Research Article

Distributed Constrained Stochastic Subgradient Algorithms Based on Random Projection and Asynchronous Broadcast over Networks

Junlong Zhu,¹ Ping Xie,² Qingtao Wu,² Mingchuan Zhang,²
Ruijuan Zheng,² Jianfeng Guan,¹ and Changqiao Xu¹

¹State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, China

²Information Engineering College, Henan University of Science and Technology, Luoyang, China

Correspondence should be addressed to Junlong Zhu; jlzhu@bupt.edu.cn

Received 22 February 2017; Accepted 17 July 2017; Published 28 September 2017

Academic Editor: Thomas Hanne

Copyright © 2017 Junlong Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We consider a distributed constrained optimization problem over a time-varying network, where each agent only knows its own cost functions and its constraint set. However, the local constraint set may not be known in advance or consists of huge number of components in some applications. To deal with such cases, we propose a distributed stochastic subgradient algorithm over time-varying networks, where the estimate of each agent projects onto its constraint set by using random projection technique and the implement of information exchange between agents by employing asynchronous broadcast communication protocol. We show that our proposed algorithm is convergent with probability 1 by choosing suitable learning rate. For constant learning rate, we obtain an error bound, which is defined as the expected distance between the estimates of agent and the optimal solution. We also establish an asymptotic upper bound between the global objective function value at the average of the estimates and the optimal value.

1. Introduction

Distributed optimization problems have received considerable interest from industry and academia since it arises in many applications, including distributed parameter estimation, detection and source localization [1–4], distributed learning and regression [5–7], resource allocation [8, 9], and distributed power control [10]. The goal of such problems is to minimize a global objective function which is a sum of local cost functions over a network. To achieve this goal, we need to design a distributed optimization algorithm over time-varying networks, where each local cost function and constraint set are private information. Moreover, each agent can exchange information with its neighbors across the time-varying networks. Hence, many distributed optimization algorithms are proposed to address the distributed optimization problems [11–15].

However, each agent may not know the constraint set $\mathcal{X}_i$ beforehand in some applications. Thus, the estimate of each agent i cannot be projected onto the constraint set $\mathcal{X}_i$ by the determinate projection operation, and the determinate

projection-based distributed optimization algorithm [16–18] cannot be directly application in such optimization problems. To deal with this case, a random projection-based distributed optimization algorithm is studied in [19, 20]. Moreover, the determinate projection-based distributed optimization algorithm can be regarded as a special case of the random projection-based distributed optimization algorithm. Therefore, we consider the distributed random projection algorithm in this paper. Furthermore, the local constraint set $\mathcal{X}_i$ is assumed to have the form, $\mathcal{X}_i = \bigcap_{j \in \Theta_i} \mathcal{X}_i^j$, where Θ_i is the index set and $\mathcal{X}_i^j$ is a simple set.

In addition, each agent needs to exchange information with its neighbors over time-varying networks. Hence, the design of communication protocol is a crucial role in the design of distributed optimization algorithm. In practice, gossip communication protocol [21] and broadcast communication protocol [22] are two frequently used communication protocols. An asynchronous distributed random projection algorithm based on gossip communication protocol has been proposed in [23]. However, the broadcast is a natural

communication mode in wireless media. Moreover, compared with the gossip communication protocol, the broadcast communication protocol can improve communication in consensus algorithms. Since the consensus problem can be regarded as a special case of distributed optimization problem, the global objective function is a sum of local cost functions. Hence, the same improvement result can apply to the distributed optimization problem. Furthermore, due to bidirectional communications, the communication bottleneck is created in broadcast communication protocol. To remove this bottleneck, we use the random broadcast communication protocol to exchange information among agents over time-varying networks. Therefore, we propose an asynchronous distributed random subgradient algorithm by using random projection operation and random broadcast communication protocol. Further, we assume that the link can be randomly interrupted in our proposed algorithm. The broadcast-based consensus algorithm has been recently studied in [22, 24, 25]. Besides, [26] proposes an asynchronous distributed projection algorithm based on broadcast communication protocol, where the constraint set is available to every agent and the projection operation is a determinate projection. Unlike [26], we project the estimate of each agent onto its own constraint set by employing random projection operation. Furthermore, each agent only knows its own constraint set and does not know the constraint sets of other agents. Informally, the algorithm in [26] is a special case of our proposed algorithm.

