq \inf\{t > t_{s-1} | a(t) \le \lambda\},$$

with $t_0 = 0$*. Then,* $\lim_{s\to\infty}(t_{s+1} - t_s)t_s^{-\beta} = 0$ *almost surely for all* $\beta > 0$*.*

Proof: Let us define the martingale $S(t)$ by

$$S(t) = S(t-1) + \left(1_{\{a(t)>\lambda\}} - \mathbb{E}[1_{\{a(t)>\lambda\}} | \mathcal{F}(t-1)]\right),$$

where $S(0) = 0$. Noting $|S(t+1) - S(t)| \le 1$, from Azuma's inequality (see e.g., Theorem 7.2.1 in [1]), we have

$$P(S(t+\sigma) - S(t) > \sigma\rho) \le \exp\left(-\frac{\sigma^2\rho^2}{2\sigma}\right), \tag{4.32}$$

for all $\sigma \in \mathbb{N}$ and $\rho \in (0,1)$. For $\theta > 0$, let use define the sequences of events

$$A_\theta(t) \triangleq \left\{\omega \middle| S(t + \lfloor\theta t^\beta\rfloor) - S(t) > \lfloor\theta t^\beta\rfloor\rho\right\}.$$

From (4.32), we have

$$P(A_\theta(t)) \le \exp\left(-\frac{1}{2}(\theta t^\beta - 1)\rho^2\right),$$

implying $\sum_{t=1}^{\infty} P(A_\theta(t)) < \infty$ as $\exp\left(-t^\beta\right) \le \frac{M}{t^2}$ for sufficiently large M (depending on β).

Therefore, the Borel–Cantelli Theorem (see e.g., Theorem 2.3.1 in [20]) implies that for all $\theta > 0$

$$P(\{A_\theta(t) \quad \text{i.o.}\}) = 0.$$

For $\theta > 0$, let the sequences of events

$$B_\theta(t) \triangleq \left\{ \omega \middle| \frac{\hat{t} - t}{t^\beta} > \theta \text{ where } \hat{t} = \inf\{\tau > t | a(\tau) \leq \lambda\} \right\}.$$

We show that $B_\theta(t) \subset A_\theta(t)$ for all t, θ. Fix a constant $\rho \in (0,1)$ such that $1 - \frac{\tilde{\lambda}}{\lambda} > \rho$. Since $\mathbb{E}[a(\tau) \mid \mathcal{F}(\tau-1)] \leq \tilde{\lambda}$, we have

$$\mathbb{E}[\lambda 1_{\{a(\tau) \geq \lambda\}} \mid \mathcal{F}(\tau-1)] \leq \mathbb{E}[a(\tau) \mid \mathcal{F}(\tau-1)] \leq \tilde{\lambda} < \lambda(1-\rho)$$

and hence,

$$\mathbb{E}[1_{\{a(\tau) \geq \lambda\}} \mid \mathcal{F}(\tau-1)] < 1 - \rho. \tag{4.33}$$

Let $\sigma(t) \triangleq \lfloor \theta t^\beta \rfloor$. If $\hat{t} - t > \theta t^\beta$, then

$$\begin{aligned} S(t+\sigma(t)) - S(t) &= \sigma(t) - \sum_{\tau=t+1}^{t+\sigma(t)} \mathbb{E}[1_{\{a(\tau) > \lambda\}} | \mathcal{F}(\tau-1)] \\ &> \sigma(t) - \sigma(t)(1-\rho) = \sigma(t)\rho, \end{aligned}$$

which follows from (4.33). Therefore, we have $B_\theta(t) \subset A_\theta(t)$, and hence, $P(\{B_\theta(t) \quad \text{i.o.}\}) = 0$ for all $\theta > 0$.

Finally, by contradiction, we show that $\lim_{s\to\infty}(t_{s+1} - t_s)t_s^{-\beta} = 0$. Since, if

$$\lim_{s\to\infty}(t_{s+1} - t_s)t_s^{-\beta} \neq 0$$

almost surely, then $\limsup_{s\to\infty}(t_{s+1}-t_s)t_s^{-\beta} > 0$ almost surely, and hence,

$$P\left(\limsup_{s\to\infty}(t_{s+1}-t_s)t_s^{-\beta} > \epsilon\right) > 0$$

for some $\epsilon > 0$. Therefore, $P(\{B_\epsilon(t) \quad \text{i.o.}\}) > 0$, which is a contradiction. ∎

Proof of Lemma 23: Let $t_s \triangleq \inf\{t > t_{s-1} | a(t) \leq \lambda\}$ and $t_0 = 0$ for some $\tilde{\lambda} < \lambda < 1$, and

$$c(s) \triangleq \sum_{\tau=t_s+1}^{t_{s+1}-1} b(\tau).$$

Also, define

$$A \triangleq \left\{\omega \middle| \lim_{s\to\infty} \frac{t_{s+1}-t_s}{t_s^{\min\{\tilde{\beta}-\beta,1\}}} = 0\right\}.$$

Note that Lemma 27 implies $P(A) = 1$. On the other hand, using (4.24), we have

$$D(t_{s+1}) \leq D(t_s) \prod_{\ell=t_s+1}^{t_{s+1}} a(\ell) + \sum_{\tau=t_s}^{t_{s+1}-1} b(\tau) \prod_{\ell=\tau+2}^{t_{s+1}} a(\ell) \leq D(t_s)\lambda + c(s),$$

where the last inequality follows from $a(t) \in [0,1]$ and $a(t_{s+1}) \leq \lambda$. Letting $R(t) = D(t)t^\beta$, we have

$$R(t_{s+1}) \leq \left(\frac{t_{s+1}}{t_s}\right)^\beta R(t_s)\lambda + c(s)t_{s+1}^\beta.$$

Note that, for $\omega \in A$, we have

$$\lim_{s\to\infty} \frac{t_{s+1}}{t_s} = \lim_{s\to\infty} 1 + \frac{t_{s+1}-t_s}{t_s} = 1. \tag{4.34}$$

As a result, for any $\hat{\lambda} \in (\lambda, 1)$, there exists $\hat{s}$ such that for $s \geq \hat{s}$, we have $\left(\frac{t_{s+1}}{t_s}\right)^{\beta} \lambda \leq \hat{\lambda}$, and hence

$$R(t_{s+1}) \leq R(t_s)\hat{\lambda} + c(s) t_{s+1}^{\beta}.$$

Therefore,

$$R(t_s) \leq \hat{\lambda}^{s-\hat{s}} R(t_{\hat{s}}) + \sum_{\tau=\hat{s}}^{s-1} c(\tau) t_{\tau+1}^{\beta} \hat{\lambda}^{s-\tau-1}.$$

Taking the limits of the both sides, we have

$$\limsup_{s\to\infty} R(t_s) \leq \limsup_{s\to\infty} \hat{\lambda}^{s-\hat{s}} R(t_{\hat{s}}) + \sum_{\tau=\hat{s}}^{s-1} c(\tau) t_{\tau+1}^{\beta} \hat{\lambda}^{s-\tau-1} = \lim_{s\to\infty} c(s) t_{s+1}^{\beta},$$

which is implied by Lemma 3.1-(a) in [51]. For $\omega \in A$, we have

$$\begin{aligned} \lim_{s\to\infty} c(s) t_{s+1}^{\beta} &\overset{(a)}{\leq} \lim_{s\to\infty} \frac{K(t_{s+1} - t_s)}{t_s^{\tilde{\beta}}} t_{s+1}^{\beta} \\ &= \lim_{s\to\infty} \frac{K(t_{s+1} - t_s)}{t_s^{\tilde{\beta}-\beta}} \frac{t_{s+1}^{\beta}}{t_s^{\beta}} \\ &\overset{(b)}{=} \lim_{s\to\infty} \frac{K(t_{s+1} - t_s)}{t_s^{\tilde{\beta}-\beta}} = 0. \end{aligned} \tag{4.35}$$

where (a) follows from $b(t) \leq K t^{-\tilde{\beta}}$, and (b) follows from (4.34). Therefore, $\lim_{s\to\infty} R(t_s) = 0$. Now for any $t > 0$ with $t_s \leq t < t_{s+1}$, let $\sigma(t) = s$. By the definition of $R(t)$, we have

$$R(t) \leq \left(\frac{t}{\sigma(t)}\right)^{\beta} R(t_{\sigma(t)}) + c(\sigma(t)) t_{\sigma(t)+1}^{\beta}.$$

Therefore, $\lim_{s\to\infty} R(t_s) = 0$ and Inequality (4.35) imply

$$\limsup_{t\to\infty} R(t) \leq \lim_{t\to\infty} \left(\frac{t}{\sigma(t)}\right)^{\beta} R(t_{\sigma(t)}) + c(\sigma(t))t^{\beta}_{\sigma(t)+1} = 0,$$

which is the desired conclusion as $R(t) = D(t)t^{\beta}$. ■

Chapter 4, in full, is a reprint of the material as it appears in A. Aghajan and B. Touri, Distributed optimization over dependent random networks, being submitted for publication. The dissertation author was the primary investigator and author of this paper.

Chapter 5

Distributed Gradient-Tracking Algorithm

In this chapter, we study distributed gradient-tracking algorithm over a broad class of dependent random networks, where similar to Chapter 4, the randomness is not only time-varying but also possibly dependent on the past. For the strongly convex and smooth local objective functions, we show that the distributed gradient-tracking algorithm at each node converges to a global optimizer at a geometric rate almost surely if the weight matrices corresponding to the optimizer are row-stochastic almost surely, and the weight matrices corresponding to the gradient tracking are doubly stochastic almost surely, and satisfy a connectivity assumption over time. Also, we show that column-stochasticity is necessary for the weight matrices corresponding to the gradient. It is worth mentioning that to derive the main results, we study the linear dynamics with non-negative random matrices. We find a new sufficient condition on the convergence of the state values to zero for such dynamics, which is presented in Lemma 29.

