Distributed Optimization Over Dependent Random Networks

Adel Aghajan and Behrouz Touri, Senior Member, IEEE

Abstract—We study the averaging-based distributed optimization solvers over random networks. We show a general result on the convergence of such schemes using weight matrices that are row-stochastic almost surely and column-stochastic in expectation 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 theorem, we establish new results on the rate of convergence analysis of 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.

Index Terms—Convex optimization, directed graph, distributed optimization, random networks, spanning tree.

I. INTRODUCTION

Distributed optimization has received increasing attention in recent years due to its applications in robust sensor network control [24], distributed estimation and signal processing [4], [30], power networks [7], [9], and game theory [14].

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 function $f_i(z)$. 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-known algorithm that achieves this is, what we refer to as, the averaging-based distributed optimization solver where 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 [21]. However, to show the convergence of such an algorithm, the corresponding weight matrices are often assumed to be doubly stochastic matrices. In many distributed settings, making a row-stochastic weight matrix is an easy task, and in some cases, e.g., when each node has information about its out-degree, making a column-stochastic matrix is an easy one. However, ensuring both, i.e., making a doubly stochastic matrix distributively is a challenging task. Therefore, it is assumed that at each round, the doubly stochastic weight matrix is given, or an extra effort is needed to construct such a weight matrix. For example, if the instantaneous communication network is undirected and each node knows the degrees of itself and its neighbors, the underlying doubly stochastic weight matrix can be constructed using the Metropolis–Hastings algorithm/method.

A solution to overcome this challenge is to establish more complicated distributed algorithms that effectively reconstruct the average-state distributively. The first algorithm in this category was proposed in [34], [35], and [36], which is called push-sum or subgradient-push, and later was extended for time-varying networks [20]. 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 [17] and [37]. 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 involve division operation that results in theoretical challenges in establishing their stability in random networks [26], [27].

Another solution to address this challenge is to use gossip [13] and broadcast gossip [19] algorithms over random networks. The weight matrices of gossip algorithms are row-stochastic and in-expectation column-stochastic. This fact was generalized in [18], where it is proven that it is sufficient to have row-stochastic weight matrices, that are column-stochastic in expectation. 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.). In [16], the broader class of random networks, which is Markovian networks, was studied in distributed optimization; however, weight matrices were assumed to be doubly stochastic almost surely. In addition, our work is closely related to the existing works on distributed averaging on random networks [12], [31], [32], [33].
In this article, 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.

This article is organized as follows: We conclude this section by introducing mathematical notations that will be used subsequently. In Section II, we formulate the problem of interest and state the main result of this work and discuss some of its immediate consequences. To prove the main result, first we study the behavior of the distributed averaging dynamics over random networks in Section IV. Then, in Section V, we extend this analysis to the dynamics with arbitrary control inputs. Finally, the main result is proved in Section VI. We conclude this work in Section VII.

Notation and Basic Terminology: The following notation will be used throughout the article. We let \([n] \triangleq \{1, \ldots, n\}\). We denote the space of real numbers by \(\mathbb{R}\) and natural (positive integer) numbers by \(\mathbb{N}\). We denote the space of \(n\)-dimensional real-valued vectors by \(\mathbb{R}^n\). In this article, 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 \(i\)th coordinate of \(x\). We denote the all-one vector in \(\mathbb{R}^n\) by \(e^n = (1, 1, \ldots, 1)^T\). We drop the superscript \(n\) in \(e^n\) whenever the dimension of the space is understandable from the context. A non-negative matrix \(A\) is a row-stochastic (column-stochastic) matrix if \(A e = e (e^T A = e^T)\).

Let \((\Omega, \mathcal{F}, \mathbb{P})\) 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) \(\{x(1), \ldots, x(t)\}\), we denote the sigma-algebra generated by these random variables by \(\sigma(x(0), \ldots, x(t))\). We say that \(\{F(t)\}\) is a filtration for \((\Omega, \mathcal{F})\) if \(F(0) \subseteq F(1) \subseteq \cdots \subseteq \mathcal{F}\). We say that a random process \(\{V(t)\}\) of random variables, vectors, or matrices is adapted to \(\{F(t)\}\) if \(V(t)\) is measurable with respect to \(F(t)\). Here, \(F(t)\) essentially captures the information about the outcome of the random experience available by iteration \(t\); as time passes, we have more information about the outcome that is captured by the fact that \(F(0) \subseteq F(1) \subseteq F(t) \subseteq \cdots\).

Throughout this article, we mainly deal with 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 \(\mathcal{E} \subset [n] \times [n]\). 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 in the graph. 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 \(\mathcal{E}^\gamma (A) = \{(j, i) : i, j \in [n], a_{ij} > \gamma\}\). Later, we fix the value \(0 < \gamma < 1\) throughout the article, and hence, unless otherwise stated, for notational convenience, we use \(\mathcal{G}(A)\) and \(\mathcal{E}(A)\) instead of \(\mathcal{G}^\gamma (A)\) and \(\mathcal{E}^\gamma (A)\).

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

\[
\theta f(x) + (1 - \theta) f(y) \geq f(\theta x + (1 - \theta) y).
\]

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., [23, Th. 3.1.15]). For convenience and due to the frequent use of \(\ell_2\) norm in the article, we use \(\|\cdot\|\) to denote the \(\ell_2\) norm \(\|x\| = \sqrt{\sum_{i=1}^{m} x_i^2}\).

Also, we denote \(\ell_\infty\) norm with \(\|x\|_\infty \triangleq \max_{i \in [m]} |x_i|\).

II. Problem Formulation and Main Result

In this section, we discuss the main problem and the main result of this work. The proof of the result is provided in the subsequent sections.

A. General Framework

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 article is to solve

\[
\arg \min_{z \in \mathbb{R}^m} F(z)
\]

distributively with the following assumption on the objective function.

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

1) \(f_i\) is a convex function over \(\mathbb{R}^m\) for all \(i \in [n]\).
2) The optimizer set \(Z \triangleq \arg \min_{z \in \mathbb{R}^m} F(z)\) is nonempty.
3) The subgradients of \(f_i\) are uniformly upper bounded, i.e., for all \(g \in \nabla f_i(z)\), \(\|g\| \leq \beta_i\) for all \(z \in \mathbb{R}^m\) and all \(i \in [n]\). We let \(L \triangleq \sum_{i=1}^{n} \beta_i\).

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

A distributed solution of (1) was first proposed in [21] using the following deterministic dynamics:

\[
x_i(t+1) = \sum_{j=1}^{n} w_{ij}(t+1) x_j(t) - \alpha(t) g_i(t)
\]

for \(t \geq 0\), initial conditions \(x_i(0) \in \mathbb{R}^m\) for all \(i \in [n]\), where \(g_i \in \mathbb{R}^m\) is a subgradient of \(f_i(z)\) at \(z = x_i(t)\) for \(i \in [n]\), and \((\alpha(t))\) is a step-size sequence (in [21] 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

\[
x(t+1) = W(t+1) x(t) - \alpha(t) g(t)
\]

The dynamics work for any initial time \(t_0\), but since it does not make any difference, in this article, we set the initial to be zero.
where \( g(t) = [g_1(t), \ldots, g_n(t)]^T \) is the vector of the subgradient vectors and matrix multiplication should be understood over the vector-field \( \mathbb{R}^m \), i.e.,

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

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

### B. Our Contribution

In the article, we consider the random variation of (2), i.e., when \( \{W(t)\} \) is a chain of random matrices. This random variation was first studied in [15] 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 [16]. The dynamics (2) with i.i.d. weight matrices that are row-stochastic almost surely and column-stochastic in-expectation was studied in [18]. A special case of [18] is the asynchronous gossip algorithm that was introduced in [19].

In this work, we provide an overarching framework for the study of (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 the following.

1. For all \( t \geq 0 \), \( W(t) \) is row-stochastic almost surely.
2. For every \( t > 0 \), \( \mathbb{E}[W(t) | \mathcal{F}(t-1)] \) is column-stochastic (and hence, doubly stochastic) almost surely.

Similar to other works in this domain, our goal is to ensure that \( \lim_{t \to \infty} x_i(t) = z \) almost surely for some optimal \( z \in \mathcal{Z} \) for all \( i \in [n] \). To reach such a consensus value, we need to ensure a sufficient 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 \( t \geq 0 \)

1. Every node in \( G(W(t)) \) has a self-loop, almost surely.
2. There exists an integer \( B > 0 \) such that the random graph \( G_B(t) = ([n], \mathcal{E}_B(t)) \) where

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

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 \( \bigcup_{\tau=t+1}^{t+B} \mathcal{E}(W(\tau)) \) for time-varying networks. In a random setting, the connectivity condition for random i.i.d. networks is that \( \mathcal{E}(\mathbb{E}[W]) \) has a rooted tree. A natural generalization of this condition to dependent random networks is that \( \bigcup_{\tau=t+1}^{t+B} \mathcal{E}(\mathbb{E}[W(\tau) | \mathcal{F}(\tau-1)]) \) has a spanning rooted tree almost surely. We further generalize this condition to Assumption 3-(2). This is due to the following lemma, which is proved in Appendix.

**Lemma 1:** If the random graph with the vertex set \( [n] \) and the edge set \( \bigcup_{\tau=t+B}^{t+1} \mathcal{E}(\mathbb{E}[W(\tau) | \mathcal{F}(\tau-1)]) \) has a spanning rooted tree almost surely, then for some \( \gamma \in (0, \gamma) \) the form \( \gamma = K(n, \mathcal{B}) \gamma \) with \( K(n, \mathcal{B}) \in (0, 1] \) depending only on \( n \) and \( \mathcal{B} \), the random graph with the vertex set \( [n] \) and the edge set \( \bigcup_{\tau=t+B}^{t+1} \mathcal{E}(\mathbb{E}[W(\tau) | \mathcal{F}(t)]) \) 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

\[
\sum_{t=0}^{\infty} \alpha(t) = \infty \quad \text{and} \quad \sum_{t=0}^{\infty} \frac{\alpha^2(t)}{2} < \infty.
\]

The main result of this article is the following theorem.