Our objective is to analyze the convergence properties of our proposed algorithm and establish some asymptotic error bounds. The main contributions of this paper are as follows:

- (i) We propose an asynchronous distributed random subgradient algorithm based on random projection operation and randomized broadcast communication protocol, where we consider the case that the link failures are randomly occurring over time-varying networks. Moreover, we assume that each agent only knows its local cost function and its own constraint set.
- (ii) We analyze the convergence properties of our proposed algorithm by appropriately chosen step sizes; we show that the estimates of all agents converge to an optimal solution with probability 1.
- (iii) We also establish some asymptotic error bounds when the step sizes are constant.

The remainder of this paper is organized as follows. We describe the optimization problem of our interest, present the algorithm, and give some assumptions in Section 2. We state the main results of the paper in Section 3. In Section 4, the convergence rate of the algorithm and their proofs are provided. The analysis of error bounds is presented in Section 5. The conclusion of the paper is given in Section 6.

Notation. In this paper, all vectors are column vectors. We use boldface to denote the vectors in $\mathbb{R}^d$ and use normal font to denote scalars or vectors of different dimensions. $\mathbf{x}^\top$ and $A^\top$ denote the transpose operation of a vector $\mathbf{x}$ and a

matrix A , respectively. We use $\|\mathbf{x}\|$ to denote the standard Euclidean norm of a vector $\mathbf{x}$. The notations $\mathbf{1}$ and I denote a vector whose all entries are 1 and the identity matrix of size $n \times n$, respectively. $\mathbb{E}[X]$ denotes the expectation of a random variable X .

2. Algorithm Description and Assumptions

We consider a network which consists of n agents (or nodes), indexed by $1, \dots, n$. At each time t , the network topology is denoted by an undirected graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$, where $\mathcal{V} = \{1, \dots, n\}$ denotes the set of agents and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ denotes the set of edges. We assume that the undirected graph is simple. If there exists a directed edge from agent i to j at time t , then $(i, j) \in \mathcal{E}(t)$. In a connected network, two agents are said to be neighbor of if the agents may be connected directly by an edge; that is, the agents can share information with each other. At time t , we denote the set of neighbors of agent i by $\mathcal{N}_i$; that is, $\mathcal{N}_i = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}(t)\}$. We also assume that communication links may be interrupted at random times.

We consider an optimization problem as follows:

$$\begin{aligned} & \text{minimize} \quad \phi(\mathbf{x}) \triangleq \sum_{i=1}^n \phi_i(\mathbf{x}) \\ & \text{subject to} \quad \mathbf{x} \in \mathcal{X} \triangleq \bigcap_{i=1}^n \mathcal{X}_i, \end{aligned} \tag{1}$$

where $\phi_i : \mathbb{R}^d \mapsto \mathbb{R}$ denotes the convex objective function of agent i and $\mathcal{X}_i$ denotes constraint set of agent i .

To solve problem (1), we propose an asynchronous distributed subgradient random projection algorithm based on randomized broadcast communication protocol. In this paper, we employ the asynchronous time model as in [21] and the randomized broadcast model as in [22].