This chapter is organized as follows: in Section 5.1, we first formulate the problem of interest. Then, we state the main results of this chapter, which is a necessary condition (Theorem 6) in Section 5.2 and a sufficient condition (Theorem 7) in Section 5.3 for the convergence of the distributed gradient-tracking algorithm to the optimizer of $F(\cdot)$ over dependent random networks. In Section 5.3, after stating Theorem 7, we discuss implications of this theorem, and finally, we prove it.

5.1 Problem Formulation

Similar to Chapter 4, the goal of this chapter is to solve

$$\arg\min_{z\in\mathbb{R}^m} F(z) \tag{5.1}$$

distributively where $F(z) \triangleq \sum_{i=1}^{n} f_i(z)$ and the function $f_i : \mathbb{R}^m \to \mathbb{R}$ is the cost function of node i. However, in this chapter, we assume that the functions are strongly convex and smooth. More precisely, we consider the following assumption on the objective functions in this chapter.

Assumption 5 (Assumption on the Objective Function)**.** *We assume that:*

(a) All f_is are β-smooth functions over $\mathbb{R}^m$.

(b) All f_is are α-strongly convex functions over $\mathbb{R}^m$.

As in the previous chapter, the goal is to find distributed dynamics $\mathbf{x}_i(t)$s such that $\lim_{t\to\infty}\mathbf{x}_i(t) = z^*$ for all $i \in [n]$ where z^* is a minimizer of $F(\cdot)$. In Chapter 4, we discussed about averaging-based distributed optimization solvers, i.e.,

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \eta(t)\mathbf{g}(t), \tag{5.2}$$

where $\{\eta(t)\}$ is a step-size sequence, and $\{W(t)\}$ is a sequence of doubly stochastic matrices. If the objective function is β-smooth and α-strongly convex, the centralized variation of (5.2), which is the gradient descent algorithm, geometrically converges to the minimizer with a properly chosen constant step-size (see e.g., Theorem 2.1.15 in [48]). However, it was shown that the dynamics (5.2) cannot converge to the minimizer with a constant step-size. Furthermore, Theorem

6 in [49] proves that any dynamics similar to (5.2) cannot converge to the minimizer geometrically. To remedy this, in [82, 49, 44], the following dynamics is proposed

$$\mathbf{x}(t+1) = W^x(t+1)\mathbf{x}(t) - \eta\mathbf{s}(t), \tag{5.3}$$

$$\mathbf{s}(t+1) = W^s(t+1)\mathbf{s}(t) + \mathbf{g}(t+1) - \mathbf{g}(t), \tag{5.4}$$

where both $\{W^x(t)\}$ and $\{W^s(t)\}$ are doubly stochastic sequences and

$$\mathbf{g}_i(t) \triangleq \nabla f_i(\mathbf{x}_i(t)) \in \mathbb{R}^m,$$

for $i \in [n]$. Also, the geometric convergence rate of the dynamics (5.3)-(5.4) is established there. Later, in [81] (time-invariant setting) and [60] (time-varying setting), it was shown that having a row-stochastic sequence of $\{W^x(t)\}$ and a column-stochastic sequence of $\{W^s(t)\}$ is enough for the convergence of the dynamics (5.3)-(5.4) to the minimizer.

5.2 Necessary Condition

In this chapter, we consider the random variation of (5.3)-(5.4), i.e., when $\{W^x(t)\}$ and $\{W^s(t)\}$ are sequences of possibly (time-) dependent random matrices. This is motivated by our study in Chapter 4, where we show that the doubly stochastic assumption is not necessary for the averaging-based distributed optimization algorithm (5.2) for a broad class of random sequences that are row-stochastic almost surely by only column-stochasticity in expectation (as opposed to almost surely). This observation facilitates new algorithm designs for distributed optimization and also implies robustness of (5.2) to a broad class of link failure. It is natural to conjecture that a similar extension should hold for (5.3)-(5.4), i.e., for the algorithm/dynamics to work, it is sufficient to have almost surely row-stochastic $\{W^x(t)\}$ and $\{W^s(t)\}$ that are only column stochastic in-expectation. However, the following result shows the algorithm does not converge to the minimizer if $e^T W^s(t) \neq e^T$ for all $t > 0$ almost surely.

Theorem 6 (Necessary Condition)**.** *Consider the parameterized family of scalar quadratic functions $f_i : \mathbb{R} \to \mathbb{R}$ (i.e., $m = 1$) defined by $f_i(z) = (z - c_i)^2$ where $c_i \in \mathbb{R}$ and let $z^* = \frac{1}{n}\sum_{i=1}^n c_i$ be the minimizer of $F(z) = \sum_{i=1}^n f_i(z)$. If for all initial time $t_0 > 0$, initial condition $\mathbf{x}(t_0) \in \mathbb{R}^n$, and $c \in \mathbb{R}^n$, and some $i \in [n]$, we have $\lim_{t\to\infty} \mathbf{x}_i(t) = z^*$ for the dynamics* (5.3)-(5.4)*, then $e^T W^s(t) = e^T$ for all $t > 0$.*

Proof: Let $\mathbf{u}(t)$ be the concatenation of $\mathbf{x}(t)$ and $\mathbf{s}(t)$, i.e.,

$$\mathbf{u}(t) \triangleq \begin{bmatrix} \mathbf{x}(t) \\ \mathbf{s}(t) \end{bmatrix}.$$

Noting $\mathbf{g}(t) = 2(\mathbf{x}(t) - c)$, we have $\mathbf{u}(t+1) = \Gamma(t+1)\mathbf{u}(t)$ where

$$\Gamma(t) = \begin{bmatrix} W^x(t) & -\eta I \\ 2W^x(t) - 2I & W^s(t) - 2\eta I \end{bmatrix}.$$

Let $\Phi(t,\tau) = [\phi_{ij}(t,\tau)]$ be the transition matrix of $\Gamma(t)$, and $\Phi_i(t,\tau)$ be the ith row of $\Phi(t,\tau)$. Since $\lim_{t\to\infty} \mathbf{x}_i(t) = z^*$ for all initial time $\tau > 0$, we have

$$\lim_{t\to\infty} \Phi_i(t,\tau)\mathbf{u}(\tau) = \frac{1}{n}\sum_{j=1}^n c_j.$$

Therefore, noting

$$\mathbf{u}(\tau) \triangleq \begin{bmatrix} \mathbf{x}(\tau) \\ 2(\mathbf{x}(\tau) - c) \end{bmatrix},$$

we have

$$\lim_{t\to\infty}\sum_{j=1}^{n}(\phi_{ij}(t,\tau)+2\phi_{ij+n}(t,\tau))\mathbf{x}_j(\tau)-\sum_{j=1}^{n}2\phi_{ij+n}(t,\tau)c_j=\frac{1}{n}\sum_{j=1}^{n}c_j. \tag{5.5}$$

Since (5.5) holds for all initial time $\tau > 0$, initial condition $\mathbf{x}(\tau)\in\mathbb{R}^n$, and $c\in\mathbb{R}^n$, therefore for all $j\in[n]$

$$\lim_{t\to\infty}(\phi_{ij}(t,\tau)+2\phi_{ij+n}(t,\tau))=0, \text{ and}$$
$$\lim_{t\to\infty}-2\phi_{ij+n}(t,\tau)=\frac{1}{n},$$

and hence,

$$\begin{aligned}\lim_{t\to\infty}\Phi_i(t,\tau)&=\Big[\underbrace{\frac{1}{n},\ldots,\frac{1}{n}}_{n},\underbrace{-\frac{1}{2n},\ldots,-\frac{1}{2n}}_{n}\Big]\\&=\left[\frac{1}{n}e^T,-\frac{1}{2n}e^T\right],\end{aligned}$$

for all $\tau > 0$. Therefore

$$\lim_{t\to\infty}\Phi_i(t,\tau)\Gamma(\tau)=\lim_{t\to\infty}\Phi_i(t,\tau-1),$$

implies

$$\frac{1}{n}e^T(-\eta I)-\frac{1}{2n}e^T(W^s(\tau)-2\eta I)=-\frac{1}{2n}e^T,$$

and hence, we need to have $e^TW^s(\tau)=e^T$, which completes the proof. ■

Note that the above result holds for any given deterministic sequence of $\{W^s(t)\}$ (and

$\{W^x(t)\}$). Therefore, for almost sure convergence to z^* for random $\{W^x(t)\}$ and $\{W^s(t)\}$, we should have $e^T W^s(t) = e^T$ for all $t > 0$ almost surely. This motivates the following assumption on the stochasticity of the random weight matrix sequences.

Assumption 6 (Stochastic Assumption)**.** *We assume that* $\{W^x(t)\}$ *and* $\{W^s(t)\}$ *satisfy*

(a) For all $t > 0$*,* $W^x(t)$ *is row-stochastic almost surely.*

(b) For all $t > 0$*,* $W^s(t)$ *is doubly stochastic almost surely.*

5.3 Sufficient Condition

Similar to other works in random networks, our goal is to ensure that $\lim_{t\to\infty} \mathbf{x}_i(t) = z^*$ *almost surely* for all $i \in [n]$, where z^* is the minimizer of $F(\cdot)$. To ensure this, the associated graph sequences of $\{W^x(t)\}$ and $\{W^s(t)\}$ both need to satisfy the Conditional B-Connectivity conditions, i.e., Assumption 3, which is introduced in Chapter 4. With these background information and assumptions, we are ready to present the main result of this chapter.