**Theorem 1:** Under the Assumptions 1–4 on the model and the dynamics (2), \( \lim_{t \to \infty} x_i(t) = z^* \) almost surely for all \( i \in [n] \) and all initial conditions \( x_i(0) \in \mathbb{R}^m \), where \( z^* \) is a random vector that is supported on the optimal set \( \mathcal{Z} \).

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

1) **Gossip-based sequential solvers:** Gossip algorithms, which were originally studied in [2] and [3], have been used in solving distributed optimization problems [13], [19]. 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 use Theorem 1 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 1:** Consider a connected undirected network\(^2\) \( G = ([n], E) \). 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 \( \frac{1}{2} \), it 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

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

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

\[
x_{\ell(t+1)}(t+1) = \frac{1}{2}(x_{s(t+1)}(t) + x_{\ell(t)}(t)) - \alpha(t)g_{\ell(t+1)}(t).
\]

\(^2\)Indeed, the underlying graph do not need to be time-invariant. It can be shown that this example and its assertion can be extended to underlying undirected time-varying graphs that are \( B \)-strongly connected (as defined in [20]).
For the other agents $i \neq \ell(t + 1)$, we set
\[ x_i(t + 1) = x_i(t) - \alpha(t)g_i(t). \]

Let $F(t) = \sigma(x(0), \ldots, x(t), \ell(t))$, and the weight matrix $W(t) = [w_{ij}(t)]$ be
\[
w_{ij}(t) = \begin{cases} \frac{1}{2}, & i = j = \ell(t) \\ \frac{1}{2}, & i = \ell(t), j \in \{s(t), \ell(t - 1)\} \setminus \{\ell(t)\} \\ 1, & i \neq \ell(t) \end{cases}
\]
which is the weight matrix of this scheme. Note that $\mathbb{E}[W(t)|F(t - 1)] = V(\ell(t - 1))$ where $R(h) = [r_{ij}(h)]$ with
\[
r_{ij}(h) = \begin{cases} \frac{1}{\delta}, & i = j = h \\ \frac{1}{\delta_i}, & i = h, (i, j) \in E \\ 1, & i \neq h \end{cases}
\]
where $\delta_i$ is the degree of the node $i$. Note that the matrix $\mathbb{E}[W(t)|F(t - 1)]$ is doubly stochastic, satisfies Assumption 3-(1), and only depends on $\ell(t - 1)$. Now, we need to check whether $\{W(t)\}$ satisfies Assumption 3-(2). We have
\[
\mathbb{E}[W(t + n)|F(t)] = \mathbb{E}\left[\mathbb{E}[W(t + n) | F(t + n - 1)] | F(t)\right]
\]
\[
= \mathbb{E}\left[\frac{1}{n} \left(\sum_{z=1}^{n} R(\ell(t + n - 1))1_{1_{\ell(t + n - 1)} = z}\right) | F(t)\right]
\]
\[
= \frac{1}{n} \left(\sum_{z=1}^{n} \mathbb{E}[R(1)1_{1_{\ell(t + n - 1)} = z}] | F(t)\right)
\]
\[
= \frac{1}{n} \left(\sum_{z=1}^{n} \mathbb{E}[R(1)1_{1_{\ell(t + n - 1)} = z}] | \mathbb{F}(t)\right).
\]

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 $\Delta = \max_{i \in [n]} \delta_i$. Therefore, we have $\mathbb{E}[1_{1_{\ell(t + n - 1)} = z}] | \mathbb{F}(t) > 0$ for all $i \in [n]$ and $t$, and hence, Assumption 3-(2) is satisfied with $B = n$.

2) **Robustness to link failure:** Our result shows that (2) is robust to random link-failures. Note that the results such as [15] 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 $p(t)$ 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
\[
\mathbb{E}[b_{ij}(t)|F(t - 1)] = p(t)
\]
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). Define $W(t) = [w_{ij}(t)]$ as follows:
\[
w_{ij}(t) = \begin{cases} a_{ij}(t)b_{ij}(t), & i \neq j \\ 1 - \sum_{t' \neq i} a_{t'j}(t)b_{t'j}(t), & i = j \end{cases}
\]
Note that $W(t)$ is row-stochastic, and since $A(t)$ is column-stochastic, $\mathbb{E}[W(t)|F(t - 1)]$ is column-stochastic. Thus, Theorem 1, using $W(t)$, translates to a theorem on robustness of the distributed dynamics (2): As long as the connectivity conditions of Theorem 1 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 [22, Prop. 4] (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 [18] would not imply such a robustness result.

### III. Theorem 1: Sketch of the Proof

Here, we provide the sketch of the proof of the main result (see Theorem 1) to assist with its readability. We can divide the proof into two main steps as follows.

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

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

To show Step 1, we fix a $z \in Z$ and use the Lyapunov (like) function $V(t) = \|\mathbb{x}(t) - z\|^2$. In Lemma 13, we show that this Lyapunov function satisfies
\[
\mathbb{E}[V(t + 1)|F(t)] \leq V(t) - b(t) + c(t)
\]
where
\[
b(t) \triangleq -2\alpha(t)\frac{F(\mathbb{x}(t)) - F(z)}{n}, \quad \text{and} \quad c(t) \triangleq \alpha^2(\frac{L^2}{n} + \sum_{i=1}^{n} \|x_i(t) - \mathbb{x}(t)\|^2 + \frac{4\alpha(t)}{n} \sum_{i=1}^{n} \mathbb{E}[L_i|\mathbb{x}_i(t) - \mathbb{x}(t)]](6)
\]
To analyze (5), we apply Robbins–Siegmund Theorem [28], which plays a key role in the proof of the above Step 1.

**Theorem 2 (Robbins–Siegmund Theorem [28]):** Suppose that a nonnegative random process $\{V(t)\}$ (adapted to a filtration $\{F(t)\}$) satisfies
\[
\mathbb{E}[\mathbb{V}(t + 1)|F(t)] \leq (1 + \bar{\alpha}(t))\mathbb{V}(t) - \bar{b}(t) + \bar{c}(t)
\]
where \( \tilde{a}(t), \tilde{b}(t), \tilde{c}(t) \geq 0 \) almost surely for all \( t \). Then if \( \sum_{t=0}^{\infty} \tilde{a}(t) < \infty \) and \( \sum_{t=0}^{\infty} \tilde{b}(t) < \infty \) almost surely, \( \lim_{t \to \infty} V(t) \) exists and \( \sum_{t=0}^{\infty} \tilde{b}(t) < \infty \) almost surely.

Note that in (6), \( b(t), c(t) \geq 0 \) for all \( t \). To apply the Robbins–Siegmund result for (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 \( \sum_{t=0}^{\infty} \alpha^2(t) < \infty \), and

\[
\sum_{t=0}^{\infty} \alpha(t) \| x_i(t) - \bar{x}(t) \| < \infty
\]

almost surely, for all \( i \in [n] \). To establish this, in Lemma 9, we find an upper bound on the diameter of \( x(t) \), i.e., \( d(x(t)) \), which is defined in (15). Combining \( \| x_i(t) - \bar{x}(t) \| \leq d(x) \) [see Lemma 3-(c)] and Lemma 9, we arrive at

\[
\mathbb{E}\left[ \| x_i(t) - \bar{x}(t) \|^2 \right] \leq M\alpha^2(t)
\]

for some constant \( M > 0 \).

While in (8), we show that \( x_i(t) \) converges to \( \bar{x}(t) \) (in the second moment) with the convergence rate \( \alpha^2(t) \), to prove Step II, we need to show that it converges to \( \bar{x}(t) \) almost surely. However, we provide a stronger result and in Lemma 11, we show that

\[
\| x_i(t) - \bar{x}(t) \| \leq \tilde{M}\alpha(t)
\]

for some constant \( \tilde{M} > 0 \).

To show (8) and (9), we study the conditional expectation of the diameter of \( x(t) \). To do so, we derive

\[
\mathbb{E}[d(x(t))|\mathcal{F}(\tau)] \leq \mathbb{E}[\text{diam}(\Phi(t, \tau))|\mathcal{F}(\tau)]d(x(\tau))
\]

\[
+ \tilde{L} \sum_{s=\tau}^{t-1} \mathbb{E}[\text{diam}(\Phi(t, s + 1))|\mathcal{F}(\tau)]\alpha(s)
\]

from the main dynamics (2) for some constant \( \tilde{L} > 0 \) where \( \text{diam}(A) \) is the diameter of the matrix \( A \) and defined in (13). Therefore, we need to investigate \( \mathbb{E}[\text{diam}(\Phi(t, \tau))|\mathcal{F}(\tau)] \) for all \( t \geq \tau \). In Lemma 6, we show that \( \mathbb{E}[\text{diam}(\Phi(t, \tau))|\mathcal{F}(\tau)] \) goes to zero fast, i.e.,

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

To prove (11), using Assumption 3 and Assumption 2-(a), in Lemma 4, we show that for a large enough \( T \),

\[
\mathbb{E}[\text{diam}(\Phi(T, \tau))|\mathcal{F}(\tau)] \leq 1 - \theta
\]

for some \( \theta > 0 \). Using (11), in Lemma 9, we prove (8) and complete the proof of Step I.

To prove Step II, using Assumption 4 \( (\alpha(t) \leq Kt^{-\beta}) \), first we simplify (10) to arrive at a recursive stochastic inequality

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

for some \( \tilde{K} > 0 \). However, since \( \sum_{t=0}^{\infty} t^{-\beta} \) is not necessarily summable, we cannot use the standard Robbins–Siegmund Theorem [28] to argue \( d(x(t)) \to 0 \) based on this inequality. To remedy this, we prove a martingale-type result in Lemma 10, which helps us to prove (9) (in Lemma 11) using the facts that \( \mathbb{E}[\text{diam}(\Phi(t, \tau))|\mathcal{F}(\tau)] < 1 \) for large enough \( t - \tau \), and \( \text{diam}(\Phi(t, \tau)) \leq 1 \) for all \( t \geq \tau \). This step completes the proof of Step II.