2.1. Algorithm Description. In asynchronous model, each agent has a virtual clock, and the virtual clock ticks is a Poisson process with rate n . We assume that one agent is waked up at a time. Thus, if an agent is waked up at time t , we use I_t to denote the index of the agent. Since the link can be randomly interrupted, we also use J_t to denote the subset of the neighbors of the agent I_t . Hence, the agent $j \in J_t$ receives the broadcast information with probability p_{ij} ; that is, if $(i, j) \in \mathcal{E}(t)$, then $p_{ij} > 0$. Each agent $i \in J_t$ hears the broadcast information from agent I_t at time t . Hence, if $i \notin J_t$, then

$$\mathbf{x}_i(t) = \mathbf{x}_i(t-1), \tag{2}$$

where $\mathbf{x}_i(t)$ denotes the estimate of agent i at time t . If $i \in J_t$, then the estimate of agent i is updated as follows:

$$\mathbf{z}_i(t) = \beta \mathbf{x}_{I_k}(t-1) + (1-\beta) \mathbf{x}_i(t-1) \tag{3}$$

$$\mathbf{x}_i(t) = P_{\mathcal{X}_i^{\omega_i(t)}} [\mathbf{z}_i(t) - \alpha_i(t) (\nabla \phi_i(\mathbf{z}_i(t)) + \epsilon_i(t))],$$

where $\beta \in (0, 1)$ is a mixing parameter, $\alpha_i(t)$ denotes the stepsize of agent i at time t , $\omega_i(t)$ denotes a random variable,

$\nabla\phi_i(\mathbf{z}_i(t))$ denotes subgradient of function at $\mathbf{x} = \mathbf{z}_i(t)$, and $\boldsymbol{\epsilon}_i(t)$ is a stochastic subgradient error of agent i at time t . Moreover, let $\omega_i(t)$ be independent of each other and be independent from the random broadcast process. Note that $\omega_i(t)$ is drawn from the index set Θ_i . $P_{\mathcal{X}_i^{\omega_i(t)}}$ denotes the Euclidean projection operation on the constraint set $\mathcal{X}_i^{\omega_i(t)}$.

In order to analyze conveniently, we first define the following matrix:

$$Q(t) = I - \beta \sum_{j \in J_t} (e_j e_j^\top - e_j e_{i_t}^\top), \quad (4)$$

where I is the identity matrix of size $n \times n$ and $e_j \in \mathbb{R}^n$ denotes a vector with j th entry being 1 and the other entries being 0. From Lemma 2 in [26], each random matrix $Q(t)$ is not doubly stochastic, but the expectation of matrix $Q(t)$ is doubly stochastic. Let $D(t) \triangleq Q(t) - (1/n)\mathbf{1}\mathbf{1}^\top Q(t)$. We note that the matrices $D(t)$ are i.i.d. and $\lambda \triangleq \lambda_1(\mathbb{E}[D(t)^\top D(t)]) < 1$, where $\lambda_1(A)$ denotes the largest eigenvalue of a symmetric matrix A . Hence, by (4), relations (2)-(3) can be rewritten as follows:

$$\begin{aligned} \mathbf{z}_i(t) &= \sum_{j=1}^n [Q(t)]_{ij} \mathbf{x}_j(t-1) \\ \mathbf{x}_i(t) &= \mathbf{z}_i(t) - \mathbf{z}_i(t) \chi_{\{i \in J_t\}} \\ &\quad + P_{\mathcal{X}_i^{\omega_i(t)}} [\mathbf{z}_i(t) - \alpha_i(t) (\nabla\phi_i(\mathbf{z}_i(t)) + \boldsymbol{\epsilon}_i(t))] \chi_{\{i \in J_t\}}, \end{aligned} \quad (5)$$

where $\chi_{\{i \in J_t\}} = 1$ if $\{i \in J_t\}$ occurs, or else $\chi_{\{i \in J_t\}} = 0$.

2.2. Assumptions. In order to analyze the convergence properties of algorithm (5), we need to give some standard assumptions as follows.