Theorem 7 (Sufficient Condition)**.** *Suppose that the Assumptions 5-6 hold on the model and the dynamics* (5.3)-(5.4)*, and* $\{W^x(t)\}$ *and* $\{W^s(t)\}$ *satisfy Assumption 3. Then, as* $t \to \infty$*,* $\mathbf{x}_i(t)$ *converges geometrically to* z^* *almost surely for all* $i \in [n]$ *and all initial conditions* $\mathbf{x}_i(0) \in \mathbb{R}^m$*, where* z^* *is the minimizer of* $F(\cdot)$*. More precisely, there exists* $0 < \rho < 1$ *(independent of the sample point/path) such that*

$$\lim_{t\to\infty} \frac{\|\mathbf{x}_i(t) - z^*\|}{\rho^t} < \infty,$$

for all $i \in [n]$ *and all initial conditions* $\mathbf{x}_i(0) \in \mathbb{R}^m$*.*

5.3.1 Implications

Before proving the sufficient condition (Theorem 7), let us discuss an example of how one can synthesize distributed and random dynamics that enables fast distributed optimization algorithms using our main result.

For this, consider a connected *undirected* network[1] $\mathcal{G} = ([n], E)$. Consider a simple link-failure model where each edge of the graph can be dropped independent of the past and other edges. Let $p_{ij}(\tau)$ be the probability of link-failure of edge[2] $\{i, j\}$ at time τ where $p_{ij}(\tau) \leq \bar{p}$ for some $1 > \bar{p}$, and all τ and $\{j, i\} \in E$. Define the time window $\Delta(t) \triangleq \{tB+1, \ldots, (t+1)B\}$, and let $\mathcal{N}_i \triangleq \{j \in [n] \mid (i, j) \in E\}$ be the neighbors of node $i \in [n]$. To implement the distributed gradient-tracking algorithm, we need to construct the matrix sequences $\{W^x(\tau)\}$ and $\{W^s(\tau)\}$ satisfying Assumptions 6-3.

For constructing the sequence $\{W^x(\tau)\}$, for each $t > 0$, each node $j \in [n]$ picks at least one time instance $\tau \in \Delta(t)$ and shares $\mathbf{x}_j(\tau)$ with one of its neighbors, which is chosen uniformly randomly. Moreover, at any time $\tau \in \Delta(t)$, any node $\ell \in [n]$ updates $\mathbf{x}_\ell$ as follows

$$\mathbf{x}_\ell(\tau+1) = \frac{1}{|\mathcal{N}_\ell^x(\tau)|} \sum_{j \in \mathcal{N}_\ell^x(\tau)} \mathbf{x}_j(\tau) - \eta \mathbf{s}_i(\tau),$$

where

$$\mathcal{N}_\ell^x(\tau) \triangleq \{\ell\} \cup \{j \mid \text{node } j \text{ sends } \mathbf{x}_j(\tau) \text{ to node } \ell\}.$$

Therefore, for $i, j \in [n]$ with $i \neq j$, we have

$$w_{ij}^x(\tau) = \begin{cases} \frac{1}{|\mathcal{N}_i^x(\tau)|}, & \text{if } j \in \mathcal{N}_i^x(\tau) \\ 0, & \text{otherwise} \end{cases},$$

and $w_{ii}^x(\tau) = 1 - \sum_{j \neq i} w_{ij}^x(\tau)$.

Unlike $W^x(\tau)$, $W^s(\tau)$ needs to be doubly stochastic, and hence, there are more consider-

[1] The graphs do not need to be time-invariant, and this example can be extended to processes over underlying time-varying graphs.

[2] Since the underlying graph is assumed to be undirected, the failed graphs are also assumed to be undirected.

ations for constructing $\{W^s(\tau)\}$. To do so, in our algorithm, each node $j \in [n]$ decides to request information for at least one time instance $\tau \in \Delta(t)$. In that case, j shares $\mathbf{s}_j(\tau)$ with a uniformly and randomly chosen neighbor of itself, say node $i \in [n]$, and requests $\mathbf{s}_i(\tau)$ from i. In this case, we say that j is submitting a request to i. Let

$$\mathcal{N}_i^s(\tau) \triangleq \{j \mid \text{node } j \text{ sends } \mathbf{s}_j(\tau) \text{ to node } i \text{ and requests } \mathbf{s}_i(\tau)\}$$

denote all the nodes that submit requests to node i at time τ. Then, at time[3] τ, node i picks a uniformly chosen random node $j_i(\tau) \in \mathcal{N}_i^s(\tau)$ from the received requests and sends $\mathbf{s}_i(\tau)$ to $j_i(\tau)$. For consistency, let $j_i(\tau) = 0$ if node i does not receive any request at time τ and let $T^s(\tau) = \{i \mid j_i(\tau) \neq 0\}$. Let $R^s(\tau)$ be the collection of all nodes that submit a request at time τ whose requests have been approved, i.e., $R^s(\tau) = \{j_i(\tau) \mid i \in T^s(\tau)\}$. We simply refer to the nodes in $R^s(\tau)$ as requesters and nodes in $T^s(\tau)$ as responders. Then, at any time $\tau \in \Delta(t)$, any node $\ell \in [n]$ updates its gradient tracking information conditioned on one of the three cases

(i) $\ell \in (R^s(\tau) \setminus T^s(\tau)) \cup (T^s(\tau) \setminus R^s(\tau))$: Here, ℓ is a requester or a responder, but not both. Let $j \in [n]$ be the node that has requested information or responded to ℓs request. In this case, we let

$$\mathbf{s}_\ell(\tau+1) = \frac{2}{3}\mathbf{s}_\ell(\tau) + \frac{1}{3}\mathbf{s}_j(\tau) + \mathbf{g}_\ell(\tau+1) - \mathbf{g}_\ell(\tau).$$

(ii) $\ell \in R^s(\tau) \cap T^s(\tau)$: Here, ℓ is both a requester from a node $i \in [n]$, and responding to a request from node $j \in [n]$. In this case, we let

$$\mathbf{s}_\ell(\tau+1) = \frac{1}{3}(\mathbf{s}_\ell(\tau) + \mathbf{s}_j(\tau) + \mathbf{s}_i(\tau)) + \mathbf{g}_\ell(\tau+1) - \mathbf{g}_\ell(\tau).$$

[3]The algorithm still works, even if node i chooses uniformly randomly τ_i from $\Delta(t)$ and only responds at τ_i, i.e., $j_i(\tau) = 0$ for all $\tau \in \Delta(t) \setminus \{\tau_i\}$ and $j_i(\tau_i) \in [n]$.

(iii) $\ell \notin R^s(\tau) \cup T^s(\tau)$: In this case, ℓ is not requesting nor responding to any requests at time τ. Here, we simply let

$$\mathbf{s}_\ell(\tau+1) = \mathbf{s}_\ell(\tau) + \mathbf{g}_\ell(\tau+1) - \mathbf{g}_\ell(\tau).$$

From the above update rule, it is clear that for any τ, $\mathbf{s}(\tau+1)$ can be written as of a linear form of (5.4) for some row stochastic matrix $W^s(\tau)$. Next we argue that this matrix is also column stochastic. To show this, fix τ. Then for any node $i \in [n]$, consider the sequence of nodes $i_1, i_2, \ldots, i_\kappa$ given by $i_{\ell+1} \triangleq j_{i_\ell}(\tau)$, where κ is the first time where either $j_{i_\kappa}(\tau) = 0$ or $j_{i_\kappa}(\tau) = i_1$. Note that $j_{i_\kappa}(\tau) \neq i_\ell$ for $1 < \ell < \kappa$ as otherwise, i_ℓ has responded to two requests (from $i_{\ell-1}$ and i_κ) which is not possible in our algorithm. Then, depending on whether $j_{i_\kappa} = 0$ or $j_{i_\kappa}(\tau) = i_1$, the induced block matrix by the nodes (indices) $i_1, \ldots, i_\kappa$ is of the form

$$\begin{bmatrix} \frac{2}{3} & \frac{1}{3} & 0 & \cdots & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \cdots & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{2}{3} \end{bmatrix} \text{ or } \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \cdots & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \cdots & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{3} & 0 & 0 & \cdots & \frac{1}{3} \end{bmatrix}, \tag{5.6}$$

respectively. Therefore, the matrix $W_s(\tau)$ consists of induced block matrices of the form of (5.6), and hence, it is doubly stochastic.

Since $W^x(t)$ is row-stochastic, and $W^s(t)$ is doubly stochastic, to check that this scheme satisfies Theorem 7 conditions, we need to investigate whether they satisfy Assumption 3. First note that although the process is potentially dependent, the weight matrices from current window are independent of the ones from the past windows. Let $\gamma \triangleq \frac{1-\bar{p}}{3n^2}$. If at time $\tau \in \Delta(t)$, node j

sends $\mathbf{x}_j(\tau)$ to one of its neighbors, for all $i \in \mathcal{N}_j$, we have

$$\mathbb{E}[w_{ij}^x(\tau)|\mathcal{F}(tB)] = \mathbb{E}[w_{ij}^x(\tau)] = \frac{1-p_{ij}(\tau)}{|\mathcal{N}_j||\mathcal{N}_i^x(\tau)|} \geq \frac{1-\bar{p}}{n^2}.$$

Therefore, $\mathcal{N}_j \in \mathcal{E}^\gamma(\mathbb{E}[W^x(\tau)|\mathcal{F}(tB)])$, and hence

$$\bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}^\gamma(\mathbb{E}[W^x(\tau)|\mathcal{F}(tB)]) = \bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}^\gamma(\mathbb{E}[W^x(\tau)]) = E.$$

Also, if at time $\tau \in \Delta(t)$, node j is submitting a request by sending $\mathbf{s}_j(\tau)$ to one of its neighbors, for all $i \in \mathcal{N}_j$, we have

$$\mathbb{E}[w_{ij}^s(\tau) \mid \mathcal{F}(tB)] = \mathbb{E}[w_{ij}^s(\tau)] = \frac{1-p_{ij}(\tau)}{3|\mathcal{N}_j||\mathcal{N}_i^s(\tau)|} \geq \frac{1-\bar{p}}{3n^2},$$

and hence, similarly, $\bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}^\gamma(\mathbb{E}[W^s(\tau)|\mathcal{F}(tB)]) = E$. Therefore, since $\mathcal{G} = ([n], E)$ is connected, the conditions of Theorem 7 are satisfied, and hence, $\mathbf{x}(t)$ converges to the minimizer of $F(\cdot)$ with a geometric rate.