IV. AUTONOMOUS AVERAGING DYNAMICS

To prove Theorem 1, we need to study the time-varying distributed averaging dynamics with a particular control input (gradientlike 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

\[
x(t + 1) = W(t + 1)x(t)
\]

where \( \{W(t)\} \) satisfying Assumption 3. Defining transition matrix \( \Phi(t, \tau) \triangleq W(t) \cdots W(\tau + 1) \) with \( \Phi(\tau, \tau) = I \), we have \( x(t) = \Phi(t, \tau)x(\tau) \). Note that since \( W(t) \) is 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 \( i \), \( \lim_{t \to \infty} \| x_i(t) - \bar{x} \| = 0 \) almost surely, for all choices of initial condition \( x(0) \in \mathbb{R}^n \) in (12) and some random vector \( \bar{x} = \bar{x}(\bar{x}(0)) \). It can be shown that an equivalent condition for consensus is to have \( \lim_{t \to \infty} \Phi(t, 0) = e \pi^T \phi(0) \) for a random stochastic vector \( \pi(0) \in \mathbb{R}^n \), almost surely.

For a matrix \( A = [a_{ij}] \), let

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

and the mixing parameter

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

Note that for a row-stochastic matrix \( A, \text{diam}(A) \in [0, 1] \). For a vector \( x = [x_i] \) where \( x_i \in \mathbb{R}^m \) for all \( i \), let

\[
d(x) = \max_{i,j\in[n]} \| x_i - x_j \|.
\]

Note that \( d(x) \leq 2 \max_{x\in[0]} \| x \| \). Also, if we have consensus, then \( \lim_{t \to \infty} \| x(t) \| = 0 \). And in fact, the reverse implications are true [6], i.e., a chain achieves consensus if and only if \( \lim_{t \to \infty} \| x(t) \| = 0 \) for all \( x(0) \in \mathbb{R}^n \) or \( \lim_{t \to \infty} \| x(t) \| = 0 \).

The following results relating to the above quantities are useful for our future discussions. First, we start with reviewing some existing results relating to the above quantities.

Lemma 2: For \( n \times n \) row-stochastic matrices \( A, B \), we have the following.

1) \( \text{diam}(AB) \leq (1 - \Lambda(A))\text{diam}(B) \) [11], [29].

2) \( \text{diam}(A) = 1 - \Lambda(A) \), and hence [8]

\[
\text{diam}(AB) \leq \text{diam}(A)\text{diam}(B).
\]

3) \( d(Ax) \leq \text{diam}(A)d(x) \) for all \( x \in \mathbb{R}^n \).
The following result also relates several of these quantities. The proofs of most parts are based on similar results established in [5] for $m = 1$.

**Lemma 3:** For any $n \times n$ row-stochastic matrices $A, B$, we have the following.

1. $d(Ax) \leq \text{diam}(A) d(x)$ for all $x \in (\mathbb{R}^n)^n$.
2. $d(x + y) \leq d(x) + d(y)$ for all $x, y \in (\mathbb{R}^n)^n$.
3. $\|x_i - \sum_{j=1}^{n} x_j \pi_j\| \leq \sqrt{n} d(x)$ for all $i \in [n]$, $x \in (\mathbb{R}^n)^n$, and any stochastic vector $\pi \in [0,1]^n$ (i.e., $\sum_{j=1}^{n} \pi_j = 1$).

**Proof:** The proof is provided in Appendix.

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 $\text{diam}(\Phi(t_2, \tau_2) \text{diam}(\Phi(t_1, \tau_1)))$.

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

$$
\mathbb{E}[\mathcal{A}(\Phi(n^2 + s)B, sB)] \leq \theta/s.
$$

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

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

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 $V = \{\omega \ | \ \forall k \in \mathbb{T}, \mathcal{T}_k \in \mathbb{T}\}$. 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

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

and $\mathcal{T}_k$ 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 (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

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

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

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

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

$$
\begin{align*}
1 & \{U_k\} \geq 1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \in \mathbb{T}\} 1 \{u_k \in \mathcal{D}_B(k)\} + 1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} \\
\text{(17)}
\end{align*}
$$

holds for $k \geq 1$.

On the other hand, from Tower rule (see, e.g., [10, Th. 5.1.6]), we have

$$
\begin{align*}
\mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} 1 \{u_k \in \mathcal{D}_B(k)\} & | \mathcal{F} \\
& = \mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} \mathbb{E}1 \{u_k \in \mathcal{D}_B(k)\} | \mathcal{F}_{k-1} | \mathcal{F} \\
& = \mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} | \mathcal{F}_{k-1} | \mathcal{F}.
\end{align*}
$$

(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

$$
\begin{align*}
\mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} | \mathcal{F}_{k-1} | \mathcal{F} \\
& = \mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} | \mathcal{F}_{k-1} | \mathcal{F}.
\end{align*}
$$

Therefore,

$$
\begin{align*}
\mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} | \mathcal{F}_{k-1} | \mathcal{F} & = \mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} | \mathcal{F}_{k-1} | \mathcal{F} \\
& \geq \mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} | \mathcal{F}_{k-1} | \mathcal{F}.
\end{align*}
$$

Therefore, (17) implies

$$
\begin{align*}
\mathbb{E}1 \{U_{k-1}\} | \mathcal{F} \\
& \geq \mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} | \mathcal{F}_{k-1} | \mathcal{F} + \mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \in \mathbb{T}\} | \mathcal{F} \\
& \geq \mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} | \mathcal{F}_{k-1} | \mathcal{F} + \mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \in \mathbb{T}\} | \mathcal{F} \\
& \geq \mathbb{E}1 \{U_{k-1}\} 1 \{\mathcal{T}_{k-1} \notin \mathbb{T}\} | \mathcal{F}_{k-1} | \mathcal{F}.
\end{align*}
$$

and hence, $\mathbb{E}1 \{U_{k-1}\} | \mathcal{F} \geq p^k$. Finally, since $\mathcal{T}_{n^2}$ has a spanning rooted tree, from [15, Lemma 1], we have

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

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

$$
\begin{align*}
\mathbb{E} \Lambda(\Phi(n^2 + s)B, sB)) | \mathcal{F}(sB) & \geq \nu^{n^2B} \mathbb{E}1 \{U_{n^2}\} | \mathcal{F} \\
& \geq \nu^{n^2B} p^{n^2} \triangleq \theta > 0.
\end{align*}
$$

\[\square\]
Finally, we need the following result that is a straightforward implication of the product and tower rules for conditional expectation and is proved in Appendix.

**Lemma 5:** For a nonnegative random process \(\{Y(k)\}\), adapted to a filtration \(\{\mathcal{F}(k)\}\), let
\[
E[Y(k) | \mathcal{F}(k-1)] \leq a(k)
\]
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. Then, for any \(\sigma\)-algebra \(\bar{\mathcal{F}} \subseteq \mathcal{F}(K_1-1)\), we have almost surely
\[
E \left[ \prod_{k=K_1}^{K_2} Y(k) | \bar{\mathcal{F}} \right] \leq \prod_{k=K_1}^{K_2} a(k).
\]

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

**Lemma 6:** Under Assumptions 2-(a) and 3, there exist \(0 < C\) and \(0 \leq \lambda < 1\) such that for every \(0 \leq \tau_1 \leq t_1\) and \(0 \leq \tau_2 \leq t_2\) with \(\tau_1 \leq \tau_2\), we have almost surely
\[
E[\text{diam}(\Phi(t_2, \tau_2)) | \mathcal{F}(\tau_1)] \leq C^\lambda t_2 - \tau_1, t_2 - \tau_2.
\]

**Proof:** First, we prove
\[
E[\text{diam}(\Phi(t_2, \tau_2)) | \mathcal{F}(\tau_1)] \leq C\lambda t_2 - \tau_1, t_2 - \tau_2
\]
for some \(0 < C\) and \(0 \leq \lambda < 1\). Let \(s \triangleq \left[ \frac{t}{n^2} \right] \) and \(K \triangleq \left[ \frac{t-sB}{n^2 B} \right]\). Note that
\[
\text{diam}(\Phi(t_2, \tau_2)) = \text{diam} \left( \prod_{k=1}^{K} \Phi(sB + kn^2B, sB + (k-1)n^2B) \right)
\]
\[\leq \text{diam}(\Phi(t_2, sB + Kn^2B)) \text{diam}(\Phi(sB, \tau_2)) \prod_{k=1}^{K} \text{diam}(\Phi(sB + kn^2B, sB + (k-1)n^2B)) \]
\[\leq \prod_{k=1}^{K} (1 - \Lambda(\Phi(sB + kn^2B, sB + (k-1)n^2B))) \]
where \((a)\) follows from Lemma 2-(b), and \((b)\) follows from the fact that \(\text{diam}(A) \leq 1\) for all row-stochastic matrices \(A\) and Lemma 2-(b). Therefore, we have
\[
E[\text{diam}(\Phi(t_2, \tau_2)) | \mathcal{F}(\tau_1)] \leq \prod_{k=1}^{K} (1 - \Lambda(\Phi(sB + kn^2B, sB + (k-1)n^2B))) \mathcal{F}(\tau_1)
\]
\[\leq (1 - \theta)^K \leq C(1 - \theta) \frac{t_2 - \tau_2}{n^2 B}
\]
where \(C = (1 - \theta)^{-1} - \frac{1}{n^2} \) and \((a)\) follows from Lemmas 4 and 5 with \(Y(k) = 1 - \Lambda(\Phi(sB + kn^2B, sB + (k-1)n^2B))\), and \(\bar{\mathcal{F}} = \mathcal{F}(\tau)\). Since \(\theta > 0\), we have \(\tilde{\lambda} \triangleq (1 - \theta) \frac{t_2 - \tau_2}{n^2 B} < 1\).

