Distributed Optimization With Uncertain Communications

Pouya Rezaeinia, Bahman Gharesifard, and Tamás Linder

***Abstract*—In this article, we consider a distributed optimization problem for the sum of convex functions where the underlying communication network connecting nodes at each time epoch is drawn at random from a collection of directed graphs. We propose a modified version of the subgradient-push algorithm that provably almost surely converges to an optimizer on any such sequence of random directed graphs. We also prove that the convergence rate of our proposed algorithm is upper bounded as $O(\frac{1}{\sqrt{t}})$, where t is the time horizon.**

***Index Terms*—Distributed control, distributed optimization, ergodic chains, random networks.**

I. Introduction

Distributed optimization of a sum of convex functions involves solving the problem where a network of nodes $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$, each with a private local convex function $f_i : \mathbb{R}^d \to \mathbb{R}$, aims to solve the problem

$$\text{minimize } F(\mathbf{z}) := \sum_{i=1}^{n} f_i(\mathbf{z}), \quad \mathbf{z} \in \mathbb{R}^d \tag{1}$$

in a *distributed manner*, exchanging only limited information on their estimate of the optimizer. The significance of this problem lies in its wide range of applications, including sensor localization [3], statistical learning [4], and Big Data [5].

Distributed optimization is a well-established subject with extensive literature, as noted in [6]. This work focuses on communication network conditions, especially with randomly drawn communication graphs, to ensure convergence to a solution of (1). Practical implementations of distributed optimization algorithms aim for guaranteed performance despite communication deficiencies or uncertainties, as emphasized in [7]. Thus, our literature review in Section II-A centers on communication networks, and so it may overlook some important results in other areas.

Received 28 May 2024; revised 11 October 2024; accepted 4 November 2024. Date of publication 8 November 2024; date of current version 31 March 2025. The work of Tamás Linder was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada. An early version of this paper's first part was presented at the 2018 7th IFAC Workshop on Distributed Estimation and Control in Networked Systems, (NECSYS) [DOI: 10.1016/j.ifacol.2018.12.067]. The convergence rate results were initially announced on arXiv and have been significantly improved in this manuscript. Recommended by Associate Editor Z. Shu. *(Corresponding author: Pouya Rezaeinia.)*

Pouya Rezaeinia is an independent researcher, formerly with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: rezaeinia.pouya@queensu.ca).

Bahman Gharesifard is with the Department of Electrical and Computer Engineering, University of California, Los Angeles, Los Angeles, CA 90095 USA (e-mail: gharesifard@ucla.edu).

Tamás Linder is with the Department of Electrical and Computer Engineering... Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: tamas.linder@queensu.ca).

Digital Object Identifier 10.1109/TAC.2024.3495456

A. Mathematical Preliminaries

Let $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{Z}_{\geq 0}$ denote the sets of real numbers, nonnegative real numbers, and integers, respectively. For a set $\mathbb{A}$, $S \subset \mathbb{A}$ indicates that S is a proper subset of $\mathbb{A}$; the empty set and $\mathbb{A}$ are trivial subsets of $\mathbb{A}$. The complement of S is S^c. The cardinality of a finite set S is $|S|$. Vectors in $\mathbb{R}^n$ are column vectors, where n is a positive integer. The standard Euclidean norm and the 1-norm on $\mathbb{R}^n$ are denoted by $\|\cdot\|$ and $\|\cdot\|_1$, respectively. The transpose of a matrix A and a vector v are denoted by A' and v'. The set of $n \times n$ nonnegative real-valued matrices is denoted by $\mathbb{R}_{\geq 0}^{n\times n}$. A matrix $A \in \mathbb{R}_{\geq 0}^{n\times n}$ is column-stochastic if each columns sums to 1, it is row-stochastic if each row sums to 1, and it is doubly stochastic if both of these conditions are satisfied. For $A \in \mathbb{R}_{\geq 0}^{n\times n}$ and any nontrivial $S \subset [n] := \{1, \ldots, n\}$, define $A_{SS^c} := \sum_{i\in S, j\in S^c} A_{ij}$.

1) Graph Theory: A (weighted) *directed graph* $G := (\mathcal{V}, \mathcal{E}, W)$ consists of a node set $\mathcal{V} := \{v_1, v_2, \ldots, v_n\}$, an edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, and a weighted *adjacency matrix* $W \in \mathbb{R}_{\geq 0}^{n\times n}$, with $W_{ji} > 0$ if and only if $(v_i, v_j) \in \mathcal{E}$, indicating that v_i is connected to v_j. Similarly, given a matrix $W \in \mathbb{R}_{\geq 0}^{n\times n}$, one can associate a directed graph $G = (\mathcal{V}, \mathcal{E})$, where $(v_i, v_j) \in \mathcal{E}$ if and only if $W_{ji} > 0$, making W the corresponding weighted adjacency matrix. The in-neighbors and the out-neighbors of v_i are the sets of nodes $N_i^{\text{in}} = \{j \in [n] : W_{ij} > 0\}$ and $N_i^{\text{out}} = \{j \in [n] : W_{ji} > 0\}$, respectively. The out-degree of v_i is $d_i = |N_i^{\text{out}}|$; for undirected graphs we drop the superscripts "in" and "out." In the directed graph $G = (\mathcal{V}, \mathcal{E}, W)$, a path is the sequence of distinct nodes $v_{i_1}, \ldots, v_{i_k}$ for some $k \in [n]$ such that $(v_{i_j}, v_{i_{j+1}}) \in \mathcal{E}$ for all $j \in [k-1]$. A directed graph is *strongly connected* if there is a path between any pair of nodes. If $G = (\mathcal{V}, \mathcal{E}, W)$ is strongly connected, W is called irreducible. For graphs $G_1 = (\mathcal{V}, \mathcal{E}_1)$ and $G_2 = (\mathcal{V}, \mathcal{E}_2)$ on the node set $\mathcal{V}$, $G = G_1 \cup G_2$ is the graph on $\mathcal{V}$ with the edge set $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2$.

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

II. Algorithm

Consensus-based optimization has been extensively explored in literature, often under the assumption that the underlying network is either doubly stochastic or weight-balanced, as discussed in various studies [8], [9], [10], [11], [12], [13]. This article focuses on scenarios where the network is not weight-balanced. In such cases, the subgradient-push (SP) algorithm is employed to achieve average consensus [11]. In particular, this method combines the push-sum protocol [14] with subgradient flow, for more recent literature see [15], ensuring convergence to a solution of (1). We introduce this dynamic as a foundation for subsequent discussions.

A. Subgradient-Push Algorithm

The SP algorithm involves each node v_i maintaining and updating two vector variables, $\mathbf{x}_i(t)$ and $\mathbf{w}_i(t) \in \mathbb{R}^d$, and a scalar variable $y_i(t) \in \mathbb{R}$. Initially, $\mathbf{x}_i(0)$ is set to an estimate of the optimal solution for node v_i, and $y_i(0)$ is set to 1. Nodes send $\mathbf{x}_i(t)$ and $y_i(t)$ to out-neighbors in a (deterministic) directed graph of the available communication channels $\bar{G}(t) = (\mathcal{V}, \bar{\mathcal{E}}(t))$, assumed to contain self-loops. For each $i \in [n]$, $\bar{N}_i^{\text{in}}(t)$ is the set of in-neighbors of v_i and $\bar{d}_i(t)$ is the out-degree of v_i in $\bar{G}(t)$. Each node updates its variables at time $t+1$ as follows:

$$\begin{cases} \mathbf{w}_i(t+1) = \sum_{j \in \bar{N}_i^{\text{in}}(t)} \frac{\mathbf{x}_j(t)}{\bar{d}_j(t)} \\ y_i(t+1) = \sum_{j \in \bar{N}_i^{\text{in}}(t)} \frac{y_j(t)}{\bar{d}_j(t)} \\ \mathbf{z}_i(t+1) = \frac{\mathbf{w}_i(t+1)}{y_i(t+1)} \\ \mathbf{x}_i(t+1) = \mathbf{w}_i(t+1) - \alpha(t+1)\mathbf{g}_i(t+1) \end{cases}$$

where $\mathbf{g}_i(t+1)$ is a subgradient of the convex function f_i at $\mathbf{z}_i(t+1)$ and $\alpha(t) = \frac{1}{t^\gamma}$ for some $\frac{1}{2} < \gamma < 1$. This choice of $\alpha(t)$ ensures $\sum_{t=1}^{\infty} \alpha(t) = \infty$, and $\sum_{t=1}^{\infty} \alpha^2(t) < \infty$. At each time t, $\mathbf{z}_i(t)$ is node v_i's estimate of a minimizer of $F(\mathbf{z})$. The functions f_i are assumed to be Lipschitz continuous, i.e., for all $i \in [n]$ there exists L_i such that $\|\mathbf{g}_i\| \le L_i$. For our future analysis we define $L = \sum_{i=1}^n L_i$. A key assumption is that each node is aware of its out-degree, deemed necessary for algorithmic success according to [16].