Assumption 1. The network topology $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$ is connected and without self-loops at time t . If $\{i, j\} \in \mathcal{E}(t)$, then $p_{ij} > 0$. If $\{i, j\} \notin \mathcal{E}(t)$, then $p_{ij} = 0$. Furthermore, the link failure process is independent and identically distribution.

Assumption 2. We assume that the constraint sets $\mathcal{X}_i^j$ are nonempty closed and convex for all $i \in \mathcal{V}$. We also assume that the cost function $\phi_i : \mathbb{R}^d \rightarrow \mathbb{R}$ is convex. Moreover, we assume that the subgradient $\nabla\phi_i(\mathbf{x})$ of $\phi_i(\mathbf{x})$ is uniformly bounded over $\mathcal{X}$ for every $i \in \mathcal{V}$; namely, $\|\nabla\phi_i(\mathbf{x})\| \leq L_{\max}$ for all $\mathbf{x} \in \mathcal{X}$.

Assumption 3. For any random variable $\omega_i(t) \in \Theta_i$, $i \in \mathcal{V}$, and for all $\mathbf{x} \in \mathbb{R}^d$, we assume that the following relation hold:

$$\rho^2(\mathbf{x}, \mathcal{X}) \leq c_i \mathbb{E} [\rho^2(\mathbf{x}, \mathcal{X}_i^{\omega_i(t)})], \quad (6)$$

where c_i is a positive constant and $\rho(\mathbf{x}, \mathcal{X})$ denotes the distance between a point $\mathbf{x}$ and a set $\mathcal{X}$.

In Assumption 3, random variable $\omega_i(t)$ obeys a positive probability distribution. Moreover, if the set $\mathcal{X}$ has a

nonempty interior, then Assumption 3 can be seen to hold [27].

Let $\mathcal{F}_t$ denote all the information generated by the entire history of algorithm (5). Next, we give the assumption for stochastic subgradient error $\boldsymbol{\epsilon}_i(t)$ as follows.

Assumption 4. For each agent i at time t , we assume that the error $\boldsymbol{\epsilon}_i(t)$ satisfies

$$\begin{aligned} \mathbb{E} [\boldsymbol{\epsilon}_i(t) \mid \mathcal{F}_{t-1}, I_t, J_t] &= 0, \\ \mathbb{E} [\|\boldsymbol{\epsilon}_i(t)\| \mid \mathcal{F}_{t-1}, I_t, J_t] &\leq \nu \end{aligned} \quad (7)$$

with probability 1, where ν is a positive constant.

In this section, we propose an asynchronous distributed subgradient projection algorithm based on random projection operation and randomized broadcast communication protocol. Moreover, we also provide some standard assumptions to analyze the convergence properties of the algorithm. We will present main results of this paper in Section 3.

3. Main Results

In this section, we provide the convergence properties of algorithm (5). The detailed proofs of main results are given in next section.

We first define the optimal value and optimal solutions set of problem (1) as follows:

$$\begin{aligned} \phi^* &= \min_{\mathbf{x} \in \mathcal{X}} \phi(\mathbf{x}), \\ \mathcal{X}^* &= \{\mathbf{x} \in \mathcal{X} \mid \phi(\mathbf{x}) = \phi^*\}. \end{aligned} \quad (8)$$

The first result states that our proposed algorithm is convergent with probability 1.

Theorem 5. Under Assumptions 1–4, let the set of optimal point $\mathcal{X}^*$ be nonempty. Let estimate sequence $\{\mathbf{x}_i(t)\}$, $i = 1, \dots, n$, be generated by algorithm (5) with positive stepsize $\alpha_i(t) = 1/\Phi_i(t)$, where $\Phi_i(t)$ is the update number of agent i until time t . Then, the estimates of all agents converge to some optimal points $\mathbf{x}^* \in \mathcal{X}^*$ with probability 1.

Theorem 5 shows that the iterations of all agents asymptotic converge to some optimal points over time-varying networks; that is, for all $i \in \mathcal{V}$, $\lim_{t \rightarrow \infty} \mathbf{x}_i(t) = \mathbf{x}^*$ with probability 1.