To prove the main statement, we consider the following two cases:

- **i)** intervals \((\tau_1, t_1)\) and \((\tau_2, t_2)\) do not have an intersection;
- **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
\[
E[\text{diam}(\Phi(t_2, \tau_2)) | \mathcal{F}(\tau_1)] \leq \tilde{C}\lambda^{t_2 - \tau_1} \tilde{C}\lambda^{t_2 - \tau_2}
\]
which follows from (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
\[\left( \tau_1, t_1 \right) \cup (\tau_2, t_2) = (s_1, s_2) \cup (s_2, s_3) \cup (s_3, s_4)\]
for \(s_1 \leq s_2 \leq s_3 \leq s_4\) where \((s_2, s_3) \triangleq (\tau_1, t_1) \cap (\tau_2, t_2)\), \((s_1, s_2) \triangleq (\tau_1, t_1) \cup (\tau_2, t_2)\). Using this, it can be verified that
\[
E[\text{diam}(\Phi(t_2, \tau_2)) | \mathcal{F}(\tau_1)] \leq \tilde{C}\lambda^{s_2 - s_1} \tilde{C}\lambda^{s_3 - s_2} \tilde{C}\lambda^{s_4 - s_3}
\]
\[\leq \tilde{C}\lambda^{s_2 - s_1} \sqrt{\lambda} \tilde{C}\lambda^{s_3 - s_2} \sqrt{\lambda} \tilde{C}\lambda^{s_4 - s_3} \]
\[\leq \tilde{C}\lambda^{t_1 - \tau_1} \sqrt{\lambda} \tilde{C}\lambda^{t_2 - \tau_2}\]
where \((a)\) follows from Lemma 2-(b), \((b)\) follows from \(\text{diam}(A) \leq 1\) for all row-stochastic matrices \(A\), and \((c)\) follows from (20) and Lemma 5. Letting \(C \triangleq \max\{C^2, C^3\}\) and \(\lambda \triangleq \sqrt{\lambda}\), we arrive at the conclusion.

**V. AVERAGING DYNAMICS WITH GRADIENT-FLOW LIKE FEEDBACK**

In this section, we study the controlled linear time-varying dynamics
\[
x(t + 1) = W(t + 1)x(t) + u(t).
\]
Note that the feedback \(u(t) = -\alpha(t)g(t)\) leads to the dynamics (2). The goal of this section is to establish bounds on the convergence-rate of \(d(x)\) (to zero) in-expectation and almost surely for a class of regularized input \(u(t)\).

We start with the following two lemmas.

**Lemma 7:** For dynamics (21) and every \(0 \leq \tau \leq t\), we have
\[
d(x(t)) \leq \text{diam}(\Phi(t, \tau)) d(x(\tau)) + \sum_{s=\tau}^{t-1} \text{diam}(\Phi(t, s + 1)) d(u(s)).
\]
Proof: Note that the general solution for the dynamics (21) is given by
\[ x(t) = \Phi(t, \tau)x(\tau) + \sum_{s=\tau}^{t-1} \Phi(t, s+1)u(s). \] (23)
Therefore, using the sublinearity property of \( d(\cdot) \) [see Lemma 3-(b)], we have
\[ d(x(t)) \leq d(\Phi(t, \tau)x(\tau)) + \sum_{s=\tau}^{t-1} d(\Phi(t, s+1)u(s)) \]
\[ \leq \text{diam}(\Phi(t, \tau))d(x(\tau)) \]
\[ + \sum_{s=\tau}^{t-1} \text{diam}(\Phi(t, s+1))d(u(s)) \]
where the last inequality follows from Lemma 3-(a).

Lemma 8: Let \( \{\beta(t)\} \) be a positive (scalar) sequence such that \( \lim_{t \to \infty} \frac{\beta(t)}{\mu(t)} = 1 \). Then, for any fixed \( \theta \in [0, 1) \) and \( \tau > 0 \), there exists some \( M > 0 \) independent of \( t \) 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 \( E[d(x(t))] \) and \( E[d^2(x(t))] \) approach to zero when the diameter of the control input \( d(u(t)) \) goes to zero. Since \( E[d^2(x(t))] \geq E^2[d(x(t))] \), it suffice to study convergence rate of \( E[d^2(x(t))] \).

Lemma 9: Under Assumptions 2-(a), 3, and 4, if almost surely \( d(u(t)) < q\alpha(t) \) for some \( q > 0 \), then \( \frac{E[d^2(x(t))]}{\alpha^2(t)} \leq M \) for some \( M > 0 \) and all \( t \geq 0 \).

Proof: Taking the square of both sides of (22), for \( t \geq \tau \geq 0 \), we have
\[ d^2(x(t)) \leq \text{diam}^2(\Phi(t, \tau))d^2(x(\tau)) \]
\[ + 2\text{diam}(\Phi(t, \tau))d(x(\tau)) \sum_{s=\tau}^{t-1} \text{diam}(\Phi(t, s+1))d(u(s)) \]
\[ + \sum_{s=\tau}^{t-1} \sum_{\ell=\tau}^{s-1} \text{diam}(\Phi(t, s+1))d(u(s))diam(\Phi(t, \ell+1))d(u(\ell)). \]
Taking the expectation of both sides of the above inequality, and using \( d(u(t)) < q\alpha(t) \) almost surely, we have
\[ E[d^2(x(t))] \leq E[\text{diam}^2(\Phi(t, \tau))d^2(x(\tau))] \]
\[ + 2 \sum_{s=\tau}^{t-1} E[\text{diam}(\Phi(t, \tau))d(x(\tau))diam(\Phi(t, s+1))d(u(s))] \]
\[ + \sum_{s=\tau}^{t-1} \sum_{\ell=\tau}^{s-1} E[\text{diam}(\Phi(t, s+1))d(u(s))diam(\Phi(t, \ell+1))d(u(\ell))] \]
\[ \leq E\left[ \text{diam}^2(\Phi(t, \tau))d^2(x(\tau)) \right] \]
\[ + 2 \sum_{s=\tau}^{t-1} E[E[\text{diam}(\Phi(t, s+1))diam(\Phi(t, s+1))d(u(s))] \]
\[ + \sum_{s=\tau}^{t-1} \sum_{\ell=\tau}^{s-1} E[\text{diam}(\Phi(t, s+1))d(u(s))diam(\Phi(t, \ell+1))d(u(\ell))] \]
\[ \leq \frac{CqM^2}{\lambda^2}. \]
Therefore, using Lemma 6, we have
\[ E[d^2(x(t))] \]
\[ \leq C\lambda^2(t-\tau)E[d^2(x(\tau))] + \frac{2Cq}{\lambda} \lambda^{t-\tau}E[d(x(\tau))][\sum_{s=\tau}^{t-1} \lambda^{t-s}\alpha(s)] \]
\[ + \frac{Cq^2}{\lambda^2} \sum_{s=\tau}^{t-1} \sum_{\ell=\tau}^{s-1} \alpha(s)\alpha(\ell)\lambda^{t-s}\lambda^{t-\ell} \]
\[ \leq C\lambda^2(t-\tau)E[d^2(x(\tau))] + \frac{2CqM}{\lambda} \lambda^{t-\tau}E[d(x(\tau))]\alpha(t) \]
\[ + \frac{Cq^2M^2}{\lambda^2} \]
where the last inequality follows from Lemma 8 and the fact that
\[ \sum_{s=\tau}^{t-1} \sum_{\ell=\tau}^{s-1} \alpha(s)\alpha(\ell)\lambda^{t-s}\lambda^{t-\ell} = \left( \sum_{s=\tau}^{t-1} \alpha(s)\lambda^{t-s} \right)^2. \]
Dividing both sides of the above inequality by \( \alpha^2(t) \) and noting \( \frac{\alpha(\tau)}{\alpha(\tau+1)} = \prod_{k=\tau}^{t-1} \frac{\alpha(k+1)}{\alpha(k)} \), we have
\[ \frac{E[d^2(x(t))]}{\alpha^2(t)} \leq C \frac{E[d^2(x(\tau))]}{\alpha^2(\tau)} \left( \prod_{k=\tau}^{t-1} \frac{\alpha(k+1)}{\alpha(k)} \right)^2 \]
\[ + \frac{2CqM E[d(x(\tau))] \alpha(t)}{\lambda^\tau} \left( \prod_{k=\tau}^{t-1} \frac{\alpha(k+1)}{\alpha(k)} \right) + \frac{Cq^2 M^2}{\lambda^2}. \]
Since \( \lim_{t \to \infty} \frac{\alpha(\tau)}{\alpha(\tau+1)} = 1 \), for any \( \bar{\lambda} \in (\lambda, 1) \), there exists \( \bar{\tau} \) such that for \( \tau \geq \bar{\tau} \), we have \( \frac{\mu(t)}{\mu(t+1)} \leq \bar{\lambda} \). Therefore
\[ \frac{E[d^2(x(t))]}{\alpha^2(t)} \leq C \frac{E[d^2(x(\tau))]}{\alpha^2(\tau)} \left( \frac{\alpha(\tau)}{\alpha(\tau+1)} \right)^2 \]
\[ + \frac{2CqM E[d(x(\tau))] \alpha(t)}{\lambda^\tau} \left( \frac{\alpha(\tau)}{\alpha(\tau+1)} \right) + \frac{Cq^2 M^2}{\lambda^2}. \]
Taking the limit of the above inequality, we get
\[ \lim_{t \to \infty} \frac{E[d^2(x(t))]}{\alpha^2(t)} \leq C \frac{E[d^2(x(\tau))]}{\alpha^2(\tau)} \left( \frac{\alpha(\tau)}{\alpha(\tau+1)} \right)^2 \]
\[ + \frac{2CqM E[d(x(\tau))] \alpha(t)}{\lambda^\tau} \left( \frac{\alpha(\tau)}{\alpha(\tau+1)} \right) + \frac{Cq^2 M^2}{\lambda^2} \]
\[ = \frac{Cq^2 M^2}{\lambda^2}. \]
As a result, there exists an \( \bar{M} > 0 \) such that \( \frac{E[d^2(x(t))]}{\alpha^2(t)} \leq \bar{M}. \) ■
To prove the main theorem, we also need to show that $d(x(t))$ converges to zero almost surely (as will be proved in Lemma 11). To do so, we apply the following result, which is proved in Appendix.