The SP algorithm has been proven to converge in deterministic, time-varying settings with strong uniform connectivity [11]. Although consensus-based optimization has been studied in random undirected settings [17], extending these results to random directed graphs presents significant challenges. Recent studies have concentrated on optimization in environments characterized by noisy and imperfectly connected time-varying networks [7], [18].

In [19], we established an ergodicity criterion for column-stochastic matrices in the push-sum protocol, showing that a broad range of time-varying random directed graphs meet these criteria. We applied these findings to random graphs with Bernoulli-like edge probabilities, deriving convergence rates for the push-sum algorithm under milder ergodicity and infinite flow conditions (see [20] for details). Building on this foundation, this article examines minimization problem (1) over a sequence of random directed graphs.

B. Modified Subgradient-Push Algorithm

Here, we present our algorithm designed to extend deterministic convergence rates to scenarios with random graphs. The difficulty lies in the algorithm's dependence on the ratio of correlated random variables and the need to bound the denominator away from zero.

In the modified SP (MSP) algorithm, each node $v_i \in \mathcal{V}$ sends its values to its out-neighbors in $\bar{G}(t)$ only if $y_i(t) \ge \frac{1}{n^{2n}} =: \delta$, making v_i an *active* node. At time t, node v_i receives information from its active in-neighbors $N_i^{\text{in}}(t) = \bar{N}_i^{\text{in}}(t) \backslash \{v_j \in \bar{N}_i^{\text{in}}(t) \mid y_j(t) < \delta\}$. This forms an effective communication network graph $G(t) = (\mathcal{V}, \mathcal{E}(t), W(t))$ at time t with nodes $\mathcal{V}$ and edges $\mathcal{E}(t) \subseteq \bar{\mathcal{E}}(t)$. The MSP algorithm is

$$\begin{cases} \mathbf{w}_i(t+1) = \sum_{j \in N_i^{\text{in}}(t)} \frac{\mathbf{x}_j(t)}{d_j(t)} \\ y_i(t+1) = \sum_{j \in N_i^{\text{in}}(t)} \frac{y_j(t)}{d_j(t)} \\ \mathbf{z}_i(t+1) = \frac{\mathbf{w}_i(t+1)}{y_i(t+1)} \\ \mathbf{x}_i(t+1) = \mathbf{w}_i(t+1) - \alpha(t+1)\mathbf{g}_i(t+1) \end{cases} \tag{2}$$

where $\bar{N}_i^{\text{in}}(t)$ is replaced with $N_i^{\text{in}}(t)$ and $d_i(t)$ is the out-degree of v_i in $G(t)$. In what follows, whenever we write $G(t)$, or $N_i^{\text{in}}(t)$, we refer to the subgraph of $\bar{G}(t)$ resulting from this modification.

The MSP algorithm addresses information imbalances by using a threshold mechanism that allows nodes to selectively participate based on their connectivity and "information quality." This selective participation differentiates an effective communication network graph from conventional network graphs that do not filter connections based on these criteria.

III. Random Setting and Main Result

As described in Section II-B, $\bar{G}(t) = (\mathcal{V}, \bar{\mathcal{E}}(t))$ denotes the graph of the available communication channels at time t and $G(t) = (\mathcal{V}, \mathcal{E}(t), W(t))$ denotes the effective communication network graph, i.e., those channels that are actively contributing to the network's objectives at any given time.

Assumption 1: Let $\mathcal{G} = \{\bar{G}_1, \bar{G}_2, \ldots, \bar{G}_{2^{n^2-n}}\}$ be the set of all possible graphs of available communication channels on $\mathcal{V}$ with self-loops at all nodes. The sequence of communication graphs $\{\bar{G}(t)\}$ satisfies the following.

i) At each time $t \ge 0$, $\bar{G}(t)$ is drawn randomly from $\mathcal{G}$ with distribution $p_b := \mathbb{P}(\bar{G}(t) = \bar{G}_b)$, where $b \in [2^{n^2-n}]$.
ii) $\bigcup_{b: p_b > 0} \bar{G}_b$ is strongly connected.
iii) $\{\bar{G}(t)\}$ is an independent and identically distributed sequence.

These assumptions state that the graphs of available communication channels are drawn independently from the set of all possible graphs on $\mathcal{V}$. Part (ii) imposes a mild connectivity assumption, analogous to deterministic settings in [11]. Note that this assumption only involves the underlying probabilities p_b, $b \in [2^{n^2-n}]$.

We define the following auxiliary quantities used in the main results of this article.

1) $S(0) = 0$; and $S(t) = \sum_{s=1}^t \alpha(s)$ for all $t \ge 1$.
2) $B = 2n - 2$.
3) $\lambda = \left(1 - \frac{1}{n^{\left(\frac{4nB}{p}\right)}}\right)^{\frac{p}{2nB}}$.

Theorem 1: Consider MSP algorithm (2) and suppose that the sequence of available communication channels $\{\bar{G}(t)\}$ satisfies Assumption 1.

i) We have $\lim_{t\to\infty} \mathbf{z}_i(t) = \mathbf{z}^*$, for all $i \in [n]$, almost surely (a.s.), where $\mathbf{z}^*$ is a solution (minimizer) of (1).
ii) Assume that every node i maintains the vector $\tilde{\mathbf{z}}_i(t) \in \mathbb{R}^d$, initialized with an arbitrary $\tilde{\mathbf{z}}_i(0) \in \mathbb{R}^d$, following the update rule:

$$\tilde{\mathbf{z}}_i(t+1) = \frac{\alpha(t+1)\mathbf{z}_i(t+1) + S(t)\tilde{\mathbf{z}}_i(t)}{S(t+1)}, \quad t \ge 0.$$

For all $t \ge 1$, $i \in [n]$, and $\mathbf{z}^* \in Z^*$, we have

$$\mathbb{E}\left[F(\tilde{\mathbf{z}}_i(t+1)) - F(\mathbf{z}^*)\right] \le \Gamma(t)\left[\frac{n\|\bar{\mathbf{x}}(0) - \mathbf{z}^*\|_1}{2} + \left(1 + \frac{1}{2\gamma - 1}\right)\frac{L^2}{2n} + \frac{6\eta_1 L \sum_{j=1}^n \|\mathbf{x}_j(0)\|_1}{\delta(1-\lambda)} + \frac{6\eta_1 d L^2}{\delta(1-\lambda)}\left(1 + \frac{1}{2\gamma - 1}\right)\right]$$

where $\Gamma(t) = \frac{(1-\gamma)}{(t+1)^{1-\gamma} - 1}$.

Theorem 1 part (ii) provides a bound on the convergence rate of the MSP algorithm. On the right-hand side (RHS) of the inequality, all the terms except $\Gamma(t)$ are constant with respect to t. In addition, we have assumed that $\frac{1}{2} < \gamma < 1$. If we allow γ to approach $1/2$ from above, we obtain the upper bound of the order of $O(\frac{1}{\sqrt{t}})$ on the convergence rate of the MSP algorithm.

This bound pertains to the weighted average of iterates in the optimization process $\tilde{\mathbf{z}}_i(\cdot)$, rather than considering the final iterate as the solution. This method, also known as iterate averaging, is commonly employed to assess the convergence rate of stochastic gradient optimizations, effectively accounting for inherent noise and fluctuations that arise from the use of subgradients. For further details, see [21].