5.3.2 Proof of Theorem 7

In this section, we study the main dynamics (5.3)-(5.4) and prove Theorem 7. Throughout this section, we let $\bar{\mathbf{x}} \triangleq \frac{1}{n}e^T\mathbf{x}$ and $\bar{\mathbf{s}} \triangleq \frac{1}{n}e^T\mathbf{s}$ for vectors $\mathbf{x}, \mathbf{s} \in (\mathbb{R}^m)^n$,

To prove Theorem 7, we follow the method that is applied in Theorem 1 in [49]. However, this method needs to be extended as in our case, the matrices are time-varying and in the time-varying setting, after each iteration, $\mathrm{d}(\mathbf{x}(t))$ does not necessarily decrease. In other words, $\mathrm{diam}(\Phi(t+1,t))$ is not necessarily strictly less than 1. However, we show that with our connectivity assumptions, $\mathrm{diam}(\Phi(t,\tau)) < 1$ for large enough $t-\tau$, which adds complexity to the proofs. To make arguments easier, we use the following trick: we put the contraction factor in the last corresponding matrix, which is possible due to the following lemma.

Lemma 28. *For the dynamics* (5.3)-(5.4)*, let*

$$y(t) = \begin{bmatrix} \sqrt{n}\mathrm{d}(\mathbf{s}(t)) \\ \sqrt{n}\mathrm{d}(\mathbf{x}(t)) \\ \|\bar{\mathbf{x}}(t) - z^*\| \end{bmatrix},$$

and $\Phi_x(t,\tau), \Phi_s(t,\tau)$ *be the transition matrices associated with* $\{W^x(t)\}$ *and* $\{W^s(t)\}$*, respectively. Then, under Assumption 5, for any* $t \geq \tau$*, we have*

$$y(t) \leq A_\eta(t,\tau) C_\eta^{t-\tau-1} y(\tau). \tag{5.7}$$

where the matrix $A_\eta(t,\tau)$ *is given by*

$$\begin{bmatrix} \mathrm{diam}(\Phi_s(t,\tau)) + 2\beta\eta\sqrt{n} & 2\beta(1+\eta\beta)\sqrt{n} & 2\eta\beta^2\sqrt{n} \\ \eta & \mathrm{diam}(\Phi_x(t,\tau)) & 0 \\ 0 & \eta\beta+1 & \lambda_\eta \end{bmatrix}, \tag{5.8}$$

C_η *is given by*

$$C_\eta \triangleq \begin{bmatrix} 1+2\beta\eta\sqrt{n} & 2\beta(1+\eta\beta)\sqrt{n} & 2\eta\beta^2\sqrt{n} \\ \eta & 1 & 0 \\ 0 & \eta\beta+1 & \lambda_\eta \end{bmatrix} \tag{5.9}$$

and $\lambda_\eta = \max\{|1-\eta\alpha|, |1-\eta\beta|\}$.

Proof: For notational convenience, we remove the subscript η from matrices $A_\eta(t,\tau)$ and

C_η in the proof. Consider the dynamics

$$\begin{cases} u(k+1) = Cu(k) & \text{if } t-1 > k \geq \tau \\ u(k+1) = A(t,\tau)u(k) & \text{if } t-1 = k \end{cases}, \tag{5.10}$$

where $u(\tau) \triangleq y(\tau)$. To prove the lemma, we show that $y(k) \leq u(k)$ for $t \geq k \geq \tau$ by induction on k. Note that the assertion holds for $k = \tau$ as $u(\tau) = y(\tau)$. We divide the proof into two parts: $i)$ $t > k \geq \tau$, and $ii)$ $k = t$.

$i)$ $\boldsymbol{t > k \geq \tau}$: The proof of this case follows the main proof idea of Theorem 1 in [49]. Here, we need to show

$$y(k+1) \leq Cy(k) \tag{5.11}$$

that for all k in this range. In (5.11), we have three inequalities corresponding to each coordinate of $y(k)$. We show each inequality separately.

To establish the inequality for $y_3(k+1)$, multiply $\frac{1}{n}e^T$ from left to both sides of (5.4). Since $W_s(k)$ is doubly stochastic a.s., from Lemma 7-(a) in [49], we have

$$\begin{aligned} \bar{\mathbf{s}}(k+1) &= \bar{\mathbf{s}}(k) + \bar{\mathbf{g}}(k+1) - \bar{\mathbf{g}}(k) \\ &= \bar{\mathbf{g}}(k+1). \end{aligned} \tag{5.12}$$

Let $\overline{W^x}(k) \triangleq \frac{1}{n}e^T W^x(k)$. Multiplying $\frac{1}{n}e^T$ from left to both sides of (5.3), we have

$$
\begin{aligned}
\bar{\mathbf{x}}(k+1) &= \overline{W^x}(k+1)\mathbf{x}(k) - \eta\bar{\mathbf{s}}(k) \\
&\overset{(a)}{=} \overline{W^x}(k+1)\mathbf{x}(k) - \eta\bar{\mathbf{g}}(k), \\
&= \bar{\mathbf{x}}(k) - \eta\bar{\mathbf{g}}(k) + \overline{W^x}(k+1)\mathbf{x}(k) - \bar{\mathbf{x}}(k) \\
&= (\bar{\mathbf{x}}(k) - \eta\nabla F(\bar{\mathbf{x}}(k))) + (\eta\nabla F(\bar{\mathbf{x}}(k)) - \eta\bar{\mathbf{g}}(k)) + \left(\overline{W^x}(k+1)\mathbf{x}(k) - \bar{\mathbf{x}}(k)\right),
\end{aligned}
$$

where (a) follows from (5.12). Therefore, the triangle inequality implies

$$
\begin{aligned}
\|\bar{\mathbf{x}}(k+1) - z^*\| \le\ & \|\bar{\mathbf{x}}(k) - \eta\nabla F(\bar{\mathbf{x}}(k)) - z^*\| + \eta\|\nabla F(\bar{\mathbf{x}}(k)) - \bar{\mathbf{g}}(k)\| \\
& + \|\overline{W^x}(k+1)\mathbf{x}(k) - \bar{\mathbf{x}}(k)\| \\
\le\ & \lambda_\eta\|\bar{\mathbf{x}}(k) - z^*\| + \eta\|\nabla F(\bar{\mathbf{x}}(k)) - \bar{\mathbf{g}}(k)\| + \|\overline{W^x}(k+1)\mathbf{x}(k) - \bar{\mathbf{x}}(k)\|,
\end{aligned}
\tag{5.13}
$$

where the last inequality follows from Lemma 10 in [49]. According to the definition of $F(\cdot)$ and $\mathbf{g}(\cdot)$, we have

$$
\begin{aligned}
\|\nabla F(\bar{\mathbf{x}}(k)) - \bar{\mathbf{g}}(k)\| &= \left\| \sum_{i=1}^{n} \frac{\nabla f_i(\bar{\mathbf{x}}(k)) - \nabla f_i(\mathbf{x}_i(k))}{n} \right\| \\
&\overset{(a)}{\le} \beta \sum_{i=1}^{n} \frac{\|\mathbf{x}_i(k) - \bar{\mathbf{x}}(k)\|}{n} \\
&\overset{(b)}{\le} \beta\sqrt{n}\mathrm{d}(\mathbf{x}(k)),
\end{aligned}
\tag{5.14}
$$

where (a) follows form β-smoothness of f_is and the triangle inequality, and (b) follows from

Lemma 16-(e). Due to the convexity of norm $\|\cdot\|$, we have

$$\begin{aligned}\|\overline{W^x}(k+1)\mathbf{x}(k)-\bar{\mathbf{x}}(k)\| &\le \sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n} w_{ij}^x(k+1)\|\mathbf{x}_i(k)-\bar{\mathbf{x}}(k)\| \\ &\overset{(a)}{\le} \max_{i\in[n]}\|\mathbf{x}_i(k)-\bar{\mathbf{x}}(k)\| \\ &\le \sqrt{n}\mathrm{d}(\mathbf{x}(k)),\end{aligned} \tag{5.15}$$

where (a) follows from the fact that $\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n} w_{ij}^x(k+1)=1$, which is because of row-stochasticity of $W^x(k+1)$. Therefore, from (5.13)-(5.15), we get

$$\|\bar{\mathbf{x}}(k+1)-z^*\| \le \lambda_\eta\|\bar{\mathbf{x}}(k)-z^*\|+(\eta\beta+1)\sqrt{n}\mathrm{d}(\mathbf{x}(k)),$$

which proves (5.11) for $y_3(k)$.