**Lemma 10:** Suppose that $\{D(t)\}$ is a nonnegative random (scalar) process such that

$$D(t + 1) \leq a(t + 1)D(t) + b(t), \quad \text{almost surely}$$

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

$$E|a(t + 1) | F(t) | < \chi$$

almost surely for some $\chi < 1$ 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 $\lim_{t \to \infty} d(x(t)) = 0$ (and more) under our connectivity assumption and a regularity condition on the input $u(t)$ for the controlled averaging dynamics (21).

**Lemma 11:** Suppose that $\{W(t)\}$ satisfies Assumption 2-(a) and 3. Then, if $d(u(t)) < qT^{1-\beta}$ almost surely for some $q \geq 0$, we have $\lim_{t \to \infty} d(x(t))t^{\beta} = 0$, almost surely, for $\beta < \tilde{\beta}$.

**Proof:** From inequality (22), we have

$$d(x(k)) \leq \text{diam}(\Phi(k, \tau))d(x(\tau)) + \sum_{s=\tau}^{k-1} \text{diam}(\Phi(s, s + 1))d(u(s))$$

$$(a) \leq \text{diam}(k, \tau))d(x(\tau)) + \sum_{s=\tau}^{k-1} qs^{-\tilde{\beta}}$$

$$\leq \text{diam}(\Phi(k, \tau))d(x(\tau)) + (k - \tau) qT^{-\tilde{\beta}} \quad (25)$$

where $(a)$ follows from $\text{diam}(\Phi(., .)) \leq 1$. Let $C > 0$ and $\lambda \in [0, 1)$ be the constants satisfying the statement of Lemma 6. Since $\lambda < 1$, for $T = \lceil \frac{\log C}{\log \lambda} \rceil + 1$, we have $\lambda \leq C\lambda T < 1$. Then, Lemma 6 implies that

$$E[\text{diam}(\Phi(T(t + 1), T(t)) | F(T)) | F(T(t)) | F(T(t)) \leq C\lambda T = \lambda < 1. \quad (26)$$

Let $D(t) \triangleq d(x(T(t)))$. From inequality (25), for $\tau = Tt$ and $k = T(t + 1)$, we have

$$D(t + 1) \leq \text{diam}(\Phi(T(t + 1), T(t)) | F(T)) | F(T(t)) + T^{1-\beta} qT^{-\tilde{\beta}}.$$ 

Taking conditional expected value of both sides of the above inequality given $F(Tt)$, we have

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

By letting $a(t + 1) \triangleq \text{diam}(\Phi(T(t + 1), T(t)))$ and $b(t) \triangleq TqT^{-\tilde{\beta}}$, we are in the setting of Lemma 10. Therefore, by Inequality (26) and $\text{diam}(\Phi(T(t + 1), T(t)) \leq 1$, the conditions of Lemma 10 hold, and hence

$$\lim_{t \to \infty} D(t)t^{\beta} = 0 \quad (27)$$

almost surely.

On the other hand, letting $\tau = T \lfloor \frac{k}{T} \rfloor$ in (25), we have

$$d(x(k)) \leq D\left(\lfloor \frac{k}{T} \rfloor\right) + T^{1-\beta} q\lfloor \frac{k}{T} \rfloor^{-\tilde{\beta}}. \quad (28)$$

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

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

Similarly

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

Therefore, using these inequalities and (28), we get

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

where the last equality follows from (27) and $\tilde{\beta} > \beta$.

**VI. CONVERGENCE ANALYSIS OF THE MAIN DYNAMICS**

Finally, in this section, we will study the main dynamics (2), i.e., the dynamics (21) with the feedback policy $u_i(t) = -\alpha_i(t)g_i(t)$ where $g_i(t) \in \nabla f_i(x_i(t))$. Throughout this section, we let $\bar{x} \triangleq \frac{1}{n}e^T x$ for a vector $x \in \mathbb{R}^m$.

First, we prove an inequality (see Lemma 13) that plays a key role in the proof of Theorem 1 and to do so, we make use of the following result, which is proven as a part of the proof of Lemma 8 ([20, Eq. (27)].

**Lemma 12** (see [20]): Under Assumption 1, for all $v \in \mathbb{R}^m$, we have

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

**Lemma 13:** For the dynamics (2), under Assumption 1 and 2, for all $v \in \mathbb{R}^m$, we have

$$E[\|\bar{x}(t + 1) - v, t\|^2 \|F(t)\|] \leq \|\bar{x}(t) - v\|^2 + \alpha^2(t) \frac{L_i^2}{n^2} + \sum_{i=1}^{n} \|x_i(t) - \bar{x}(t)\|^2$$

$$- \frac{2\alpha(t)}{n}(F(\bar{x}(t)) - F(v)) + \frac{4\alpha(t)}{n} \sum_{i=1}^{n} L_i \|x_i(t) - \bar{x}(t)\|. $$
Proof: Multiplying $\frac{1}{n}e^T$ from left to both sides of (2), we have
\[
\bar{x}(t+1) = \bar{W}(t+1)x(t) - \alpha(t)g(t)
\]
\[
= \bar{x}(t) - \alpha(t)g(t) + \bar{W}(t+1)x(t) - \bar{x}(t)
\]
where $\bar{W}(t) \triangleq \frac{1}{n}E[W(t)]$. Therefore, we can write
\[
\|\bar{x}(t+1) - v\|^2
\]
\[
= \|\bar{x}(t) - v - \alpha(t)g(t) + \bar{W}(t+1)x(t) - \bar{x}(t)\|^2
\]
\[
= \|\bar{x}(t) - v\|^2 + \|\alpha(t)g(t)\|^2 + \|\bar{W}(t+1)x(t) - \bar{x}(t)\|^2
\]
\[
- 2\langle \alpha(t)g(t), \bar{W}(t+1)x(t) - \bar{x}(t) \rangle
\]+ 2\langle \bar{x}(t) - v, \bar{W}(t+1)x(t) - \bar{x}(t) \rangle.
\]
Taking conditional expectation of both sides of the above equality given $\mathcal{F}(t)$, we have
\[
E[\|\bar{x}(t+1) - v\|^2|\mathcal{F}(t)] = E[\|\bar{x}(t) - v\|^2|\mathcal{F}(t)]
\]
\[
+ E [\|\bar{W}(t+1)x(t) - \bar{x}(t)\|^2|\mathcal{F}(t)]
\]
\[
- 2\langle \alpha(t)g(t), E [\bar{W}(t+1)x(t) - \bar{x}(t)|\mathcal{F}(t)] \rangle
\]+ 2\langle \bar{x}(t) - v, E [\bar{W}(t+1)x(t) - \bar{x}(t)|\mathcal{F}(t)] \rangle.
\]
The last equality follows from the assumption that, $W(t+1)$ is doubly stochastic in-expectation and hence
\[
E[\bar{W}(t+1)|\mathcal{F}(t)] = \frac{1}{n}e^T
\]
which implies
\[
\langle g(t), E [\bar{W}(t+1)x(t) - \bar{x}(t)|\mathcal{F}(t)] \rangle = \langle \alpha(t)g(t), \bar{x}(t) - v \rangle
\]
and
\[
E [\bar{W}(t+1)x(t) - \bar{x}(t)|\mathcal{F}(t)] = 0.
\]
Note that $\bar{W}(t+1)$ is a stochastic vector (almost surely); therefore, due to the convexity of norm-square $\| \cdot \|^2$, we get
\[
\|\bar{W}(t+1)x(t) - \bar{x}(t)\|^2 \leq \sum_{i=1}^{n} \bar{W}_i(t+1)\|x_i(t) - \bar{x}(t)\|^2
\]
\[
\leq \sum_{i=1}^{n} \|x_i(t) - \bar{x}(t)\|^2
\]
as $\bar{W}_i(t+1) \leq 1$ for all $i \in [n]$. Therefore
\[
E[\|\bar{x}(t+1) - v\|^2|\mathcal{F}(t)]
\]
\[
= \|\bar{x}(t) - v\|^2 + \|\alpha(t)g(t)\|^2
\]
\[
+ E [\|\bar{W}(t+1)x(t) - \bar{x}(t)\|^2|\mathcal{F}(t)]
\]
\[
- 2\langle \alpha(t)g(t), \bar{x}(t) - v \rangle
\]
\[
\leq \|\bar{x}(t) - v\|^2 + \|\alpha(t)g(t)\|^2 + \sum_{i=1}^{n} \|x_i(t) - \bar{x}(t)\|^2
\]
\[
- 2\langle \alpha(t)g(t), \bar{x}(t) - v \rangle.
\]
Finally, Lemma 12 and the fact that
\[
\|g(t)\|^2 = \frac{1}{n^2} \left( \sum_{i=1}^{n} g_i(t) \right)^2 \leq \frac{1}{n^2} \left( \sum_{i=1}^{n} \|g_i(t)\|^2 \right)^2 \leq \frac{L^2}{n^2}
\]
complete the proof.

To prove Theorem 1, we make use of the following general result, which is proved in Appendix.

**Lemma 14:** Let $F : \mathbb{R}^m \rightarrow \mathbb{R}$ be a continuous function with a nonempty set of minimizer $Z \triangleq \arg \min_{z \in \mathbb{R}^m} F(z)$ and $\{\alpha(t)\}$ be a nonnegative scalar sequence with $\sum_{t=0}^{\infty} \alpha(t) = \infty$. Suppose that for a random vector sequence $\{y(t)\}$ in $\mathbb{R}^m$, for every $z \in Z$, we almost surely have
\[
\sum_{t=0}^{\infty} \alpha(t) F(y(t) - F(z)) < \infty.
\]
Then, $\lim_{t \rightarrow \infty} y(t) = z^*$ exists almost surely and $z^* \in Z$ almost surely.