IV. Convergence Analysis

The rest of this article is dedicated to proving Theorem 1. Similar to [11], we define a perturbed version of the push-sum algorithm with our modifications. We primarily work with scalar variables. In proving Theorem 1, we apply these results to solve (1) with vector variables.

In the modified perturbed-push (MPP) algorithm each node v_i maintains and updates scalar variables $x_i(t), w_i(t)$, and $y_i(t)$, where $x_i(0)$ are arbitrary scalars and $y_i(0)$ are initialized to 1. Similar to the MSP algorithm, node v_i shares its values only if $y_i(t) \geq \delta$, and $d_i(t)$ and $N_i^{\text{in}}(t)$ are the out-degree and the in-neighbor of $v_i, i \in [n]$, in the effective communication network graph $G(t)$, respectively. To simplify the analysis, we express the algorithm in matrix form. For $t \geq 0$, let $W(t)$ be the column-stochastic matrix associated with the effective communication network graph $G(t)$ with entries

$$W_{ij}(t) = \begin{cases} \frac{1}{d_j(t)}, & \text{if } j \in N_i^{\text{in}}(t) \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The MPP algorithm in matrix form is

$$\begin{cases} w(t+1) = W(t)x(t) \\ y(t+1) = W(t)y(t) \\ z_i(t+1) = \frac{w_i(t+1)}{y_i(t+1)}, \quad \text{for all } i \in [n] \\ x(t+1) = w(t+1) + \epsilon(t+1) \end{cases} \tag{4}$$

where $\epsilon(t) = (\epsilon_1(t), \ldots, \epsilon_n(t))'$ is the vector of perturbations at time t, to be specified later. Here, $w(t) = (w_1(t), \ldots, w_n(t)) \in \mathbb{R}^n$, $z(t) = (z_1(t), \ldots, z_n(t)) \in \mathbb{R}^n$, and we treat each component of the individual node's state variables separately. We assume $\|\epsilon(t)\|_1 \leq \frac{U}{t^\gamma}$, for some $U > 0$, which holds when considering the subgradient term in the MSP algorithm as a perturbation. Throughout this article, the $W(t)$, defined in (3), denote the adjacency matrices of the effective communication graphs.

We examine the convergence properties of the MPP algorithm, focusing initially on the connectivity of the generated matrix sequence. A crucial attribute of these random matrices is the directed infinite flow property.

Definition 1 ([19, Definition 3]): A sequence of random matrices $\{W(t)\}$ has the directed infinite flow property if for any nontrivial $S \subset [n]$, $\sum_{t=0}^{\infty} W_{SS^c}(t) = \infty$, a.s.

Next, we repeat a few definitions from [19]. For a sequence of matrices $W(t)$ of the form (3) that has the directed infinite flow property, let $k_0 = 0$. For any $q \geq 1$, define

$$k_q = \arg\min_{t' > k_{q-1}} \left(\min_{S \subset [n]} \sum_{t=k_{q-1}}^{t'-1} W_{SS^c}(t) > 0 \right). \tag{5}$$

Essentially, k_q is the minimal time instance after k_{q-1}, where there is nonzero communication between any nontrivial subset of $\mathcal{V}$ and its complement. Moreover, let $\ell_0 = 0$ and for $q \geq 1$

$$\ell_q = k_{qn} - k_{(q-1)n}. \tag{6}$$

For $t > s \geq 0$, define $\mathbb{Q}_{t,s} = \{q : s \leq k_{(q-1)n}, k_{qn} \leq t\}$. As shown in [19, Proposition 1], k_q and ℓ_q, which play important roles in our convergence analysis, are sequences of random variables and are well defined for the sequence of matrices $\{W(t)\}$ that has the directed infinite flow property.

The following proposition presents an upper bound on how well the sequences $z_i(t+1)$ estimate the average $\bar{x}(t) := \frac{1}{n}\sum_{i\in[n]} x_i(t)$ for each sample path, when $\{W(t)\}$ has the directed infinite flow property. This will allow us to state our first connectivity result in a random setting.

Proposition 1: Consider MPP algorithm (4) and assume that the sequence $\{W(t)\}$ has the directed infinite flow property, a.s. Then, for all $t \geq s \geq 0$ we have

$$|z_i(t+1) - \bar{x}(t)| \leq \frac{2}{y_i(t+1)} \left(\Lambda_{t,0}\|x(0)\|_1 + \sum_{s=1}^{t} \Lambda_{t,s}\|\epsilon(s)\|_1 \right)$$

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

Proof: For $t \geq s \geq 0$, define the shorthand notation $W(t:s) = W(t)W(t-1)\cdots W(s)$. Since $\{W(t)\}$ has the directed infinite flow property, a.s., by [19, Proposition 3], there exist $\phi(t) \in \mathbb{R}^n$ such that for all $i, j \in [n]$

$$|[W(t:s)]_{ij} - \phi_i(t)| \leq \Lambda_{t,s} \tag{7}$$

where $\phi_i(t)$ is the i'th entry of the vector $\phi(t)$. Define $D(t:s) = W(t:s) - \phi(t)\mathbf{1}_n'$, where $\mathbf{1}_n \in \mathbb{R}^n$ is the vector of all ones. Following similar steps as in [11, proof of Lemma 1], we get

$$|z_i(t+1) - \bar{x}(t)| \leq \frac{2}{y_i(t+1)} \Bigg(\max_j |[D(t:0)]_{ij}| \|x(0)\|_1 + \sum_{s=1}^{t} \max_j |[D(t:s)]_{ij}| \|\epsilon(s)\|_1 \Bigg).$$

Using (7) we obtain the desired result. ■

In the following, we state our first result on the connectivity of the matrix sequence $\{W(t)\}$ for the given random setting. The main challenge, unlike in [19], is that this sequence depends not only on time but also on the states $y_i(t)$.

Lemma 1: Consider the MPP algorithm (4) with $W(t)$ as the weighted adjacency matrix of the effective communication network at time t. Suppose Assumption 1 holds. Then we have

i) defining $p = (\min_{b:p_b>0} p_b)^{2n-2}$, for all $t \geq 0$ we have

$$\mathbb{P}(W(t+2n-3:t) \text{ is irreducible}) \geq p > 0. \tag{8}$$

ii) $\{W(t)\}$ has the directed infinite flow property.

Proof: We start by proving (i). Since the communication networks are assumed to have self-loops at all nodes, if the sequence $G(t), G(t+1), \ldots, G(t+2n-3)$ is such that $\bigcup_{t'=t}^{t+2n-3} G(t)$ is strongly connected, then $W(t+2n-3:t)$ is irreducible. Thus, it is sufficient to show that such a sequence occurs with probability at least p. To do this, we construct events $\mathcal{A}_1$ and $\mathcal{A}_2$ so that under $\mathcal{A}_1 \cap \mathcal{A}_2$, $\bigcup_{t'=t}^{t+2n-3} G(t)$ is strongly connected and $\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) \geq p$.

Similar to the setting in the push-sum algorithm in [14, Proposition 2.2], it is easy to check that $\sum_{i\in[n]} y_i(t) = n$ for all t. This is known as the mass conservation property. Therefore, there exists a node $v_{i_0} \in \mathcal{V}$ with $y_{i_0}(t) \geq 1$. Since the graphs contain self-loops at all nodes, in each iteration node i sends $1/d_i(t)$ share of its values to itself. As $d_i(t) \leq n$ for all i, we have $y_i(t) \geq \frac{1}{n} y_i(t-1)$. This along with $y_{i_0}(t) \geq 1$, implies that $y_{i_0}(t') \geq \delta$, for all $t \leq t' \leq t+2n$. In other words, v_{i_0} will remain active for all $t \leq t' \leq t+2n$.