To show the inequality for $y_2(k+1)$, by applying $\mathrm{d}(\cdot)$ on (5.3), and using Lemma 16-(a) and 16-(b), we get

$$\begin{aligned}\mathrm{d}(\mathbf{x}(k+1)) &\le \mathrm{diam}(W^x(k+1))\mathrm{d}(\mathbf{x}(k))+\eta\mathrm{d}(\mathbf{s}(k)) \\ &\le \mathrm{d}(\mathbf{x}(k))+\eta\mathrm{d}(\mathbf{s}(k)),\end{aligned}$$

where the latter inequality follows from $\mathrm{diam}(W^x(k+1))\le 1$. This simply shows that the inequality in (5.11) holds for $y_2(k+1)$.

To show the inequality for the first coordinate, we apply $\mathrm{d}(\cdot)$ on the dynamics (5.4). Then, Lemma 16-(a), Lemma 16-(b), and $\mathrm{diam}(W^s(k+1))\le 1$ imply

$$\mathrm{d}(\mathbf{s}(k+1)) \le \mathrm{d}(\mathbf{s}(k))+\mathrm{d}(\mathbf{g}(k+1)-\mathbf{g}(k)). \tag{5.16}$$

From the definition of $\mathrm{d}(\cdot)$, we have

$$\begin{aligned}
\mathrm{d}(\mathbf{g}(k+1)-\mathbf{g}(k)) &\leq 2\max_{i\in[n]}\|\mathbf{g}_i(k+1)-\mathbf{g}_i(k)\| \\
&\overset{(a)}{\leq} 2\beta\max_{i\in[n]}\|\mathbf{x}_i(k+1)-\mathbf{x}_i(k)\| \\
&\overset{(b)}{=} 2\beta\max_{i\in[n]}\left\|\sum_{j=1}^{n} w_{ij}^x(k+1)\mathbf{x}_j(k)-\eta\mathbf{s}_i(k)-\mathbf{x}_i(k)\right\| \\
&\overset{(c)}{\leq} 2\beta\max_{i\in[n]}\sum_{j=1}^{n} w_{ij}^x(k+1)\,\|\mathbf{x}_j(k)-\eta\mathbf{s}_i(k)-\mathbf{x}_i(k)\| \\
&\overset{(d)}{\leq} 2\beta\max_{i\in[n]}\sum_{j=1}^{n} w_{ij}^x(k+1)\,(\|\mathbf{x}_j(k)-\mathbf{x}_i(k)\|+\eta\|\mathbf{s}_i(k)\|) \\
&\overset{(e)}{\leq} 2\beta\max_{i\in[n]}\sum_{j=1}^{n} w_{ij}^x(k+1)\left(\sqrt{n}\mathrm{d}(\mathbf{x}(k))+\eta\|\mathbf{s}_i(k)\|\right) \\
&\overset{(f)}{\leq} 2\beta\left(\sqrt{n}\mathrm{d}(\mathbf{x}(k))+\eta\max_{i\in[n]}\|\mathbf{s}_i(k)\|\right),
\end{aligned} \tag{5.17}$$

where (a) follows from the β-smoothness of f_i, (b) is derived from (5.3), (c) is due to the convexity of norm, (d) follows from Triangle inequality, (e) is due to $\|v\| \leq \sqrt{n}\|v\|_\infty$ for some vector v, and finally, (f) follows from the fact that $W^x(k+1)$ is a row-stochastic matrix. Moreover, we have

$$\begin{aligned}
\|\mathbf{s}_i(k)\| &= \|\mathbf{s}_i(k)-\bar{\mathbf{g}}(k)+\bar{\mathbf{g}}(k)-\nabla F(\bar{\mathbf{x}}(k))+\nabla F(\bar{\mathbf{x}}(k))\| \\
&\leq \|\mathbf{s}_i(k)-\bar{\mathbf{g}}(k)\|+\|\bar{\mathbf{g}}(k)-\nabla F(\bar{\mathbf{x}}(k))\|+\|\nabla F(\bar{\mathbf{x}}(k))\| \\
&\overset{(a)}{\leq} \sqrt{n}\mathrm{d}(\mathbf{s}(k))+\beta\sqrt{n}\mathrm{d}(\mathbf{x}(k))+\|\nabla F(\bar{\mathbf{x}}(k))\| \\
&\overset{(b)}{\leq} \sqrt{n}\mathrm{d}(\mathbf{s}(k))+\beta\sqrt{n}\mathrm{d}(\mathbf{x}(k))+\beta\|\bar{\mathbf{x}}(k)-z^*\|,
\end{aligned} \tag{5.18}$$

where (a) follows from $\bar{\mathbf{s}}(k)=\bar{\mathbf{g}}(k)$ (from (5.12)), (5.14), and Lemma 16-(e), and (b) follows from the fact that z^* is the minimizer of $F(\cdot)$, and hence $\nabla F(z^*) = 0$, which implies

$$\|\nabla F(\bar{\mathbf{x}}(k))\| = \|\nabla F(\bar{\mathbf{x}}(k)) - \nabla F(z^*)\| \leq \beta \|\bar{\mathbf{x}}(k) - z^*\|.$$

Therefore, (5.17) and (5.18) imply

$$\mathrm{d}(\mathbf{g}(k+1) - \mathbf{g}(k)) \leq 2\beta(1+\eta\beta)\sqrt{n}\mathrm{d}(\mathbf{x}(k)) + 2\eta\beta\sqrt{n}\mathrm{d}(\mathbf{s}(k)) + 2\eta\beta^2\|\bar{\mathbf{x}}(k) - z^*\|. \quad (5.19)$$

Combining (5.16) and (5.19), we have

$$\mathrm{d}(\mathbf{s}(k+1)) \leq (1+2\beta\eta\sqrt{n})\mathrm{d}(\mathbf{s}(k)) + 2\beta\,(1+\eta\beta)\,\sqrt{n}\mathrm{d}(\mathbf{x}(k)) + 2\eta\beta^2\|\bar{\mathbf{x}}(k) - z^*\|,$$

which proves the inequality (5.11) associated with the first row of C, and completes the proof that (5.11) holds for $t > k > \tau$.

$ii)$ $\boldsymbol{k = t}$: Again, we show each inequality corresponding to each coordinate of $y(k)$ in (5.11), separately. Since the third row of $A(t,\tau)$ and C are equal, the proof for the inequality corresponding to $y_3(t)$ is similar to the case $t > k \geq \tau$.

To establish the inequality for $y_2(t)$, note that the dynamics (5.3) is a linear system with input $\eta\mathbf{s}(t)$ which implies

$$\mathbf{x}(t) = \Phi_x(t,\tau)\mathbf{x}(\tau) - \sum_{r=\tau}^{t-1} \Phi_x(t, r+1)\eta\mathbf{s}(r).$$

Therefore, Lemma 16-(a) and 16-(b) imply

$$\begin{aligned}
\mathrm{d}(\mathbf{x}(t)) &\leq \mathrm{diam}(\Phi_x(t,\tau))\mathrm{d}(\mathbf{x}(\tau)) + \sum_{r=\tau}^{t-1} \mathrm{diam}(\Phi_x(t,r+1))\eta \mathrm{d}(\mathbf{s}(r)) \\
&\leq \mathrm{diam}(\Phi_x(t,\tau))\mathrm{d}(\mathbf{x}(\tau)) + \sum_{r=\tau}^{t-1} \eta \mathrm{d}(\mathbf{s}(r)) \\
&\overset{(a)}{\leq} \mathrm{diam}(\Phi_x(t,\tau))\frac{u_2(\tau)}{\sqrt{n}} + \sum_{r=\tau}^{t-1} \eta \frac{u_1(r)}{\sqrt{n}} \\
&\overset{(b)}{=} \frac{u_2(t)}{\sqrt{n}},
\end{aligned}$$

where (a) follows from the induction hypothesis, i.e., $\mathrm{d}(\mathbf{s}(r)) \leq u_1(r)$ for $t-1 > r \geq \tau$ and $u_2(\tau) = \mathrm{d}(\mathbf{x}(\tau))$, and (b) follows from the fact that based on (5.10), we have

$$u_2(r+1) = a_x(r+1)u_2(r) + \eta u_1(r)$$

for $t > r \geq \tau$, where

$$a_x(r) = \begin{cases} 1, & \text{if } t > r \geq \tau \\ \mathrm{diam}(\Phi_x(t,\tau)), & \text{if } r = t \end{cases}.$$

To show the inequality for $y_1(t)$, noting that the dynamics (5.4) is a linear system with input $\mathbf{g}(t+1) - \mathbf{g}(t)$, we have

$$\mathbf{s}(t) = \Phi_s(t,\tau)\mathbf{s}(\tau) + \sum_{r=\tau}^{t-1} \Phi_s(t,r+1)(\mathbf{g}(r+1) - \mathbf{g}(r)).$$