Finally, we are ready to prove the main result.

**Proof of Theorem 1:** In order to utilize Robbins–Siegmund Theorem [28] and Lemma 13, for all $t \geq 0$, let Lyapunov function $V(t) \triangleq \|x(t) - z\|^2$ where $z \in Z$ and $\alpha(t) = 0$, and consider $b(t)$ and $c(t)$, which are defined in (6). First, note that $a(t), b(t), c(t) \geq 0$ for all $t$. To invoke the Robbins–Siegmund result (7), we need 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
\[
\sum_{t=0}^{\infty} \|x_i(t) - \bar{x}(t)\|^2 < \infty
\]
and
\[
\sum_{t=0}^{\infty} \alpha(t) \|x_i(t) - \bar{x}(t)\| < \infty
\]
almost surely, for all $i \in [n]$. From Lemma 3-(c) and Lemma 9, we have
\[
E \left[ \frac{\|x_i(t) - \bar{x}(t)\|}{\alpha(t)} \right] \leq \frac{E \left[ \sqrt{n}d(x(t)) \right]}{\alpha(t)} \leq \sqrt{n\bar{M}} < \infty
\]
for some $\bar{M} > 0$. Therefore, we have
\[
\lim_{T \rightarrow \infty} E \left[ \sum_{t=0}^{T} \alpha(t) \|x_i(t) - \bar{x}(t)\| \right]
\]
\[
= \lim_{T \rightarrow \infty} E \left[ \sum_{t=0}^{T} \alpha^2(t) \left( \frac{\|x_i(t) - \bar{x}(t)\|}{\alpha(t)} \right) \right]
\]
\[
\leq \lim_{T \rightarrow \infty} \sum_{t=0}^{T} \alpha^2(t) E \left[ \frac{\|x_i(t) - \bar{x}(t)\|}{\alpha(t)} \right]
\]
\[
\leq \sqrt{n\bar{M}} \sum_{t=0}^{\infty} \alpha^2(t) < \infty
\]
which is followed by Assumption 4. Similarly, using Lemma 9, there exists some $\bar{M} > 0$ such that
\[
E \left[ \frac{\|x_i(t) - \bar{x}(t)\|^2}{\alpha^2(t)} \right] \leq \frac{E \left[ n\bar{M}d^2(x(t)) \right]}{\alpha^2(t)} \leq \frac{n\bar{M}}{\alpha^2(t)}
\]
for all \( t \geq 0 \), where the first inequality follows from Lemma 3-(c). Therefore
\[
\lim_{T \to \infty} \mathbb{E} \left[ \sum_{t=0}^{T} \| \mathbf{x}_t(t) - \bar{\mathbf{x}}(t) \|^2 \right] = \lim_{T \to \infty} \mathbb{E} \left[ \sum_{t=0}^{\infty} \alpha^2(t) \frac{\| \mathbf{x}_t(t) - \bar{\mathbf{x}}(t) \|^2}{\alpha^2(t)} \right] \\
\leq n \sum_{t=0}^{\infty} \alpha^2(t) < \infty
\]
which is followed by Assumption 4. Therefore, using Monotone Convergence Theorem (see, e.g., [10, Th. 1.5.5]), we have
\[
\mathbb{E} \left[ \sum_{t=0}^{\infty} \alpha(t) \| \mathbf{x}_t(t) - \bar{\mathbf{x}}(t) \| \right] < \infty, \quad \text{and}
\mathbb{E} \left[ \sum_{t=0}^{\infty} \| \mathbf{x}_t(t) - \bar{\mathbf{x}}(t) \|^2 \right] < \infty
\]
which implies \( \sum_{t=0}^{\infty} \alpha(t) \| \mathbf{x}_t(t) - \bar{\mathbf{x}}(t) \| < \infty \) and \( \sum_{t=0}^{\infty} \| \mathbf{x}_t(t) - \bar{\mathbf{x}}(t) \|^2 < \infty \), almost surely.

Now that we showed that \( c(t) \) is almost surely a summable sequence, Robbins–Siegmund Theorem implies that almost surely
\[
\lim_{t \to \infty} \mathbb{V}(t) = \lim_{t \to \infty} \| \bar{\mathbf{x}}(t) - z \|^2 \text{ exists}
\]
and
\[
\sum_{t=1}^{\infty} \alpha(t) (F(\mathbf{x}_t(t)) - F(z)) < \infty.
\]

Therefore, applying Lemma 14 for \( y(t) = \bar{\mathbf{x}}(t) \), we have almost surely \( \lim_{t \to \infty} \bar{\mathbf{x}}(t) = z^* \) with \( z^* \in \mathcal{Z} \) almost surely.

Finally, according to Assumptions 1 and 4, we have
\[
d(\alpha(t)\mathbf{g}(t)) \leq 2Kt^{-\beta} \max_{i \in [n]} L_i.
\]

Therefore, from Lemma 11, we conclude that \( \lim_{t \to \infty} d(\mathbf{x}(t)) = 0 \) almost surely, and hence, Lemma 3-(c) implies \( \lim_{t \to \infty} \| \bar{\mathbf{x}}(t) - \mathbf{x}_t(t) \| = 0 \) 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 \( \lim_{t \to \infty} \mathbf{x}_t(t) = z^* \) for all \( i \in [n] \) almost surely. 

\section{Conclusion and Future Research}

In this work, we showed that the averaging-based distributed optimization solving algorithm over dependent random networks converges to an optimal random point under the standard conditions on the objective function and network formation that is conditionally \( \tilde{B} \)-connected. To do so, we established a rate of convergence 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.

Further extensions of the current work to nonconvex settings, accelerated algorithms, and distributed online learning algorithms are of interest for future consideration on this topic.

\section*{Appendix}

\textbf{Proof of Lemma 1:} Let us fix \( t \geq 0 \). Throughout the proof, for the simplicity of the notations, we may drop the dependence of several objects on \( t \). Let \( T \) be the set of all collections of edges \( E \) such that the graph \( [n], E \) has a spanning rooted tree, and
\[
\mathcal{M}_B \triangleq \bigcup_{\tau = t + 1}^{(t+1)B} \mathbb{E}[\mathbb{E}[W(\tau) | F(\tau - 1)]].
\]
Let \( A = \{ \omega \mid \mathcal{M}_B \in \mathbb{T} \}. \) Since \( \mathcal{M}_B \) almost surely has a spanning rooted tree, we have \( P(A) = 1 \). Letting \( R(T) = \{ \omega \in A \mid \mathcal{M}_B = T \} \) for \( T \in \mathbb{T} \), we have almost surely \( \sum_{T \in \mathbb{T}} 1_{R(T)} = 1 \), which implies
\[
\mathbb{E} \left[ \sum_{T \in \mathbb{T}} 1_{R(T)} | F(t) \right] = \sum_{T \in \mathbb{T}} \mathbb{E} \left[ 1_{R(T)} | F(t) \right] = 1
\]
almost surely. Let
\[
Q(T) = \left\{ \omega \in A \mid \mathbb{E} \left[ 1_{R(T)} | F(t) \right] \geq \frac{1}{|T|} \right\}
\]
Note that if \( \omega \notin \bigcup_{T \in \mathbb{T}} Q(T) \), then
\[
\sum_{T \in \mathbb{T}} \mathbb{E} \left[ 1_{R(T)} | F(t) \right] (\omega) < 1.
\]
But since (29) holds almost surely, this implies that
\[
P(\bigcup_{T \in \mathbb{T}} Q(T)) = 1.
\]
Let \( M(T) \) be the adjacency matrix of \( T \in \mathbb{T} \). Then, for any \( T \in \mathbb{T} \), we have
\[
\sum_{\tau = t + 1}^{(t+1)B} \mathbb{E}[W(\tau) | F(\tau - 1)] \geq \gamma M(T) 1_{R(T)}.
\]
Taking conditional expectations given \( F(t) \) of both sides leads to
\[
\sum_{\tau = t + 1}^{(t+1)B} \mathbb{E}[W(\tau) | F(\tau - 1)] | F(t)] \\
\geq \mathbb{E}[\gamma M(T) 1_{R(T)} | F(t)] = \gamma M(T) \mathbb{E}[1_{R(T)} | F(t)].
\]
Therefore, from the tower identity for conditional expectation (e.g., [10, Th. 5.1.6]), we have
\[
\sum_{\tau = t + 1}^{(t+1)B} \mathbb{E}[W(\tau) | F(t)] \\
= \sum_{\tau = t + 1}^{(t+1)B} \mathbb{E}[W(\tau) | F(\tau - 1)] | F(t)] \\
\geq \gamma M(T) \mathbb{E}[1_{R(T)} | F(t)] \\
\geq \frac{\gamma M(T)}{|T|} 1_{Q(T)}.
\]
Therefore, for $T \in \mathbb{T}$ and for all $\omega \in Q(T)$, the edges of the graph $T$ is a subset of

$$\mathcal{M}_B \triangleq \bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}_{T^\tau} \left( \mathbb{E}[W(\tau) \mid F(TB)] \right).$$

This is because if $(j, i) \in T$ but $(j, i) \not\in \mathcal{M}_B$, then

$$\sum_{\tau=tB+1}^{(t+1)B} \mathbb{E}[w_{ij}(\tau) \mid F(TB)] < \frac{\gamma}{|T|} 1_{Q(\tau)}$$

which contradicts with (30). Since $T \in \mathbb{T}$, the graph with edges $\mathcal{M}_B$ has a spanning rooted tree for all $\omega \in Q(T)$. Finally, $P(\bigcup_{T \in \mathbb{T}} Q(T)) = 1$ establishes Lemma 1 with $K(n, B) = (B|T|)^{-1}$. \hfill \blacksquare

**Proof of Lemma 3**: For the proof of Part (a), let $y = Ax$, and $x_k^{(i)}$ and $y_i^{(k)}$ be the $k$th coordinate of $x_i$ and $y_i$, respectively. Also, let $x^{(k)} = (x^{(k)}_1, \ldots, x^{(k)}_m)^T$ and $y^{(k)} = (y^{(k)}_1, \ldots, y^{(k)}_n)^T$. Since $d(x) = \max_{i,j \in [n], \ell \in [m]} |x^{(\ell)}_i - x^{(\ell)}_j|$, we have $d(x) = \max_{\ell \in [m]} d(x^{(\ell)})$ and

$$d(Ax) = d(y) = \max_{\ell \in [m]} d(y^{(\ell)}) = \max_{\ell \in [m]} d(Ax^{(\ell)}) = \max_{\ell \in [m]} \text{diam}(A) d(x^{(\ell)}) = \text{diam}(A) \max_{\ell \in [m]} d(x^{(\ell)})$$

where (*) follows from Lemma 2-(c).