Since $\bigcup_{b:p_b>0} \bar{G}_b$ is strongly connected by assumption, there exists a graph $\bar{G}_{b_0}$ with probability $p_{b_0} > 0$, in which v_{i_0} is connected to some other node $v_{i_1} \in \mathcal{V}\backslash\{v_{i_0}\}$, i.e., $e_0 := (v_{i_0}, v_{i_1}) \in \bar{\mathcal{E}}_{b_0}$. When $\bar{G}(t) = \bar{G}_{b_0}$, v_{i_1} receives $1/d_{i_0}(t)$ share of v_{i_0}'s values. Thus, $y_{i_1}(t+1) \geq \frac{1}{n}$, which again implies $y_{i_1}(t') \geq \delta$, for all $t+1 \leq t' \leq t+2n$, i.e., v_{i_1} will remain active for all $t+1 \leq t' \leq t+2n$.

Similarly, there exists a graph $\bar{G}_{b_1}$, occurring with probability $p_{b_1} > 0$, where either $(v_{i_0}, v_{i_2}) \in \bar{\mathcal{E}}_{b_1}$ or $(v_{i_1}, v_{i_2}) \in \bar{\mathcal{E}}_{b_1}$ for some $v_{i_2} \in \mathcal{V}\backslash\{v_{i_0}, v_{i_1}\}$. We refer to this edge as e_1. When $\bar{G}(t) = \bar{G}_{b_0}$ and $\bar{G}(t+1) = \bar{G}_{b_1}$, which happens with probability $p_{b_1}p_{b_2}$, we have $y_{i_2}(t+2) \geq \frac{1}{n^2}$ and hence, $y_{i_2}(t') \geq \delta$, for all $t+2 \leq t' \leq t+2n$. In that event, v_{i_2} will remain active for all $t+2 \leq t' \leq t+2n$.

Continuing this argument, we obtain the event

$$\mathcal{A}_1 = \{\bar{G}(t) = \bar{G}_{b_0}, \bar{G}(t+1) = \bar{G}_{b_1}, \ldots, \bar{G}(t+n-2) = \bar{G}_{b_{n-2}}\}.$$

If $\mathcal{A}_1$ occurs, which has probability at least $(\min_{b:p_b>0} p_b)^{n-1}$, then $S_1 = \{e_0, \ldots, e_{n-2}\}$ represents the set of edges in a directed spanning tree of $\bigcup_{t'=t}^{t+n-2} G(t)$ rooted at v_{i_0}, noting that $e_i \in \mathcal{E}(t+i)$. Also, when $\mathcal{A}_1$ occurs

1) for $\tau = 0, \ldots, n-1$, node v_{i_τ} remains active for all time t' such that $t+\tau \leq t' \leq t+2n$ (note that $\{v_{i_0}, v_{i_1}, \ldots, v_{i_{n-1}}\} = \mathcal{V}$);
2) in $\bigcup_{t'=t}^{t+n-2} G(t)$ there is a path from v_{i_0} to all nodes in $\mathcal{V}$.

Next, we construct $\mathcal{A}_2$ where there is a path from all nodes to v_{i_0} in $\bigcup_{t'=t+n-1}^{t+2n-2} G(t)$. Thus, when $\mathcal{A}_1 \cap \mathcal{A}_2$ occurs, every pair of nodes in $\bigcup_{t'=t}^{t+2n-2} G(t)$ is connected, making it strongly connected. The probability of $\mathcal{A}_1 \cap \mathcal{A}_2$ is at least $(\min_{b:p_b>0} p_b)^{2n-2}$, yielding the desired result.

Since $\bigcup_{b:p_b>0} \bar{G}_b$ is strongly connected, at time $t+n-1$ there exists a graph $\bar{G}_{b_{n-1}}$ with probability $p_{b_{n-1}} > 0$, in which some node $v_{i_n} \in \mathcal{V}\backslash\{v_{i_0}\}$ is connected to v_{i_0}, i.e., $e_{n-1} := (v_{i_n}, v_{i_o})$. Similarly, at time $t+n$ there exists $\bar{G}_{b_n}$ with probability $p_{b_n} > 0$ such that for some $v_{i_{n+1}} \in \mathcal{V}\backslash\{v_{i_0}, v_{i_n}\}$ either $(v_{i_{n+1}}, v_{i_0}) \in \mathcal{E}_{b_n}$ or $(v_{i_{n+1}}, v_{i_n}) \in \mathcal{E}_{b_n}$. We refer to this edge as e_n. Continuing this argument, we obtain our desired event

$$\mathcal{A}_2 = \{\bar{G}(t+n-1) = \bar{G}_{b_{n-1}},$$
$$\bar{G}(t+n) = \bar{G}_{b_n}, \ldots, \bar{G}(t+2n-3) = \bar{G}_{b_{2n-3}}\}.$$

If $\mathcal{A}_2$ occurs, which has probability at least $(\min_{b:p_b>0} p_b)^{n-1}$, there is a path from all nodes to v_{i_0} in $\bigcup_{t'=t+n-1}^{t+2n-3} G(t)$. This completes the proof of part (i).

The proof of part (ii) is similar to the proof of [19, Lemma 1 part (i)]. Since every positive entry of $W(t)$ is bounded below by $1/n$, using the Borel–Cantelli lemma, (8) implies that $\sum_{t=0}^{\infty} W_{S\bar{S}}(t) = \infty$, a.s., for any nontrivial $S \subset [n]$. Thus, $\{W(t)\}$ has the directed infinite flow property. ■

Next is a useful consequence of this result and Proposition 1.

Corollary 1: Consider MPP algorithm (4) with $W(t)$ as the weighted adjacency matrix of the effective communication network at time t. Suppose Assumption 1 holds. We have

$$|z_i(t+1) - \bar{x}(t)| \leq \frac{2}{\delta}\left(\Lambda_{t,0}\|x(0)\|_1 + \sum_{s=1}^{t} \Lambda_{t,s}\|\epsilon(s)\|_1\right) \quad (9)$$

a.s., where $\Lambda_{t,s} \in (0,1)$ for all $t \geq s \geq 0$.

The next technical lemma is used in an upcoming result.

Lemma 2: Let $c \geq 2$. For all integers $n \geq 2$ and $\zeta > 0$, we have $e^{-\zeta} \leq 1 - \frac{1}{n^{\frac{c}{\zeta}}}$.

Proof: Since $\frac{1}{n^{\frac{c}{\zeta}}} \leq e^{-\frac{1}{\zeta}}$, for $c \geq 2$, it is enough to show $e^{-\zeta} + e^{-\frac{1}{\zeta}} \leq 1$, for $\zeta > 0$. By symmetry we focus on the case where $0 < \zeta \leq 1$. The function $\zeta \log \zeta$ is convex in this range, leading to $\zeta \log \zeta \geq \zeta - 1$. In addition, since $\zeta^2 \leq \zeta$ for $0 < \zeta \leq 1$, we have $\zeta \log \zeta \geq \zeta^2 - 1$. Dividing both sides by ζ, rearranging and using the monotonicity of the exponential function we obtain $e^{\log \zeta - \zeta} \geq e^{-\frac{1}{\zeta}}$, which simplifies to $\zeta e^{-\zeta} \geq e^{-\frac{1}{\zeta}}$. The well-known inequality $e^{\zeta} \geq 1 + \zeta$ implies $1 \geq e^{-\zeta} + \zeta e^{-\zeta}$. Combining this and $\zeta e^{-\zeta} \geq e^{-\frac{1}{\zeta}}$ yields $e^{-\zeta} + e^{-\frac{1}{\zeta}} \leq 1$, completing the proof for $0 < \zeta \leq 1$. ■

Lemma 3: Consider MPP algorithm (4) with $W(t)$ as the weighted adjacency matrix of the effective communication network at time t. Suppose Assumption 1 holds. Let $\theta := B + \frac{2nB}{p}$, where $B = 2n-2$. Then,

i) for all $t \geq s + \theta \geq 0$ we have

$$\mathbb{P}\left(\Lambda_{t,s} > 2\lambda^{t-s}\right) \leq 8e^{-c_1(t-s)}, \text{ and}$$
$$\mathbb{E}[\Lambda_{t,s}] \leq 10\lambda^{t-s} \quad (10)$$

where $\lambda = \left(1 - \frac{1}{n^{(\frac{4nB}{p})}}\right)^{\frac{p}{2nB}} \in (0,1)$, $c_1 = \frac{p^2}{4B}$, and $\Lambda_{t,s}$ is defined in Proposition 1;

ii) for all $s \geq 0$ we have $\mathbb{P}(\lim_{t\to\infty} \Lambda_{t,s} = 0) = 1$.