Therefore, Lemma 16-(a) and 16-(b), and $\operatorname{diam}(\Phi_s(t,\tau)) \leq 1$ imply

$$\begin{aligned}
\mathrm{d}(\mathbf{s}(t)) &\leq \operatorname{diam}(\Phi_s(t,\tau))\mathrm{d}(\mathbf{s}(\tau)) + \sum_{r=\tau}^{t-1} \mathrm{d}(\mathbf{g}(r+1)-\mathbf{g}(r)) \\
&\overset{(a)}{\leq} \operatorname{diam}(\Phi_s(t,\tau))\mathrm{d}(\mathbf{s}(\tau)) \\
&\quad + \sum_{r=\tau}^{t-1} 2\beta(1+\eta\beta)\sqrt{n}\mathrm{d}(\mathbf{x}(r)) + 2\eta\beta\sqrt{n}\mathrm{d}(\mathbf{s}(r)) + 2\eta\beta^2\|\bar{\mathbf{x}}(r)-z^*\| \\
&\overset{(b)}{\leq} \operatorname{diam}(\Phi_s(t,\tau))\frac{u_1(\tau)}{\sqrt{n}} + \sum_{r=\tau}^{t-1} 2\beta(1+\eta\beta)u_2(r) + 2\eta\beta u_1(r) + 2\eta\beta^2 u_3(r) \\
&\overset{(c)}{=} \frac{u_1(t)}{\sqrt{n}},
\end{aligned}$$

where (a) follows from (5.19), (b) follows from the induction hypothesis, i.e., $y(r) \leq u(r)$ for $t-1 > r \geq \tau$ and $u_1(\tau) = \sqrt{n}\mathrm{d}(\mathbf{x}(\tau))$, and (c) follows from the fact that based on (5.10), for $t > r \geq \tau$, we have

$$u_1(r+1) = a_s(r+1)u_1(r) + \left[(1+\eta\beta)u_2(r) + 2\eta\beta u_1(r) + 2\eta\beta^2 u_3(r)\right]\sqrt{n},$$

where

$$a_s(r) = \begin{cases} 1, & \text{if } t > r \geq \tau \\ \operatorname{diam}(\Phi_s(t,\tau)), & \text{if } r = t \end{cases}.$$

■

To prove Theorem 7, we need to prove that as $t \to \infty$, $y_3(t)$ converges to zero geometrically (almost surely) where the dynamics $y(t)$ satisfies (5.7). To do so, we study the random process $u(t)$ which satisfies

$$u(t+1) \leq Q(t+1)u(t), \tag{5.20}$$

where $Q(t+1) \triangleq A_\eta((t+1)K, tK)C_\eta^{(K-1)}$ is a random matrix. A classical way to prove that $\lim_{t\to\infty} \frac{\|u(t)\|_\infty}{\rho^t} < \infty$, for some $0 < \rho < 1$, is to consider the Lyapunov function $V(t) = u^T(t)u(t)$. However, since the largest eigenvalue of $\mathbb{E}[Q^T(t)Q(t)]$ is not necessarily less than 1, we cannot apply this Lyapunov function for this random process. Instead in the following lemma, we exploit the fact that $u(t)$ is non-negative and use an alternative Lyapunov function to prove

$$\lim_{t\to\infty} \frac{\|u(t)\|_\infty}{\rho^t} < \infty.$$

Lemma 29. *Consider a non-negative sequence of $N \times N$ random matrices $\{Q(t)\}$ that is adapted to a filtration $\{\mathcal{F}(t)\}$ such that $\mathbb{E}[Q(t+1) \mid \mathcal{F}(t)] \leq M$ for some matrix M and all t. Let ρ and π be the largest eigenvalue of M and the associated left eigenvector of M, respectively, and assume that π is strictly positive. Then, for the non-negative random process $u(t)$ satisfying $u(t+1) \leq Q(t+1)u(t)$, we have $\lim_{t\to\infty} \frac{\|u(t)\|_\infty}{\rho^t} < \infty$ almost surely.*

Proof: Let $\{u(t)\}$ be a non-negative random process $u(t)$ satisfying $u(t+1) \leq Q(t+1)u(t)$ for all $t \geq 0$. Define the Lyapunov function

$$V(t) \triangleq \frac{\pi^T u(t)}{\rho^t}.$$

Since $u(t)$ is a non-negative random process, $V(t)$ is a non-negative random process. Thus, we have

$$\begin{aligned}\mathbb{E}[V(t+1) \mid \mathcal{F}(t)] &= \mathbb{E}\left[\frac{\pi^T u(t+1)}{\rho^{t+1}}\middle|\mathcal{F}(t)\right] \\ &\leq \frac{\mathbb{E}\left[\pi^T Q(t+1)u(t) \mid \mathcal{F}(t)\right]}{\rho^{t+1}} \\ &= \frac{\pi^T \mathbb{E}\left[Q(t+1) \mid \mathcal{F}(t)\right] u(t)}{\rho^{t+1}} \\ &\leq \frac{\pi^T M u(t)}{\rho^{t+1}} \\ &= \frac{\rho \pi^T u(t)}{\rho^{t+1}} \\ &= V(t),\end{aligned}$$

and hence, $V(t)$ is a sub-martingale. Since, $V(t)$ is non-negative, we have $\mathbb{E}[|V(t)|^+] = \mathbb{E}[V(t)] \leq \mathbb{E}[V(0)] < \infty$. Therefore, Doob's Martingale Convergence Theorem (see e.g., Theorem 5.2.8 in [20]) implies that $\lim_{t\to\infty} V(t)$ exists, and it is less than infinity. Moreover, we have

$$\infty > \lim_{t\to\infty} V(t) = \lim_{t\to\infty} \sum_{i=1}^{N} \frac{\pi_i u_i(t)}{\rho^t} \geq \min_{i\in N} \pi_i \lim_{t\to\infty} \frac{\|u(t)\|_\infty}{\rho^t}.$$

Finally, $\min_{i\in N} \pi_i > 0$, implies $\lim_{t\to\infty} \frac{\|u(t)\|_\infty}{\rho^t} < \infty$. ■

Regarding (5.20) and Lemma 29, to prove $\lim_{t\to\infty} \frac{\|u(t)\|_\infty}{\rho^t} < \infty$, we need to find the largest left eigenvalue and corresponding eigenvector of $\mathbb{E}[A_\eta((t+1)K, tK) C_\eta^{(K-1)} \mid \mathcal{F}(tB)]$, which is done in the following lemma.

Lemma 30. *Consider the matrices C_η (defined in* (5.9)*) and*

$$H_\eta(\theta) \triangleq \begin{bmatrix} \theta+2\beta\eta\sqrt{n} & 2\beta(1+\eta\beta)\sqrt{n} & 2\eta\beta^2\sqrt{n} \\ \eta & \theta & 0 \\ 0 & \eta\beta+1 & \lambda_\eta \end{bmatrix}. \tag{5.21}$$

Then, for any $\theta < 1$ and $k \geq 0$, there exist an $\tilde{\eta}(\theta,k) > 0$ such that if $\eta < \tilde{\eta}(\theta,k)$, the largest left eigenvalue of $H_\eta(\theta)C_\eta^k$ is real and strictly less than 1. Also, the left eigenvector associated to the largest eigenvalue is strictly positive for $\eta > 0$.

Proof: First, note that for $\eta > 0$, $H_\eta(\theta)$ is an irreducible matrix. Therefore, since $C_\eta \geq \lambda_\eta I$, for $k \geq 0$, $H_\eta(\theta)C_\eta^k$ is an irreducible matrix. Since $H_\eta(\theta)C_\eta^k$ is a non-negative matrix, the Perron–Frobenius Theorem (see e.g., Chapter 8 in [38]) implies that the largest eigenvalue of $H_\eta(\theta)C_\eta^k$, denoted by $\rho(\theta)$, is a positive real number, and there is a strictly positive left eigenvector (Perron eigenvector) associated with that. By induction, we can show that

$$C_0^k \triangleq \begin{bmatrix} 1 & 2k\beta\sqrt{n} & 0 \\ 0 & 1 & 0 \\ 0 & k & 1 \end{bmatrix}.$$

Therefore, we have

$$H_0(\theta)C_0^k \triangleq \begin{bmatrix} \theta & (2k\theta+2)\beta\sqrt{n} & 0 \\ 0 & \theta & 0 \\ 0 & 1+k\lambda_\eta & \lambda_\eta \end{bmatrix}.$$

Since, the characteristic polynomial of $H_0(\theta)C_0^k$ is $c(\xi) = (\theta-\xi)^2(\lambda_\eta-\xi)$, the eigenvalues of $H_0(\theta)C_0^k$ are θ and λ_η. Therefore, since for $0 < \eta < \max\left\{\frac{2}{\alpha}, \frac{2}{\beta}\right\}$, λ_η is strictly less than one, we have $\rho(0) < 1$ for $0 < \eta < \max\left\{\frac{2}{\alpha}, \frac{2}{\beta}\right\}$. Since, the elements of $H_\eta(\theta)C_\eta^k$ are continuous functions of η, and the eigenvalues of $H_\eta(\theta)C_\eta^k$ are continuous functions of its elements, $\rho(\eta)$ is a continuous function of η. Therefore, there exists an $0 < \tilde{\eta}(\theta,k) < \max\left\{\frac{2}{\alpha},\frac{2}{\beta}\right\}$ such that if $0 < \eta < \tilde{\eta}(\theta,k)$, we have $\rho(\eta) < 1$. ∎

Finally, we can present the proof of Theorem 7.

Proof of Theorem 7: Let $K \triangleq n^2B$ where B is given in Assumption 3. Considering Lemma 28, we have

$$y((t+1)K) \leq A_\eta((t+1)K,tK)C_\eta^{(K-1)}y(tK),$$

From Lemma 17 and Lemma 16-(c), we have

$$\mathbb{E}[A_\eta((t+1)K,tK)C_\eta^{(K-1)} \mid \mathcal{F}(tK)] \leq H_\eta(1-\theta)C_\eta^{(K-1)},$$

for some $\theta > 0$. Therefore, Lemma 29 and 30 imply that for all $0 < \eta < \tilde{\eta}(1-\theta,K-1)$, we have $\lim_{t\to\infty}\frac{\|y(tK)\|_\infty}{\rho^t(\eta)} < \infty$ for some $\rho(\eta) < 1$. Since, K is finite, and for $tK < k < (t+1)K$,

$$y(k) \leq A_\eta(k,tB)C_\eta^{k-tK-1}y(tK),$$

we have $\lim_{k\to\infty}\frac{\|y(k)\|_\infty}{\rho^k(\eta)} < \infty$. Finally, noting $y_3(k) = \|\mathbf{x}(k) - z^*\|$ completes the proof. ∎

Chapter 5, in full, is a reprint of the material as it appears in A. Aghajan and B. Touri, Geometric convergence for distributed optimization over dependent random networks, being prepared for publication. The dissertation author was the primary investigator and author of this paper.