For Part (b), we note that for a fix $i, j \in [n]$, $g_{ij}(x) \triangleq \|x_i - x_j\|_\infty$ is a convex function of $x$. Therefore, $d(x) = \max_{i,j \in [n]} g_{ij}(x)$ is a convex function. It also follows from the definition that $d(\cdot)$ has the scaling property $d(\alpha x) = |\alpha| d(x)$ for all $\alpha \in \mathbb{R}$. Therefore, for any $x, y \in \mathbb{R}^m$, we get $d(x + y) = d(\frac{x + y}{2}) \leq \frac{1}{2}(d(x) + d(y))$.

For Part (c), due to the convexity of $\|\cdot\|_\infty$, we have $\|x_i - \sum_{j=1}^n \pi_j x_j\|_\infty \leq \sum_{j=1}^n \pi_j \|x_i - x_j\|_\infty \leq \sum_{j=1}^n \pi_j \sqrt{d(x)} = \sqrt{d(x)}$.

**Proof of Lemma 5**: 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

$$\mathbb{E} \left[ \prod_{k=K_1}^{K_2+1} Y(k) \right] = \mathbb{E} \left[ \prod_{k=K_1}^{K_2+1} \mathbb{E}[Y(k) \mid F(K_2)] \right]$$

$$= \mathbb{E} \left[ \mathbb{E}[Y(K_2 + 1) \mid F(K_2)] \prod_{k=K_1}^{K_2} Y(k) \right]$$

$$\leq \mathbb{E} \left[ a(K_2 + 1) \prod_{k=K_1}^{K_2} Y(k) \right] \leq \prod_{k=K_1}^{K_2+1} a(k).$$

**Proof of Lemma 8**: Consider $\bar{t} \geq 0$ such that $\bar{\theta} \triangleq \sup_{t \geq \bar{t}} \beta(t) < 1$, and let $D(t) \triangleq \sum_{s=t}^{t-1} \beta(s) \theta^{t-s}$. Dividing both sides by $\beta(t) > 0$, for $t > \bar{t}$, we have

$$\frac{D(t)}{\beta(t)} - \beta(t)^{-\bar{\theta}} \geq \frac{t-1}{t} \beta(t)^{-\bar{\theta}} - \beta(t)^{-\bar{\theta}}.$$
implying \( \sum_{t=1}^{\infty} P(A_\theta(t)) < \infty \) as \( \exp(-t^\beta) \leq \frac{1}{t^\alpha} \) for sufficiently large \( M \) (depending on \( \beta \)). Therefore, the Borel–Cantelli Theorem (see, e.g., [10, Th. 2.3.1]) implies that \( P(\{A_\theta(t) \text{ i.o.}\}) = 0 \) for all \( \theta > 0 \).

For \( \theta > 0 \), let the sequences of events

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

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

\[
E[\lambda 1_{a(\tau) \geq \lambda} | F(\tau - 1)] \leq E[a(\tau) | F(\tau - 1)] \leq \hat{\lambda} < \lambda (1 - \rho)
\]

and hence,

\[
E[1_{a(\tau) \geq \lambda}] | F(\tau - 1)] < 1 - \rho. \tag{32}
\]

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

\[
S(t + \sigma(t)) - S(t) = \sigma(t) - \sum_{\tau=t+1}^{t+\sigma(t)} E[1_{a(\tau) \geq \lambda}] | F(\tau - 1)] \leq \sigma(t) - \sigma(t)(1 - \rho) = \sigma(t)\rho
\]

which follows from (32). Therefore, we have \( B_\theta(t) \subset A_\theta(t) \), and hence, \( P(\{B_\theta(t) \text{ i.o.}\}) = 0 \) for all \( \theta > 0 \).

Finally, by contradiction, we show that \( \lim_{t \to \infty} (t+1-s)\beta - t \beta \neq 0 \). Since, if \( \lim_{t \to \infty} (t+1-s)\beta - t \beta \neq 0 \) almost surely, then \( \limsup_{t \to \infty} (t+1-s)\beta - t \beta \neq 0 \) almost surely, and hence, \( P(\limsup_{t \to \infty} (t+1-s)\beta - t \beta > \epsilon) > 0 \) for some \( \epsilon > 0 \). Therefore, \( P(\{\lim_{t \to \infty} (t+1-s)\beta - t \beta > 0\}) = 0 \), which is a contradiction.

Proof of Lemma 10: Let \( t_\beta \triangleq \inf\{t > t_{s+1} - t : a(t) \leq \lambda\} \) and \( t_0 = 0 \) for some \( \hat{\lambda} < \lambda < 1 \), and \( c(\theta) \triangleq \sum_{\tau=t+s+1}^{t_\beta} b(\tau) \). Also, define

\[
A \triangleq \left\{ \omega : \lim_{s \to \infty} \frac{t_{s+1} - t_s}{t_0} = 0 \right\}.
\]

Note that Lemma 15 implies \( P(A) = 1 \). On the other hand, using (24), we have

\[
D(t_{s+1}) \leq D(t_s) \prod_{\ell=t_s+1}^{t_{s+1}} a(\ell) + \sum_{\tau=t_s}^{t_{s+1}} b(\tau) \prod_{\ell=\tau+2}^{t_{s+1}} a(\ell)
\]

\[
\leq D(t_s) \lambda + c(s),
\]

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

\[
\lim_{s \to \infty} \frac{t_{s+1}}{t_s} = \lim_{s \to \infty} \frac{t_{s+1} - t_s}{t_s} = 1.
\]

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, \( R(t_s) \leq \hat{\lambda}_s^{\beta-\epsilon} R(t_s) + \sum_{s-1}^{s} c(t) t_s^{\beta} \hat{\lambda}_s^{\beta-\epsilon} - 1 \).

Taking the limits of the both sides, we have

\[
\limsup_{s \to \infty} R(t_s) \leq \limsup_{s \to \infty} \hat{\lambda}_s^{\beta-\epsilon} R(t_s) + \sum_{s-1}^{s} c(t) t_s^{\beta} \hat{\lambda}_s^{\beta-\epsilon} - 1
\]

which is implied by [25, Lemma 3.1-(a)]. For \( \omega \in A \), we have

\[
\lim_{s \to \infty} c(s) t_s^{\beta} \leq \lim_{s \to \infty} \frac{K(t_{s+1} - t_s)}{t_s^{\beta}} \leq \frac{K(t_{s+1} - t_s)}{t_{s+1}^{\beta}} = \frac{K(t_{s+1} - t_s)}{t_{s+1}^{\beta}} = 0
\]

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

Proof of Lemma 14: For \( z \in Z \), let us define

\[
\Omega_z \triangleq \left\{ \omega : \lim_{t \to \infty} \|y(t, \omega) - z\| \text{ exists} \right\}
\]

where \( F^* = \min_{z \in \mathbb{R}} F(z) \). From the assumption in the lemma, we know that \( P(\Omega_z) = 1 \). Now, let \( \Omega_d \subset Z \) be a countable dense subset of \( Z \) and let

\[
\Omega_d \triangleq \bigcap_{z \in \Omega_d} \Omega_z.
\]

Since \( \Omega_d \) is a countable set, we have \( P(\Omega_d) = 1 \) and for \( \omega \in \Omega_d \), since \( \sum_{t=1}^{\infty} \alpha(t) (F(y(t, \omega)) - F^*) < \infty \) and \( \alpha(t) \) is not summable, we have \( \liminf_{t \to \infty} F(y(t, \omega)) = F^* \). This fact and the fact that \( F(\cdot) \) is a continuous function implies that for all \( \omega \in \Omega_d \), we have \( \liminf_{t \to \infty} \|y(t, \omega) - z(\omega)\| = 0 \) for some \( z(\omega) \in Z \). To show this, let \( \{y(t_k)\} \) be a subsequence that \( \lim_{k \to \infty} F(y(t_k, \omega)) = F^* \) (such a subsequence depends on the sample path \( \omega \)). Since \( \omega \in \Omega_d \) and

\[
\lim_{t \to \infty} \|y(t, \omega) - z\| \text{ exists}
\]

for some \( z \in \Omega_d \), we conclude that \( \{y(t, \omega)\} \) is a bounded sequence. Therefore, \( \{y(t_k, \omega)\} \) is also bounded and it has an accumulation point \( z^* \in \mathbb{R}^m \) and hence, there is a subsequence \( \{y(t_k, \omega)\}_{k \geq 0} \) of \( \{y(t_k, \omega)\}_{k \geq 0} \) with \( \lim_{t \to \infty} y(t_k, \omega) = z^* \). As a result of continuity of \( F(\cdot) \), we have

\[
\lim_{t \to \infty} F(y(t_k, \omega)) = F(z^*) = F^*
\]
and hence, \( z^* \in Z \). Note that the point \( z^* = z^*(\omega) \) depends on the sample path \( \omega \).