We also establish asymptotic error bound between some optimal points in the optimal set and the estimates of algorithm (5).

Theorem 6. Under Assumptions 1–4, let estimate sequence $\{\mathbf{x}_i(t)\}$, $i = 1, \dots, n$, be generated by algorithm (5) and let $\alpha_i(t) = \alpha_i$ for $t \geq 1$. Moreover, assume that each function $\phi_i(\mathbf{x})$ is δ_i -strongly convex, where the constant δ_i satisfies $0 < 2\alpha_i\delta_i < 1$. Furthermore, let the set $\mathcal{X}$ be compact. Then, one has with probability 1

(a)

$$\begin{aligned}
& \limsup_{t \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2] \\
& \leq \frac{\Delta_{\gamma\alpha} L_{\max} L_{\mathcal{X}} + 2(1+c) \gamma_{\max} \alpha_{\max}^2 L_{\max}^2}{\min_i \{\gamma_i \delta_i \alpha_i\}} \\
& \quad + \frac{\gamma_{\max} \alpha_{\max}^2 L_{\max}}{(1-\sqrt{\lambda}) \min_i \{\gamma_i \delta_i \alpha_i\}} \\
& \quad \times \sqrt{2d \left((L_{\max} + \nu)^2 + \frac{8(1+c) \gamma_{\max} L_{\max}^2}{\min_i \{\gamma_i \delta_i \alpha_i\}} \right)},
\end{aligned} \tag{9}$$

(b)

$$\begin{aligned}
& \limsup_{t \rightarrow \infty} \mathbb{E} [\phi(\tilde{\mathbf{x}}_i(T)) - \phi^*] \\
& \leq \frac{n \gamma_{\max} \alpha_{\max}^2 L_{\max}}{(1-\sqrt{\lambda}) \min_i \{\gamma_i \alpha_i\}} \\
& \quad \times \sqrt{2d \left((L_{\max} + \nu)^2 + \frac{8(1+c) \gamma_{\max} L_{\max}^2}{\min_i \{\gamma_i \delta_i \alpha_i\}} \right)} \\
& \quad + \frac{n \Delta_{\gamma\alpha} L_{\max} L_{\mathcal{X}} + 2n(1+c) \gamma_{\max} \alpha_{\max}^2 L_{\max}^2}{\min_i \{\gamma_i \alpha_i\}},
\end{aligned} \tag{10}$$

where $\tilde{\mathbf{x}}_i(T) = (1/T) \sum_{t=1}^T \mathbf{x}_i(t)$, d is dimension of vector, $\gamma_i = (1/n) \sum_{j \in \mathcal{N}_i(t)} p_{ij}$ for all $i \in \mathcal{V}$, $\gamma_{\max} = \max_i \gamma_i$, $\alpha_{\max} = \max_i \alpha_i$, $L_{\mathcal{X}} = \max_{\mathbf{x}, \mathbf{y}} \|\mathbf{x} - \mathbf{y}\|$, and $\Delta_{\gamma\alpha} = \max_i \gamma_i \alpha_i - \min_i \gamma_i \alpha_i$.

Theorem 6 establishes asymptotic error bound, which is defined as the average of expected distances between some optimal points $\mathbf{x}^* \in \mathcal{X}^*$ and the estimates of algorithm (5). The asymptotic error is defined as the difference between the global cost function at $\tilde{\mathbf{x}}_i(T)$ and the global cost function at optimal point $\mathbf{x}^*$.

In this section, we provide the main results of this paper. The detailed proofs of main results are given in Sections 4 and 5.

4. Analysis of Convergence Results

In this section, we provide the proof of Theorem 5. For this purpose, we first establish a basic iterate relation for the estimates of algorithm (5).