Chapter 6

Conclusion and Future Research Direction

This thesis was on distributed averaging dynamics and its main application in distributed optimization.

In the first part of the thesis, first we considered the discrete-time distributed averaging dynamics. We showed that while having spanning rooted tree in the infinite graph is not sufficient in general, it is necessary for ergodicity of products of inhomogeneous stochastic matrices. In addition, we showed that if we consider the time-varying leader-follower dynamics among groups of agents that the averaging dynamics restricted to each group is $\mathcal{P}^*$, and the groups are connected using a directed acyclic graph containing a spanning directed rooted tree (in the directed infinite flow graph of the original process), then all agents' values will converge to the consensus value of the leading group. Then, we considered the continuous-time counterpart of the time-varying leader-follower dynamics and showed that a similar sufficient condition for consensus of time-varying continuous-time distributed averaging dynamics holds.

On this topic, closing the gap between the necessary and sufficient conditions for ergodicity of both discrete-time and continuous-time distributed averaging dynamics is of interest for future research direction.

In the second part of the thesis, we considered distributed optimization over dependent random networks. First, we considered convex objective functions and showed that the averaging-based distributed optimization solving algorithm over dependent random networks converges to an optimal random point if the underlying network structure is conditionally B-connected. To do so, we established a rate of convergence estimate for the second moment of the autonomous averaging dynamics over such networks and used that to study the convergence of the sample-paths and second moments of the controlled variation of those dynamics. Then, we studied the distributed gradient-tracking algorithm for a faster convergence rate for problems with smooth and strongly convex objective functions. We proved that, under the conditional B-connectivity condition, the distributed gradient-tracking algorithm converges geometrically to the optimal point for strongly convex and smooth function.

A future research direction for this part is further extensions of the current work to non-convex settings, accelerated algorithms, and distributed online learning algorithms.

Bibliography

[1] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

[2] Claudio Altafini. "Consensus problems on networks with antagonistic interactions". In: *IEEE transactions on automatic control* 58.4 (2012), pp. 935–946.

[3] Tuncer Can Aysal, Mehmet Ercan Yildiz, Anand D Sarwate, and Anna Scaglione. "Broadcast gossip algorithms for consensus". In: *IEEE Transactions on Signal processing* 57.7 (2009), pp. 2748–2761.

[4] Pascal Bianchi and Jérémie Jakubowicz. "Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization". In: *IEEE transactions on automatic control* 58.2 (2012), pp. 391–405.

[5] Vincent D Blondel, Julien M Hendrickx, and John N Tsitsiklis. "On Krause's multi-agent consensus model with state-dependent connectivity". In: *IEEE transactions on Automatic Control* 54.11 (2009), pp. 2586–2597.

[6] Sadegh Bolouki and Roland P Malhamé. "Consensus algorithms and the decomposition-separation theorem". In: *IEEE Transactions on Automatic Control* 61.9 (2016), pp. 2357–2369.

[7] Sadegh Bolouki and Roland P Malhamé. "On consensus with a general discrete time convex combination based algorithm for multi-agent systems". In: *2011 19th Mediterranean Conference on Control & Automation (MED)*. IEEE. 2011, pp. 668–673.

[8] Sadegh Bolouki, Roland P Malhamé, Milad Siami, and Nader Motee. "Éminence grise coalitions: On the shaping of public opinion". In: *IEEE Transactions on Control of Network Systems* 4.2 (2015), pp. 133–145.

[9] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. "Randomized gossip algorithms". In: *IEEE transactions on information theory* 52.6 (2006), pp. 2508–2530.

[10] Francesco Bullo, Jorge Cortes, and Sonia Martinez. *Distributed control of robotic networks: a mathematical approach to motion coordination algorithms*. Vol. 27. Princeton University Press, 2009.

[11] Ming Cao, A Stephen Morse, and Brian DO Anderson. "Reaching a consensus in a dynamically changing environment: A graphical approach". In: *SIAM Journal on Control and Optimization* 47.2 (2008), pp. 575–600.

[12] Ming Cao, A Stephen Morse, and Brian DO Anderson. "Reaching a consensus in a dynamically changing environment: convergence rates, measurement delays, and asynchronous events". In: *SIAM Journal on Control and Optimization* 47.2 (2008), pp. 601–623.

[13] Federico S Cattivelli and Ali H Sayed. "Diffusion LMS strategies for distributed estimation". In: *IEEE Transactions on Signal Processing* 58.3 (2010), pp. 1035–1048.

[14] Bernadette Charron-Bost. "Orientation and connectivity based criteria for asymptotic consensus". In: *arXiv preprint arXiv:1303.2043* (2013).

[15] Samprit Chatterjee and Eugene Seneta. "Towards consensus: Some convergence theorems on repeated averaging". In: *Journal of Applied Probability* 14.1 (1977), pp. 89–97.

[16] Ashish Cherukuri and Jorge Cortés. "Distributed generator coordination for initialization and anytime optimization in economic dispatch". In: *IEEE Transactions on Control of Network Systems* 2.3 (2015), pp. 226–237.

[17] Morris H DeGroot. "Reaching a consensus". In: *Journal of the American Statistical Association* 69.345 (1974), pp. 118–121.

[18] Alejandro D Dominguez-Garcia, Stanton T Cady, and Christoforos N Hadjicostis. "Decentralized optimal dispatch of distributed energy resources". In: *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE. 2012, pp. 3688–3693.

[19] Alejandro D Dominguez-Garcia and Christoforos N Hadjicostis. "Distributed algorithms for control of demand response and distributed energy resources". In: *2011 50th IEEE Conference on Decision and Control and European Control Conference*. IEEE. 2011, pp. 27–32.

[20] Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2010.

[21] Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.

[22] Bahman Gharesifard and Jorge Cortés. "Distributed continuous-time convex optimization on weight-balanced digraphs". In: *IEEE Transactions on Automatic Control* 59.3 (2013), pp. 781–786.

[23] John Hajnal and MS Bartlett. "Weak ergodicity in non-homogeneous Markov chains". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 54.2 (1958), pp. 233–246.

[24] Rainer Hegselmann, Ulrich Krause, et al. "Opinion dynamics and bounded confidence models, analysis, and simulation". In: *Journal of Artificial Societies and Social simulation* 5.3 (2002).

[25] Julien M Hendrickx and John N Tsitsiklis. "Convergence of type-symmetric and cut-balanced consensus seeking systems". In: *IEEE Transactions on Automatic Control* 58.1 (2013), pp. 214–218.

[26] Julien M Hendrickx and John N Tsitsiklis. "Convergence of type-symmetric and cut-balanced consensus seeking systems (extended version)". In: *arXiv preprint arXiv:1102.2361* (2011).

[27] A. Jadbabaie, and A. S. Morse. “Coordination of groups of mobile autonomous agents using nearest neighbor rules”. In: *IEEE Transactions on Automatic Control* 48.6 (2003), pp. 988–1001.

[28] Björn Johansson. “On distributed optimization in networked systems”. PhD thesis. KTH, 2008.

[29] Soummya Kar and José MF Moura. “Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs”. In: *IEEE Journal of Selected Topics in Signal Processing* 5.4 (2011), pp. 674–690.

[30] Solmaz S Kia, Jorge Cortés, and Sonia Martínez. “Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication”. In: *Automatica* 55 (2015), pp. 254–264.

[31] S. Lee and A. Nedić. “Asynchronous Gossip-Based Random Projection Algorithms Over Networks”. In: *IEEE Transactions on Automatic Control* 61.4 (2016), pp. 953–968.

[32] Q. Liu and J. Wang. “A Second-Order Multi-Agent Network for Bound-Constrained Distributed Optimization”. In: *IEEE Transactions on Automatic Control* 60.12 (2015), pp. 3310–3315. ISSN: 0018-9286. DOI: 10.1109/TAC.2015.2416927.

[33] Ilan Lobel and Asuman Ozdaglar. “Distributed subgradient methods for convex optimization over random networks”. In: *IEEE Transactions on Automatic Control* 56.6 (2010), pp. 1291–1306.

[34] Ilan Lobel, Asuman Ozdaglar, and Diego Feijer. “Distributed multi-agent optimization with state-dependent communication”. In: *Mathematical programming* 129.2 (2011), pp. 255–284.

[35] Jan Lorenz. “A stabilization theorem for dynamics of continuous opinions”. In: *Physica A: Statistical Mechanics and its Applications* 355.1 (2005), pp. 217–223.

[36] Van Sy Mai and Eyad H Abed. “Distributed optimization over weighted directed graphs using row stochastic matrix”. In: *2016 American Control Conference (ACC)*. IEEE. 2016, pp. 7165–7170.

[37] Samuel Martin and Julien M Hendrickx. “Continuous-time consensus under noninstantaneous reciprocity”. In: *IEEE Transactions on Automatic Control* 61.9 (2016), pp. 2484–2495.

[38] Carl D Meyer. *Matrix analysis and applied linear algebra*. Vol. 71. Siam, 2000.

[39] Soheil Mohajer and Behrouz Touri. “On convergence rate of scalar Hegselmann-Krause dynamics”. In: *2013 American Control Conference*. IEEE. 2013, pp. 206–210.