Since \( Z_d \subseteq Z \) is dense, there is a sequence \( \{q^*(s, \omega)\}_{s \geq 0} \) in \( Z_d \) such that \( \lim_{s \to \infty} \|q^*(s, \omega) - z^*(\omega)\| = 0 \). Note that since \( \omega \in \Omega_d \), \( \lim_{t \to \infty} \|y(t, \omega) - q^*(s, \omega)\| \) exists for all \( s \geq 0 \) and we have

\[
\lim_{t \to \infty} \|y(t, \omega) - q^*(s, \omega)\| = \lim_{t \to \infty} \|y(t, \omega) - z^*(\omega) + z^*(\omega) - q^*(s, \omega)\|
\leq \lim_{t \to \infty} \|y(t, \omega) - z^*(\omega)\| + \|z^*(\omega) - q^*(s, \omega)\|
\leq \|z^*(\omega) - q^*(s, \omega)\|.
\]

Therefore, we have

\[
\lim_{s \to \infty} \lim_{t \to \infty} \|y(t, \omega) - q^*(s, \omega)\| = 0. \tag{35}
\]

On the other hand, we have

\[
\limsup_{t \to \infty} \|y(t, \omega) - z^*(\omega)\| = \limsup_{t \to \infty} \|y(t, \omega) - q^*(s, \omega) + q^*(s, \omega) - z^*(\omega)\|
\leq \limsup_{t \to \infty} \|y(t, \omega) - q^*(s, \omega)\| + \|q^*(s, \omega) - z^*(\omega)\|
= \left( \limsup_{t \to \infty} \|y(t, \omega) - q^*(s, \omega)\| \right) + \|q^*(s, \omega) - z^*(\omega)\|.
\]

Therefore

\[
\limsup_{t \to \infty} \|y(t, \omega) - z^*(\omega)\| = \lim_{s \to \infty} \limsup_{t \to \infty} \|y(t, \omega) - z^*(\omega)\|
\leq \lim_{s \to \infty} \|y(t, \omega) - q^*(s, \omega)\| + \lim_{s \to \infty} \|q^*(s, \omega) - z^*(\omega)\|
= 0 \tag{36}
\]

where the last equality follows by combining (35) and \( \lim_{s \to \infty} \|q^*(s, \omega) - z^*(\omega)\| = 0 \). Note that (36) implies that almost surely (i.e., for all \( \omega \in \Omega_d \), \( \lim_{t \to \infty} y(t) = z^*(\omega) \) exists and it belongs to \( Z \).

**ACKNOWLEDGMENT**

The authors would like to thank the reviewers and the associate editor’s comments for improving this work.

**REFERENCES**

[1] N. Alon and J. H. Spencer, *The Probabilistic Method*. Hoboken, NJ, USA: Wiley, 2016.

[2] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, “Broadcast gossip algorithms for consensus,” *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2748–2761, Jul. 2009.

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

[4] F. S. Cattellini and A. H. Sayed, “Diffusion LMS strategies for distributed estimation,” *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, Mar. 2010.

[5] B. Charbon-Bost, “Orientation and connectivity based criteria for asymptotic consistency,” 2013, arXiv:1303.2043.

[6] S. Chatterjee and E. Seneta, “Towards consensus: Some convergence theorems on repeated averaging,” *J. Appl. Probability*, vol. 14, no. 1, pp. 89–97, 1977.

[7] A. Cherukuri and J. Cortés, “Distributed generator coordination for initialization and anytime optimization in economic dispatch,” *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 3, pp. 226–237, Sep. 2015.

[8] R. L. Dobrushin, “Central limit theorem for nonstationary Markov chains,” *II. Theory Probability Appl.*, vol. 1, no. 4, pp. 329–383, 1956.

[9] I. Lobel and A. Ozdaglar, “Distributed subgradient methods for optimization over random networks,” in *Proc. IEEE 51st Conf. Decis. Control*, 2012, pp. 3688–3693.

[10] D. P. Bertsekas, *Convex Optimization*, Athena Scientific, 2009.

[11] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, 1st ed. Boston, MA, USA: MIT Press, 1997.

[12] R. Durrett, *Probability: Theory and Examples*. Cambridge, U.K.: Cambridge Univ. Press, 2010.

[13] J. Hajnal and M. Bartlett, “Weak ergodicity in non-homogeneous Markov chains,” *Math. Proc. Cambridge Philos. Soc.*, vol. 54, no. 2, pp. 233–246, 1958.

[14] S. Kar and J. M. Moura, “Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and trade-offs,” *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 4, pp. 674–690, Aug. 2011.

[15] S. Lee and A. Nedić, “Asynchronous broadcast-based random projection algorithms over networks,” *IEEE Trans. Autom. Control*, vol. 61, no. 4, pp. 953–968, Apr. 2016.

[16] Y. Nesterov, “Linear methods for optimization over directed graphs,” *Proc. IEEE*, vol. 98, no. 1, pp. 140–143, Jan. 2010.

[17] I. Lobel, A. Ozdaglar, and D. Feijer, “Distributed distributed optimization with state-dependent communication,” *Math. Program.*, vol. 129, no. 2, pp. 255–284, 2011.

[18] V. S. Mai and E. H. Abed, “Distributed optimization over weighted directed graphs using row stochastic matrices,” in *Proc. IEEE Conf. Decis. Control*, 2016, pp. 7165–7170.

[19] A. Nedić, “Asynchronous broadcast-based convex optimization over a network,” *IEEE Trans. Autom. Control*, vol. 56, no. 6, pp. 1291–1306, Jun. 2011.

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

[21] A. Nedić and A. Ozdaglar, “Distributed subgradient methods for multi-agent optimization,” *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[22] A. Nedić, A. Ozdaglar, and P. A. Parrilo, “Constrained consensus and optimization in multi-agent networks,” *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.

[23] M. Rabbat and R. Nowak, “Distributed optimization in sensor networks,” in *Proc. 3rd Int. Symp. Inf. Process. Sensor Netw.*, 2004, pp. 20–27.

[24] S. S. Ram, A. Nedić, and V. V. V. Veeravalli, “Distributed stochastic subgradient projection algorithms for convex optimization,” *J. Optim. Theory Appl.*, vol. 147, no. 3, pp. 516–545, 2010.

[25] P. Rezaeinia and B. Gharesifard, “Rate of convergence for distributed optimization with uncertain communications,” in *Proc. IEEE Conf. Decis. Control*, 2020, arXiv:2003.10342.

[26] P. Rezaeinia, B. Gharesifard, T. Linder, and B. Touri, “Push-sum on random graphs: Almost sure convergence and convergence rate,” *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 1295–1302, Mar. 2020.

[27] H. Robbins and D. Siegmund, “A convergence theorem for non-negative almost supermartingales and some applications,” in *Optimizing Methods and Software*, J. S. Rustagi, Ed. New York, NY, USA: Academic, 1971, pp. 233–257.

[28] J. Shen, “A geometric approach to ergodic non-homogeneous Markov chains,” in *Proc. Lecture Notes Pure Appl. Math.*, 2000, pp. 341–366.

[29] S. S. Stankovic, M. S. Stankovic, and D. M. Stipanovic, “Decentralized parameter estimation by consensus based stochastic approximation,” *IEEE Trans. Autom. Control*, vol. 56, no. 3, pp. 531–543, Mar. 2011.

[30] A. Talboub-Salehi and A. Jadbabaie, “A necessary and sufficient condition for consensus over random networks,” *IEEE Trans. Autom. Control*, vol. 53, no. 3, pp. 791–795, Apr. 2008.
[32] B. Touri and C. Langbort, “On endogenous random consensus and averaging dynamics,” IEEE Trans. Control Netw. Syst., vol. 1, no. 3, pp. 241–248, Sep. 2014.

[33] B. Touri and A. Nedić, “Product of random stochastic matrices,” IEEE Trans. Autom. Control, vol. 59, no. 2, pp. 437–448, Feb. 2014.

[34] K. Tsianos, “The role of the network in distributed optimization algorithms: Convergence rates, scalability, communication/computation trade-offs and communication delays,” Ph.D. dissertation, Dept. Elect. Comput. Eng., McGill Univ. Libraries, 2013.

[35] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, “Consensus-based distributed optimization: Practical issues and applications in large-scale machine learning,” in Proc. 50th Annu. Allerton Conf. Commun., Control, Comput., 2012, pp. 1543–1550.

[36] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, “Push-sum distributed dual averaging for convex optimization,” in Proc. IEEE 51st Conf. Decis. Control, 2012, pp. 5453–5458.

[37] C. Xi, V. S. Mai, R. Xin, E. H. Abed, and U. A. Khan, “Linear convergence in optimization over directed graphs with row-stochastic matrices,” IEEE Trans. Autom. Control, vol. 63, no. 10, pp. 3558–3565, Oct. 2018.

Behrouz Touri (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Isfahan University of Technology, Isfahan, Iran, in 2006, the M.Sc. degree in communications, systems, and electronics from Jacobs University, Bremen, Germany, in 2008, and the Ph.D. degree in industrial engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2011.

He is currently an Associate Professor of the Electrical and Computer Engineering with the University of California San Diego, San Diego, CA, USA. Between 2011 and 2014, he was a Postdoctoral Researcher with the ECE departments of the University of Illinois and Georgia Institute of Technology. Prior to joining the University of California San Diego, he was an Assistant Professor of Electrical Engineering with the University of Colorado Boulder from 2014 to 2017. His research interests include applied probability theory, distributed optimization, control and estimation, population dynamics, and evolutionary game theory.

Dr. Touri is the recipient of American Control Council’s Donald P. Eckman Award in 2018.

Adel Aghajan received the B.Sc. and M.Sc. degrees in electrical engineering from the Isfahan University of Technology, Isfahan, Iran, in 2009 and 2012, respectively, and the Ph.D. degree in electrical engineering from the University of California San Diego, San Diego, CA, USA, in 2021.

He is currently a Postdoctoral Scholar with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, Santa Barbara, CA. His research interests include game theory, distributed computation and optimization, random dynamics, and information theory.