Lemma 7. Under Assumptions 1–4, the estimate sequence $\{\mathbf{x}_i(t)\}$, $i \in \mathcal{V}$, is generated by algorithm (5). Let $\alpha_i(t) = 1/\Phi_i(t)$ and $\gamma_i = (1/n) \sum_{j \in \mathcal{N}_i(t)} p_{ij}$ for all $i \in \mathcal{V}$. Then, for any constant $q \in (0, 1/2)$ and all vectors $\mathbf{u} \in \mathcal{X}$, one has with probability 1

$$\begin{aligned}
& \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{u}\|^2 \mid \mathcal{F}_{t-1}] \\
& \leq - \frac{2\mathbb{E} [\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u}) \mid \mathcal{F}_{t-1}]}{t}
\end{aligned}$$

$$\begin{aligned}
& - \frac{\gamma_i \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}]}{2c} \\
& + \frac{2\gamma_i \mathbb{E} [\|\mathbf{y}_i(t) - \mathbf{u}\|^2 \mid \mathcal{F}_{t-1}]}{t^{3/2-q} p_{\min}^2} \\
& + \mathbb{E} [\|\mathbf{z}_i(t) - \mathbf{u}\|^2 \mid \mathcal{F}_{t-1}] + \frac{2L_{\max}^2}{t^{3/2-q} p_{\min}^2} \\
& + \frac{16n^2(1+c)L_{\max}^2}{t^2 p_{\min}^2}
\end{aligned} \tag{11}$$

for $t \geq \hat{t}$, where $\mathbf{y}_i(t) = P_{\mathcal{X}}[\mathbf{z}_i(t)]$, $p_{\min} = \min_{(i,j)} p_{ij}$, and $\hat{t}$ is a sufficiently large positive constant.

Proof. Let $\mathbf{u} \in \mathcal{X}$. Following from the nonexpansive property [16], we have

$$\begin{aligned}
& \|\mathbf{x}_i(t) - \mathbf{u}\|^2 \\
& \leq \|\mathbf{z}_i(t) - \mathbf{u}\|^2 - \|\mathbf{x}_i(t) - \mathbf{z}_i(t)\|^2 \\
& \quad - 2\alpha_i(t) \langle \nabla \phi_i(\mathbf{z}_i(t)) + \boldsymbol{\epsilon}_i(t), \mathbf{z}_i(t) - \mathbf{u} \rangle \\
& \quad + 2\alpha_i(t) \langle \nabla \phi_i(\mathbf{z}_i(t)) + \boldsymbol{\epsilon}_i(t), \mathbf{z}_i(t) - \mathbf{x}_i(t) \rangle.
\end{aligned} \tag{12}$$

Further, from the definition of inner product and Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}
& 2\alpha_i(t) \langle \nabla \phi_i(\mathbf{z}_i(t)), \mathbf{z}_i(t) - \mathbf{x}_i(t) \rangle \\
& = 2\alpha_i(t) \nabla \phi_i(\mathbf{z}_i(t))^\top (\mathbf{z}_i(t) - \mathbf{x}_i(t)) \\
& \leq 2\alpha_i(t) \|\nabla \phi_i(\mathbf{z}_i(t))\| \|\mathbf{z}_i(t) - \mathbf{x}_i(t)\| \\
& = 2(2\alpha_i(t) \|\nabla \phi_i(\mathbf{z}_i(t))\|) \left(\frac{\|\mathbf{z}_i(t) - \mathbf{x}_i(t)\|}{2} \right) \\
& \leq 4\alpha_i^2(t) \|\nabla \phi_i(\mathbf{z}_i(t))\|^2 + \frac{\|\mathbf{z}_i(t) - \mathbf{x}_i(t)\|^2}{4} \\
& \leq 4\alpha_i^2(t) L_{\max}^2 + \frac{1}{4} \|\mathbf{z}_i(t) - \mathbf{x}_i(t)\|^2,
\end{aligned} \tag{13}$$