Proof: To prove part (i) we first provide a series of definitions. Define a sequence of events

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

By Lemma 1 part (i), we have $\mathbb{P}(X_B(t) = 1) \geq p > 0$. Define the partial sums $H_B(T) := \sum_{t=0}^{T} X_B(t)$ for all $T \geq 0$, and let $q_t := \max\{q : k_q \leq t\}$, where k_q is defined in (5). Note that $H_B(\cdot)$ is a counter indicating the number of consecutive time windows of length B with communication between each subset of nodes and its complement. On the other hand, q_t counts the number of consecutive time windows of any length with communication between each subset of nodes and its complement. By definition of $H_B(\cdot)$ and q_t, we have that

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

and therefore,

$$\mathbb{P}\left(q_t \leq \frac{pt}{2B}\right) \leq \mathbb{P}\left(H_B\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) \leq \frac{pt}{2B}\right).$$

We find an upper bound for the two sides of the inequality by finding a bound on the RHS for which we have

$$\mathbb{P}\left(H_B\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) \leq \frac{pt}{2B}\right)$$
$$= \mathbb{P}\left(\sum_{t'=0}^{\lfloor t/B \rfloor - 1} X_B(t') - p\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) \leq -\beta_t\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right)\right)$$
$$\leq \mathbb{P}\left(\sum_{t'=0}^{\lfloor t/B \rfloor - 1} (X_B(t') - \mathbb{E}[X_B(t')]) \leq -\beta_t\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right)\right)$$

where $\beta_t = p - \frac{\frac{pt}{2B}}{\lfloor \frac{t}{B} \rfloor - 1}$. The last inequality follows from the fact that for all t, $\mathbb{E}[X_B(t)] = \mathbb{P}(X_B(t) = 1) \geq p$. Since $\beta_t > 0$ for $t \geq \theta$, using Hoeffding's inequality, we obtain

$$\mathbb{P}\left(H_B\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) \leq \frac{pt}{2B}\right) \leq e^{-\beta_t^2(\lfloor \frac{t}{B} \rfloor - 1)}.$$

To simplify this inequality and find an upper bound on the RHS, we will find a lower bound on positive part of the power of the exponent function. We have

$$\beta_t^2\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) = p^2\left(\left\lfloor \frac{t}{B} \right\rfloor - 1\right) - \frac{p^2 t}{B} + \frac{\frac{p^2 t^2}{4B^2}}{\left(\lfloor \frac{t}{B} \rfloor - 1\right)}$$

$$\geq p^2\left(\frac{t}{B}-2\right)-\frac{p^2t}{B}+\frac{\frac{p^2t^2}{4B^2}}{\frac{t}{B}}$$

$$= -2p^2+\frac{p^2t}{4B}\geq -2+\frac{p^2t}{4B}.$$

Therefore,

$$\mathbb{P}\left(H_B\left(\left\lfloor\frac{t}{B}\right\rfloor-1\right)\leq\frac{pt}{2B}\right)\leq 8e^{-c_1t}$$

and consequently $\mathbb{P}(q_t\leq\frac{pt}{2B})\leq 8e^{-c_1t}$.

Now, we need to "translate" the term in parentheses on the left hand side to a bound on $\Lambda_{t,0}$. When $q_t>\frac{pt}{2B}$, we have $|Q_{t,0}|\geq\lfloor\frac{pt}{2nB}\rfloor$; therefore, using [19, Lemma A.4], $\Lambda_{t,0}$ can be bounded above as

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

In [19, Proof of Lemma 3], it is shown that the RHS is bounded above by $2\lambda^t$. Therefore, when $q_t>\frac{pt}{2B}$, we have $\Lambda_{t,0}\leq 2\lambda^t$. This implies that

$$\mathbb{P}\left(\Lambda_{t,0}>2\lambda^t\right)\leq 8e^{-c_1t}$$

for $t\geq\theta$. Following similar arguments, we can also obtain for all $t\geq s+\theta$

$$\mathbb{P}\left(\Lambda_{t,s}>2\lambda^{t-s}\right)\leq 8e^{-c_1(t-s)}$$

finishing the proof of the first part.

To prove the second part we use the law of total expectations. For all $t\geq\theta$

$$\begin{aligned}\mathbb{E}[\Lambda_{t,0}]&=\mathbb{E}\left[\Lambda_{t,0}\mid q_t>\frac{pt}{2B}\right]\mathbb{P}\left(q_t>\frac{pt}{2B}\right)\\&\quad+\mathbb{E}\left[\Lambda_{t,0}\mid q_t\leq\frac{pt}{2B}\right]\mathbb{P}\left(q_t\leq\frac{pt}{2B}\right)\\&\leq\mathbb{E}\left[\Lambda_{t,0}\mid q_t>\frac{pt}{2B}\right]+\mathbb{P}\left(q_t\leq\frac{pt}{2B}\right)\\&\leq 2\lambda^t+8e^{-c_1t}.\end{aligned}\tag{11}$$

Using Lemma 2, $e^{-\zeta}\leq\left(1-\frac{1}{n^{\frac{2n^2B}{\zeta}}}\right)$ for all $\zeta>0$. Therefore,

$$e^{-c_1t}=e^{-\frac{p^2t}{4B}}=e^{-\frac{pn}{2}\cdot\frac{pt}{2nB}}\leq\left(1-\frac{1}{n^{\frac{2n^2B}{\frac{pn}{2}}}}\right)^{\frac{pt}{2nB}}=\lambda^t.$$

This along with (11) implies for all $t\geq\theta$ that $\mathbb{E}[\Lambda_{t,0}]\leq 10\lambda^t$. Again, by following similar arguments, we can also obtain $\mathbb{E}[\Lambda_{t,s}]\leq 10\lambda^{t-s}$ for all $t\geq s+\theta$. Part (ii) is a direct consequence of part (i) and the Borel–Cantelli lemma. ∎

In Lemma 3, we found an upper bound on $\mathbb{E}[\Lambda_{t,s}]$ for $t\geq s+\theta$. When $t<s+\theta$, since $\Lambda_{t,s}\in(0,1)$, we have $\mathbb{E}[\Lambda_{t,s}]\leq 1$. Therefore, for all t we can write

$$\mathbb{E}[\Lambda_{t,s}]\leq\eta_1\lambda^{t-s}\tag{12}$$

where $\eta_1=\frac{10}{\lambda^\theta}$.

Lemma 3 provides the foundation for the subsequent technical result, which is instrumental in proving our convergence rate results in Theorem 1.

Lemma 4: Consider MPP algorithm (4) with $W(t)$ as the weighted adjacency matrix of the effective communication network at time t. Suppose Assumption 1 holds. In addition, assume that the perturbations $\epsilon_i(t)$ are bounded as follows:

$$\|\epsilon(t)\|_1\leq\frac{U}{t^\gamma},\quad\text{for all } t\geq 1$$

for some $U>0$. Then,

i) $\lim_{t\to\infty}|z_i(t+1)-\bar{x}(t)|=0$, a.s.;

ii) $\sum_{t=0}^{\infty}\alpha(t+1)|z_i(t+1)-\bar{x}(t)|<\infty$, a.s.;

iii)

$$\begin{aligned}&\mathbb{E}\left[\frac{1}{\sum_{k=0}^{t}\alpha(k+1)}\sum_{k=0}^{t}\alpha(k+1)\,|z_i(k+1)-\overline{x}(k)|\right]\\&\quad\leq\frac{2\eta_1\Gamma(t)}{\delta(1-\lambda)}\cdot\left(\|x(0)\|_1+U\left(1+\frac{1}{2\gamma-1}\right)\right)\end{aligned}$$

where $\Gamma(t)=\frac{1-\gamma}{(t+1)^{1-\gamma}-1}$.