[40] Luc Moreau. “Stability of continuous-time distributed consensus algorithms”. In: *2004 43rd IEEE conference on Decision and Control (CDC)*. Vol. 4. IEEE. 2004, pp. 3998–4003.

[41] G. Morral, P. Bianchi, and G. Fort. “Success and Failure of Adaptation-Diffusion Algorithms With Decaying Step Size in Multiagent Networks”. In: *IEEE Transactions on Signal Processing* 65.11 (2017), pp. 2798–2813.

[42] A. Nedic. “Asynchronous Broadcast-Based Convex Optimization Over a Network”. In: *IEEE Transactions on Automatic Control* 56.6 (2011), pp. 1337–1351.

[43] Angelia Nedić and Alex Olshevsky. “Distributed optimization over time-varying directed graphs”. In: *IEEE Transactions on Automatic Control* 60.3 (2014), pp. 601–615.

[44] Angelia Nedic, Alex Olshevsky, and Wei Shi. “Achieving geometric convergence for distributed optimization over time-varying graphs”. In: *SIAM Journal on Optimization* 27.4 (2017), pp. 2597–2633.

[45] Angelia Nedić and Asuman Ozdaglar. “Distributed subgradient methods for multi-agent optimization”. In: *IEEE Transactions on Automatic Control* 54.1 (2009), p. 48.

[46] Angelia Nedić and Asuman Ozdaglar. “On the rate of convergence of distributed subgradient methods for multi-agent optimization”. In: *2007 46th IEEE Conference on Decision and Control*. IEEE. 2007, pp. 4711–4716.

[47] Angelia Nedić, Asuman Ozdaglar, and Pablo A Parrilo. “Constrained consensus and optimization in multi-agent networks”. In: *IEEE Transactions on Automatic Control* 55.4 (2010), pp. 922–938.

[48] Yurii Nesterov. *Lectures on convex optimization*. Vol. 137. Springer, 2018.

[49] Guannan Qu and Na Li. “Harnessing smoothness to accelerate distributed optimization”. In: *IEEE Transactions on Control of Network Systems* 5.3 (2017), pp. 1245–1260.

[50] Michael Rabbat and Robert Nowak. “Distributed optimization in sensor networks”. In: *Proceedings of the 3rd international symposium on Information processing in sensor networks*. ACM. 2004, pp. 20–27.

[51] S Sundhar Ram, Angelia Nedić, and Venugopal V Veeravalli. “Distributed stochastic subgradient projection algorithms for convex optimization”. In: *Journal of optimization theory and applications* 147.3 (2010), pp. 516–545.

[52] Amirhossein Reisizadeh, Aryan Mokhtari, Hamed Hassani, and Ramtin Pedarsani. “An exact quantized decentralized gradient descent algorithm”. In: *IEEE Transactions on Signal Processing* 67.19 (2019), pp. 4934–4947.

[53] Hadi Reisizadeh, Behrouz Touri, and Soheil Mohajer. “Adaptive Bit Allocation for Communication Efficient Distributed Optimization”. In: (2021). To appear in IEEE Conference on Decision and Control 2021.

[54] Hadi Reisizadeh, Behrouz Touri, and Soheil Mohajer. “Distributed Optimization over Time-varying Graphs with Imperfect Sharing of Information”. In: *arXiv preprint arXiv:2106.08469* (2021).

[55] Wei Ren and Randal W Beard. “Consensus seeking in multiagent systems under dynamically changing interaction topologies”. In: *IEEE Transactions on Automatic Control* 50.5 (2005), pp. 655–661.

[56] Wei Ren, Randal W Beard, and Timothy W McLain. “Coordination variables and consensus building in multiple vehicle systems”. In: *Cooperative control*. Springer, 2005, pp. 171–188.

[57] Pouya Rezaeinia and Bahman Gharesifard. “Distributed Optimization with Uncertain Communications”. In: *arXiv preprint arXiv:2003.10342* (2020).

[58] Pouya Rezaeinia, Bahman Gharesifard, Tamas Linder, and Behrouz Touri. “Push-sum on random graphs: almost sure convergence and convergence rate”. In: *IEEE Transactions on Automatic Control* (2019).

[59] H. Robbins and D. Siegmund. “A CONVERGENCE THEOREM FOR NON-NEGATIVE ALMOST SUPERMARTINGALES AND SOME APPLICATIONS”. In: *Optimizing Methods in Statistics*. Ed. by Jagdish S. Rustagi. Academic Press, 1971, pp. 233 –257. ISBN: 978-0-12-604550-5. DOI: https://doi.org/10.1016/B978-0-12-604550-5.50015-8. URL: http://www.sciencedirect.com/science/article/pii/B9780126045505500158.

[60] Fakhteh Saadatniaki, Ran Xin, and Usman A Khan. “Decentralized optimization over time-varying directed graphs with row and column-stochastic matrices”. In: *IEEE Transactions on Automatic Control* 65.11 (2020), pp. 4769–4780.

[61] Jianhong Shen. “A geometric approach to ergodic non-homogeneous Markov chains”. In: *Lecture Notes in Pure and Applied Mathematics* (2000), pp. 341–366.

[62] Eduardo D Sontag. *Mathematical control theory: deterministic finite dimensional systems*. Vol. 6. Springer Science & Business Media, 2013.

[63] Srdjan S Stankovic, Miloš S Stankovic, and Dušan M Stipanovic. “Decentralized parameter estimation by consensus based stochastic approximation”. In: *IEEE Transactions on Automatic Control* 56.3 (2011), pp. 531–543.

[64] Alireza Tahbaz-Salehi and Ali Jadbabaie. “A necessary and sufficient condition for consensus over random networks”. In: *IEEE Transactions on Automatic Control* 53.3 (2008), pp. 791–795.

[65] Behrouz Touri and Cedric Langbort. “On endogenous random consensus and averaging dynamics”. In: *IEEE Transactions on Control of Network Systems* 1.3 (2014), pp. 241–248.

[66] Behrouz Touri, A Nedić, and S Sundhar Ram. “Asynchronous stochastic convex optimization over random networks: Error bounds”. In: *2010 Information Theory and Applications Workshop (ITA)*. IEEE. 2010, pp. 1–10.

[67] Behrouz Touri and Angelia Nedić. “Discrete-time opinion dynamics”. In: *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. IEEE. 2011, pp. 1172–1176.

[68] Behrouz Touri and Angelia Nedić. “On approximations and ergodicity classes in random chains”. In: *IEEE Transactions on Automatic Control* 57.11 (2012), pp. 2718–2730.

[69] Behrouz Touri and Angelia Nedić. “On ergodicity, infinite flow, and consensus in random models”. In: *IEEE Transactions on Automatic Control* 56.7 (2011), pp. 1593–1605.

[70] Behrouz Touri and Angelia Nedić. “Product of random stochastic matrices”. In: *IEEE Transactions on Automatic Control* 59.2 (2013), pp. 437–448.

[71] Behrouz Touri and Angelia Nedić. “Product of random stochastic matrices”. In: *IEEE Transactions on Automatic Control* 59.2 (2014), pp. 437–448.

[72] Behrouz Touri and Angelia Nedić. “When infinite flow is sufficient for ergodicity”. In: *Decision and Control (CDC), 2010 49th IEEE Conference on*. IEEE. 2010, pp. 7479–7486.

[73] Konstantinos Tsianos. “The role of the network in distributed optimization algorithms: Convergence rates, scalability, communication/computation tradeoffs and communication delays”. PhD thesis. McGill University Libraries, 2013.

[74] Konstantinos I Tsianos, Sean Lawlor, and Michael G Rabbat. “Consensus-based distributed optimization: Practical issues and applications in large-scale machine learning”. In: *2012 50th annual allerton conference on communication, control, and computing (allerton)*. IEEE. 2012, pp. 1543–1550.

[75] Konstantinos I Tsianos, Sean Lawlor, and Michael G Rabbat. “Push-sum distributed dual averaging for convex optimization”. In: *2012 ieee 51st ieee conference on decision and control (cdc)*. IEEE. 2012, pp. 5453–5458.

[76] John N Tsitsiklis, Dimitri Bertsekas, and Michael Athans. “Distributed asynchronous deterministic and stochastic gradient optimization algorithms”. In: *IEEE Transactions on Automatic Control* 31.9 (1986), pp. 803–812.

[77] John Nikolas Tsitsiklis. *Problems in decentralized decision making and computation.* Tech. rep. Massachusetts Inst of Tech Cambridge Lab for Information and Decision Systems, 1984.

[78] Marcos M Vasconcelos, Thinh T Doan, and Urbashi Mitra. “Improved Convergence Rate for a Distributed Two-Time-Scale Gradient Method under Random Quantization”. In: *arXiv preprint arXiv:2105.14089* (2021).

[79] Chenguang Xi, Van Sy Mai, Ran Xin, Eyad H Abed, and Usman A Khan. “Linear convergence in optimization over directed graphs with row-stochastic matrices”. In: *IEEE Transactions on Automatic Control* 63.10 (2018), pp. 3558–3565.

[80] Feng Xiao and Long Wang. “State consensus for multi-agent systems with switching topologies and time-varying delays”. In: *International Journal of Control* 79.10 (2006), pp. 1277–1284.

[81] Ran Xin and Usman A Khan. “A linear algorithm for optimization over directed graphs with geometric convergence”. In: *IEEE Control Systems Letters* 2.3 (2018), pp. 315–320.

[82] Jinming Xu, Shanying Zhu, Yeng Chai Soh, and Lihua Xie. “Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes”. In: *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE. 2015, pp. 2055–2060.