Proof: We start by proving (i). By Corollary 1, it suffices to show $\lim_{t\to\infty}(\Lambda_{t,0}\|x(0)\|_1+\sum_{s=1}^{t}\Lambda_{t,s}\|\epsilon(s)\|_1)=0$, a.s. As shown in Lemma 3 (ii), $\lim_{t\to\infty}\Lambda_{t,0}=0$, a.s. and therefore, it remains to show

$$\lim_{t\to\infty}\sum_{s=1}^{t}\Lambda_{t,s}\|\epsilon(s)\|_1=0,\quad\text{a.s.}$$

For all $t\geq 1$ define $\tau_t:=\left\lceil\frac{2}{c_1}\ln(t)\right\rceil$, where c_1 is the scalar constant given in Lemma 3 (i). To study this limit, we break the summation into the following two sums:

$$\sum_{s=1}^{t}\Lambda_{t,s}\|\epsilon(s)\|_1=\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\|\epsilon(s)\|_1+\sum_{s=t-\tau_t+1}^{t}\Lambda_{t,s}\|\epsilon(s)\|_1.\tag{13}$$

For the first sum on the RHS of (13) we have

$$\begin{aligned}&\mathbb{P}\left(\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\|\epsilon(s)\|_1>\sum_{s=1}^{t-\tau_t}2\lambda^{t-s}\|\epsilon(s)\|_1\right)\\&\quad\leq\mathbb{P}\left(\bigcup_{s=1}^{t-\tau_t}\{\Lambda_{t,s}>2\lambda^{t-s}\}\right)\leq\sum_{s=1}^{t-\tau_t}\mathbb{P}\left(\Lambda_{t,s}>2\lambda^{t-s}\right).\end{aligned}$$

To further simplify this inequality, consider two cases.

1) *Case 1:* When $t-\tau_t\leq\theta$,

$$\sum_{s=1}^{t-\tau_t}\mathbb{P}\left(\Lambda_{t,s}>2\lambda^{t-s}\right)\leq\theta.$$

2) *Case 2:* When $t-\tau_t>\theta$, by (10)

$$\sum_{s=1}^{t-\tau_t}\mathbb{P}\left(\Lambda_{t,s}>2\lambda^{t-s}\right)\leq\sum_{s=1}^{t-\tau_t}8e^{-c_1(t-s)}\leq\frac{8t^{-2}}{1-e^{-c_1}}.$$

Combining the two cases, for some $\eta_2>0$ and all t we have

$$\mathbb{P}\left(\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\ \|\epsilon(s)\|_1>\sum_{s=1}^{t-\tau_t}2\lambda^{t-s}\|\epsilon(s)\|_1\right)\leq\eta_2t^{-2}.\tag{14}$$

If we sum (14) over t we obtain

$$\sum_{t=1}^{\infty}\mathbb{P}\left(\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\|\epsilon(s)\|_1>\sum_{s=1}^{t-\tau_t}2\lambda^{t-s}\|\epsilon(s)\|_1\right)\leq\eta_2\sum_{t=1}^{\infty}t^{-2}.$$

Since the RHS is finite, by the Borel–Cantelli lemma, there exists a (random) t', a.s., such that for all $t\geq t'$

$$\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\|\epsilon(s)\|_1\leq\sum_{s=1}^{t-\tau_t}2\lambda^{t-s}\|\epsilon(s)\|_1.$$

It is shown in the proof of [11, Lemma 1 (b)] (alternatively, it can be shown by elementary estimates), that for our choice of λ and $\epsilon(s)$, $\lim_{t\to\infty}\sum_{s=1}^{t-\tau_t}\lambda^{t-s}\|\epsilon(s)\|_1 = 0$, and therefore,

$$\lim_{t\to\infty}\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\|\epsilon(s)\|_1 = 0, \quad \text{a.s.} \tag{15}$$

For the second summation on the RHS of (13) recall that by assumption $\|\epsilon(s)\|_1 \le \frac{U}{s^\gamma}$. This along with the fact that $\Lambda_{t,s} \in (0,1)$ implies that

$$\begin{aligned}\sum_{s=t-\tau_t+1}^{t}\Lambda_{t,s}\|\epsilon(s)\|_1 &\le \sum_{s=t-\tau_t+1}^{t}\|\epsilon(s)\|_1 \le \sum_{s=t-\tau_t+1}^{t}\frac{U}{s^\gamma}\\ &\le \tau_t \cdot \max_{t-\tau_t<s\le t}\frac{U}{s^\gamma}\\ &\le \frac{\frac{2U}{c_1}\ln(t)+1}{\left(t-\frac{2}{c_1}\ln(t)-1\right)^\gamma} =: \delta_t.\end{aligned}$$

It is easy to show that $\lim_{t\to\infty}\delta_t = 0$, which along with (15) gives us our desired result.

Now, we prove part (ii). By Corollary 1, we have

$$\begin{aligned}&\sum_{t=0}^{\infty}\alpha(t+1)\,|z_i(t+1)-\bar{x}(t)|\\ &\le \frac{2}{\delta}\left(\sum_{t=0}^{\infty}\alpha(t+1)\Lambda_{t,0}\|x(0)\|_1 + \sum_{t=0}^{\infty}\sum_{s=1}^{t}\Lambda_{t,s}\alpha(s)\|\epsilon(s)\|_1\right).\end{aligned}$$

Similar to part (i) using (10) and the fact that $\alpha(t+1) = \frac{1}{(t+1)^\gamma}$, $\frac{1}{2} < \gamma < 1$, it can be seen that the first term on the RHS is finite, a.s. For the second term on the RHS, we break the summation into two as follows:

$$\begin{aligned}\sum_{t=1}^{\infty}\sum_{s=1}^{t}\Lambda_{t,s}\alpha(s)\|\epsilon(s)\|_1 &= \sum_{t=1}^{\infty}\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\alpha(s)\|\epsilon(s)\|_1\\ &\quad+\sum_{t=1}^{\infty}\sum_{s=t-\tau_t+1}^{t}\Lambda_{t,s}\alpha(s)\|\epsilon(s)\|_1.\end{aligned} \tag{16}$$

For the first sum on the RHS of (16), as in part (i), we have

$$\begin{aligned}&\mathbb{P}\left(\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\alpha(s)\|\epsilon(s)\|_1 > \sum_{s=1}^{t-\tau_t}2\lambda^{t-s}\alpha(s)\|\epsilon(s)\|_1\right)\\ &\le \mathbb{P}\left(\bigcup_{s=1}^{t-\tau_t}\{\Lambda_{t,s} > 2\lambda^{t-s}\}\right) \le \eta_2 t^{-2}.\end{aligned} \tag{17}$$

If we sum (17) over t we have

$$\sum_{t=1}^{\infty}\mathbb{P}\left(\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\alpha(s)\|\epsilon(s)\|_1 > \sum_{s=1}^{t-\tau_t}2\lambda^{t-s}\alpha(s)\|\epsilon(s)\|_1\right) < \infty.$$

Hence, by the Borel–Cantelli lemma, there exists a (random) t'', a.s., such that for all $t \ge t''$

$$\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\alpha(s)\|\epsilon(s)\|_1 \le \sum_{s=1}^{t-\tau_t}2\lambda^{t-s}\alpha(s)\|\epsilon(s)\|_1.$$

It is shown in the proof of [11, Lemma 1 (c)] that for our choice of λ and $\epsilon(s)$, $\lim_{t\to\infty}\sum_{s=1}^{t-\tau_t}\lambda^{t-s}\alpha(s)\|\epsilon(s)\|_1 < \infty$, and therefore,

$$\lim_{t\to\infty}\sum_{s=1}^{t-\tau_t}\Lambda_{t,s}\|\epsilon(s)\|_1 < \infty, \quad \text{a.s.} \tag{18}$$

For the second sum on the RHS of (16) we have

$$\sum_{t=1}^{\infty}\sum_{s=t-\tau_t+1}^{t}\Lambda_{t,s}\alpha(s)\|\epsilon(s)\|_1 \le \sum_{t=1}^{\infty}\sum_{s=t-\tau_t+1}^{t}\frac{U}{s^{2\gamma}} \le \Delta$$

where

$$\Delta = U\sum_{t=1}^{\infty}\frac{\frac{2}{c_1}\ln(t)+1}{(t-\frac{2}{c_1}\ln(t)-1)^{2\gamma}}.$$

Since $\gamma > \frac{1}{2}$ we have $\Delta < \infty$, which along with (18) gives us our desired result.

Finally, we prove part (iii). Using Corollary 1 we have

$$\begin{aligned}&\sum_{k=1}^{t}\alpha(k+1)\,|z_i(k+1)-\bar{x}(k)|\\ &\le \sum_{k=1}^{t}\frac{2}{\delta(k+1)^\gamma}\left(\Lambda_{k,0}\|x(0)\|_1 + \sum_{s=1}^{k}\Lambda_{k,s}\|\epsilon(s)\|_1\right)\\ &= \frac{2\|x(0)\|_1}{\delta}\sum_{k=1}^{t}\frac{\Lambda_{k,0}}{(k+1)^\gamma} + \frac{2}{\delta}\sum_{k=1}^{t}\sum_{s=1}^{k}\frac{\Lambda_{k,s}\|\epsilon(s)\|_1}{(k+1)^\gamma}\\ &\le \frac{2\|x(0)\|_1}{\delta}\sum_{k=1}^{t}\Lambda_{k,0} + \frac{2U}{\delta}\sum_{k=1}^{t}\sum_{s=1}^{k}\frac{\Lambda_{k,s}}{s^{2\gamma}}.\end{aligned}$$

Taking expectations on both sides and using (12), we have

$$\begin{aligned}&\mathbb{E}\left[\sum_{k=1}^{t}\alpha(k+1)\,|z_i(k+1)-\bar{x}(k)|\right]\\ &\le \mathbb{E}\left[\frac{2\|x(0)\|_1}{\delta}\sum_{k=1}^{t}\Lambda_{k,0} + \frac{2U}{\delta}\sum_{k=1}^{t}\sum_{s=1}^{k}\frac{\Lambda_{k,s}}{s^{2\gamma}}\right]\\ &\le \frac{2\eta_1\|x(0)\|_1}{\delta}\sum_{k=1}^{t}\lambda^k + \frac{2\eta_1 U}{\delta}\sum_{k=1}^{t}\sum_{s=1}^{k}\frac{\lambda^{k-s}}{s^{2\gamma}}\\ &\le \frac{2\eta_1\|x(0)\|_1}{\delta}\frac{\lambda}{1-\lambda} + \frac{2\eta_1 U}{\delta}\sum_{k=1}^{t}\sum_{s=1}^{k}\frac{\lambda^{k-s}}{s^{2\gamma}}.\end{aligned}$$

For the second term on the RHS we have

$$\begin{aligned}\sum_{k=1}^{t}\sum_{s=1}^{k}\frac{\lambda^{k-s}}{s^{2\gamma}} &= \sum_{s=1}^{t}\frac{1}{s^{2\gamma}}\sum_{k=s}^{t}\lambda^{k-s} \le \sum_{s=1}^{t}\frac{1}{s^{2\gamma}}\frac{1}{1-\lambda}\\ &\le \frac{1}{1-\lambda}\left(1+\int_1^\infty\frac{du}{u^{2\gamma}}\right)\\ &= \frac{1}{1-\lambda}\left(1+\frac{1}{2\gamma-1}\right).\end{aligned}$$

In view of these two bounds and since $0 < \lambda < 1$, we obtain

$$\begin{aligned}&\mathbb{E}\left[\sum_{k=0}^{t}\alpha(k+1)\,|z_i(k+1)-\bar{x}(k)|\right]\\ &\le \frac{2\eta_1}{\delta(1-\lambda)}\left(\|x(0)\|_1 + U\left(1+\frac{1}{2\gamma-1}\right)\right).\end{aligned} \tag{19}$$

In addition, we have the following inequality

$$\begin{aligned}\sum_{k=0}^{t}\alpha(k+1) &= \sum_{k=1}^{t+1}\frac{1}{k^\gamma} \ge \int_1^{t+1}\frac{1}{u^\gamma}du\\ &= \frac{(t+1)^{1-\gamma}-1}{1-\gamma} =: \frac{1}{\Gamma(t)}.\end{aligned} \tag{20}$$

Thus, by (19) and (20), we obtain the result for part (iii). ■

Finally, the following lemmas enable us to prove our main result. Lemma 5 establishes conditions for the convergence of a sequence with a subgradient-like iteration. Using these conditions, we prove Theorem 1 by analyzing the average progress of the nodes' states. Lemma 6 confirms that the average values meet the requirements of Lemma 5.

Lemma 5 ([11, Lemma 7]): Consider a convex minimization problem $\min_{x\in\mathbb{R}^m} f(x)$, where $f:\mathbb{R}^m\to\mathbb{R}$ is continuous, and assume the solution set X^* is nonempty. Let $\{x_t\}$ be a sequence such that for all $x\in X^*$ and $t\geq 0$

$$\|x_{t+1}-x\|^2 \leq \|x_t - x\|^2 - \beta_t(f(x_t)-f(x)) + c_t$$

where $\beta_t\geq 0$ and $c_t\geq 0$ for all $t\geq 0$, with $\sum_{t=0}^{\infty}\beta_t=\infty$ and $\sum_{t=0}^{\infty} c_t<\infty$. Then, the sequence $\{x_t\}$ converges to some solution $x^*\in X^*$.

Lemma 6 ([11, Lemma 8]): Consider the function $F(\mathbf{z})=\sum_{i=1}^n f_i(\mathbf{z})$, where $f_i(\mathbf{z})$ are convex functions on $\mathbb{R}^d$ and Lipschitz continuous with constants $L_i<\infty$. Then, for all $\mathbf{v}\in\mathbb{R}^d$ and $t\geq 0$,

$$\|\bar{\mathbf{x}}(t+1)-\mathbf{v}\|^2 \leq \|\bar{\mathbf{x}}(t)-\mathbf{v}\|^2 - \frac{2\alpha(t+1)}{n}(\mathbf{F}(\bar{\mathbf{x}}(t))-\mathbf{F}(\mathbf{v})) + \frac{4\alpha(t+1)}{n}\sum_{i=1}^n L_i\|\mathbf{z}_i(t+1)-\bar{\mathbf{x}}(t)\| + \alpha^2(t+1)\frac{L^2}{n^2}$$

where $\bar{\mathbf{x}}(t)=\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i(t)$, generated by the MSP algorithm.

Now, we prove Theorem 1.

Theorem 1: We start by proving part (i). Having established the necessary results to accommodate the random nature of the underlying communication networks, the proof follows similar steps as in [11, Proof of Theorem 1]. Applying Lemma 4 (i) to each coordinate, we obtain

$$\lim_{t\to\infty}\|\mathbf{z}_i(t+1)-\bar{\mathbf{x}}(t)\| = 0 \quad \text{for all } i\in[n].$$

Moreover, by Lemma 4 (ii)

$$\sum_{t=0}^{\infty}\alpha(t+1)\|\mathbf{z}_i(t+1)-\bar{\mathbf{x}}(t)\|<\infty \quad \text{for all } i\in[n].$$

On the other hand, by Lemma 6, for all optimal solutions $\mathbf{z}^*$

$$\|\bar{\mathbf{x}}(t+1)-\mathbf{z}^*\|^2 \leq \|\bar{\mathbf{x}}(t)-\mathbf{z}^*\|^2 - \frac{2\alpha(t+1)}{n}(\mathbf{F}(\bar{\mathbf{x}}(t))-\mathbf{F}(\mathbf{z}^*)) + \frac{4\alpha(t+1)}{n}\sum_{i=1}^n L_i\|\mathbf{z}_i(t+1)-\bar{\mathbf{x}}(t)\| + \alpha^2(t+1)\frac{L^2}{n^2}.$$

Thus, with our choice of $\alpha(t)$, all the conditions of Lemma 5 are satisfied, yielding the desired result.

Next, we prove part (ii). Recall that

$$\tilde{\mathbf{z}}_i(t+1)=\frac{\alpha(t+1)\mathbf{z}_i(t+1)+S(t)\tilde{\mathbf{z}}_i(t)}{S(t+1)},\quad t\geq 0 \tag{21}$$

and $S(t)=\sum_{s=1}^t\alpha(s)$ with $S(0)=0$. Rearranging the terms, we have

$$S(t+1)\tilde{\mathbf{z}}_i(t+1)-S(t)\tilde{\mathbf{z}}_i(t)=\alpha(t+1)\mathbf{z}_i(t+1). \tag{22}$$

Summing both sides over t and using telescoping sum yield

$$\tilde{\mathbf{z}}_i(t+1)=\frac{\sum_{k=0}^t\alpha(k+1)\mathbf{z}_i(k+1)}{\sum_{k=0}^t\alpha(k+1)},\quad \text{for all } t\geq 0.$$

Since each $\mathbf{g}_i$ is bounded by L_i and $L=\sum_{i\in[n]}L_i$, we have

$$F(\tilde{\mathbf{z}}_i(t+1)) - F\left(\frac{\sum_{k=0}^t\alpha(k+1)\bar{\mathbf{x}}(k)}{\sum_{k=0}^t\alpha(k+1)}\right) = F\left(\frac{\sum_{k=0}^t\alpha(k+1)\mathbf{z}_i(k+1)}{\sum_{k=0}^t\alpha(k+1)}\right) - F\left(\frac{\sum_{k=0}^t\alpha(k+1)\bar{\mathbf{x}}(k)}{\sum_{k=0}^t\alpha(k+1)}\right) \leq \frac{L}{\sum_{k=0}^t\alpha(k+1)}\sum_{k=0}^t\alpha(k+1)\|\mathbf{z}_i(k+1)-\bar{\mathbf{x}}(k)\|.$$

Treating each coordinate of $\alpha(t+1)\mathbf{g}_i(t+1)$ as a perturbation $\epsilon_j(t)$ and recalling that $\alpha(t)=\frac{1}{t^\gamma}$ so that $\|\epsilon(t)\|_1\leq\frac{L}{t^\gamma}$, we apply Lemma 4 (iii) to the coordinates of the vectors $\mathbf{z}_i(k+1)$ and $\bar{\mathbf{x}}(t)$ to obtain

$$\mathbb{E}\left[F(\tilde{\mathbf{z}}_i(t+1)) - F\left(\frac{\sum_{k=0}^t\alpha(k+1)\bar{\mathbf{x}}(k)}{\sum_{k=0}^t\alpha(k+1)}\right)\right] \leq \frac{2\eta_1 L\Gamma(t)}{\delta(1-\lambda)}\left(\sum_{j=1}^n\|\mathbf{x}_j(0)\|_1 + dL(1+\frac{1}{2\gamma-1})\right). \tag{23}$$

Also, by Lemma 6, using telescopic summation, and dividing both sides by $2S(t+1)/n$, we obtain

$$\frac{\sum_{k=0}^t\alpha(k+1)F(\bar{\mathbf{x}}(k))}{S(t+1)} - F(\mathbf{z}^*) \leq \frac{n}{2}\frac{\|\bar{\mathbf{x}}(0)-\mathbf{z}^*\|^2}{S(t+1)} + \frac{1}{S(t+1)}\sum_{k=0}^t\alpha^2(k+1)\frac{L^2}{2n} + \frac{2}{S(t+1)}\sum_{k=0}^t\alpha(k+1)\sum_{i=1}^n L_i\|\mathbf{z}_i(k+1)-\bar{\mathbf{x}}(k)\|.$$

The convexity of F implies that

$$F\left(\frac{\sum_{k=0}^t\alpha(k+1)\bar{\mathbf{x}}(k)}{S(t+1)}\right) - F(\mathbf{z}^*) \leq \frac{n}{2}\frac{\|\bar{\mathbf{x}}(0)-\mathbf{z}^*\|^2}{S(t+1)} + \frac{1}{S(t+1)}\sum_{k=0}^t\alpha^2(k+1)\frac{L^2}{2n} + \frac{2}{S(t+1)}\sum_{k=0}^t\alpha(k+1)\sum_{i=1}^n L_i\|\mathbf{z}_i(k+1)-\bar{\mathbf{x}}(k)\|.$$

As shown in (20) we have $\frac{1}{S(t+1)}\leq\Gamma(t)$. Using this and reapplying Lemma 4 (iii) we obtain

$$\mathbb{E}\left(F\left(\frac{\sum_{k=0}^t\alpha(k+1)\bar{\mathbf{x}}(k)}{S(t+1)}\right) - F(\mathbf{z}^*)\right) \leq n\Gamma(t)\frac{\|\bar{\mathbf{x}}(0)-\mathbf{z}^*\|_1}{2} + \frac{L^2\Gamma(t)}{2n}\left(1+\frac{1}{2\gamma-1}\right) + \frac{4\eta_1 L\Gamma(t)}{\delta(1-\lambda)}\left(\sum_{j=1}^n\|\mathbf{x}_j(0)\|_1 + dL\left(1+\frac{1}{2\gamma-1}\right)\right). \tag{24}$$

By summing (23) and (24), we obtain our desired result. ■

ACKNOWLEDGMENT

The authors would like to thank Professor Behrouz Touri, University of California, San Diego, for his insightful inputs and expertise that greatly assisted this research work. His willingness to give his time so generously is much appreciated.

REFERENCES

[1] P. Rezaeinia, B. Gharesifard, T. Linder, and B. Touri, "Distributed optimization on random graphs," *IFAC-PapersOnLine*, vol. 51, no. 23, pp. 391–395, 2018.
[2] P. Rezaeinia and B. Gharesifard, "Distributed optimization with uncertain communications," 2020, *arXiv:2003.10342*.
[3] J. Fink, N. Michael, S. Kim, and V. Kumar, "Distributed pursuitevasion without global localization via local fronteirs," *Auton. Robots*, vol. 32, no. 1, pp. 81–95, 2012.
[4] A. Nedić, A. Olshevsky, and C. A. Uribe, "Fast convergence rates for distributed non-Bayesian learning," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 5538–5553, Nov. 2017.
[5] I. Necoara, Y. Nesterov, and F. Glineur, "Random block coordinate descent methods for linearly constrained optimization over networks," *J. Optim. Theory Appl.*, vol. 173, pp. 227–254, 2017.
[6] A. Nedić, A. Olshevsky, and M. G. Rabbat, "Network topology and communication-computation tradeoffs in decentralized optimization," *Proc. IEEE*, vol. 106, no. 5, pp. 953–976, May 2018.
[7] B. Gerencsér and J. M. Hendrickx, "Push-sum with transmission failures," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1019–1033, Mar. 2019.
[8] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, no. 9, pp. 803–812, Sep. 1986.
[9] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.
[10] A. Nedic, 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.
[11] 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.
[12] I. Lobel and A. Ozdaglar, "Distributed subgradient methods for convex optimization over random networks," *IEEE Trans. Autom. Control*, vol. 56, no. 6, pp. 1291–1306, Jun. 2011.
[13] B. Gharesifard and J. Cortés, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, Mar. 2014.
[14] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. 44th Annu. IEEE Symp. Found. Comput. Sci.*, 2003, pp. 482–491.
[15] B. Gerencsér and M. Kornyik, "Low complexity convergence rate bounds for the synchronous gossip subclass of push-sum algorithms," *IEEE Contr. Syst. Lett.*, vol. 8, pp. 1283–1288, 2024.
[16] J. M. Hendrickx and J. N. Tsitsiklis, "Fundamental limitations for anonymous distributed systems with broadcast communications," in *Proc. 53rd Annu. Allerton Conf. Commun., Control, Comput.*, 2015, pp. 9–16.
[17] A. Nedic, A. Olshevsky, and W. Shi, "Achieving geometric convergence for distributed optimization over time-varying graphs," *SIAM J. Optim.*, vol. 27, no. 4, pp. 2597–2633, 2017.
[18] H. Reisizadeh, B. Touri, and S. Mohajer, "Distributed optimization over time-varying graphs with imperfect sharing of information," *IEEE Trans. Autom. Control*, vol. 68, no. 7, pp. 4420–4427, Jul. 2023.
[19] 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.
[20] F. Iutzeler, P. Ciblat, and W. Hachem, "Analysis of sum-weight-like algorithms for averaging in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2802–2814, Jun. 2013.
[21] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization methods for large-scale machine learning," *SIAM Rev.*, vol. 60, no. 2, pp. 223–311, 2018.