where the second inequality follows from the inequality $2ab \leq a^2 + b^2$. Hence, following from relations (12) and (13), we obtain

$$\begin{aligned}
& \|\mathbf{x}_i(t) - \mathbf{u}\|^2 \leq \|\mathbf{z}_i(t) - \mathbf{u}\|^2 \\
& \quad - 2\alpha_i(t) (\phi_i(\mathbf{z}_i(t)) - \phi_i(\mathbf{u})) \\
& \quad - \frac{3}{4} \|\mathbf{x}_i(t) - \mathbf{z}_i(t)\|^2 + 4\alpha_i^2(t) L_{\max}^2 \\
& \quad + 2\alpha_i(t) \boldsymbol{\epsilon}_i(t)^\top (\mathbf{z}_i(t) - \mathbf{x}_i(t)) \\
& \quad - 2\alpha_i(t) \boldsymbol{\epsilon}_i(t)^\top (\mathbf{z}_i(t) - \mathbf{u}).
\end{aligned} \tag{14}$$

In addition, we also have

$$\begin{aligned}
& -2\alpha_i(t) (\phi_i(\mathbf{z}_i(t)) - \phi_i(\mathbf{u})) \\
& \leq 2\alpha_i(t) L_{\max} \|\mathbf{z}_i(t) - \mathbf{y}_i(t)\| \\
& \quad - 2\alpha_i(t) (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \leq 4\eta\alpha_i^2(t) L_{\max}^2 + \frac{1}{4\eta} \|\mathbf{z}_i(t) - \mathbf{y}_i(t)\|^2 \\
& \quad - 2\alpha_i(t) (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})),
\end{aligned} \tag{15}$$

where the last inequality follows from the inequality $2|a||b| \leq \tau a^2 + b^2/\tau$ for some $\tau > 0$ by letting $a = L_{\max}\alpha_i(t)$, $b = \|\mathbf{z}_i(t) - \mathbf{y}_i(t)\|$, and $\tau = 4\eta$ ($\eta > 0$). Hence, combining relations (14) and (15), we obtain

$$\begin{aligned}
\|\mathbf{x}_i(t) - \mathbf{u}\|^2 & \leq \|\mathbf{z}_i(t) - \mathbf{u}\|^2 \\
& \quad - 2\alpha_i(t) (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \quad - \frac{3}{4} \|\mathbf{x}_i(t) - \mathbf{z}_i(t)\|^2 \\
& \quad + \frac{1}{4\eta} \|\mathbf{z}_i(t) - \mathbf{y}_i(t)\|^2 \\
& \quad + 4\alpha_i^2(t) (1 + \eta) L_{\max}^2 \\
& \quad - 2\alpha_i(t) \boldsymbol{\epsilon}_i(t)^\top (\mathbf{z}_i(t) - \mathbf{u}) \\
& \quad + 2\alpha_i(t) \boldsymbol{\epsilon}_i(t)^\top (\mathbf{z}_i(t) - \mathbf{x}_i(t)).
\end{aligned} \tag{16}$$

Next, we consider the term $2\alpha_i(t)(\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u}))$ in (16), which can be rewritten as follows:

$$\begin{aligned}
& 2\alpha_i(t) (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& = \frac{2}{t\gamma_i} (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \quad + 2\alpha_i(t) (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \quad - \frac{2}{t\gamma_i} (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \geq \frac{2}{t\gamma_i} (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \quad - 2 \left| \alpha_i(t) - \frac{1}{t\gamma_i} \right| |\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})|.
\end{aligned} \tag{17}$$

Thus, according to Lemma 3 of [26], there exists a sufficient large positive constant $\hat{t}$, for any $t \geq \hat{t}$, we have

$$\begin{aligned}
& 2\alpha_i(t) (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \geq \frac{2}{t\gamma_i} (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \quad - 2 \frac{2}{t^{3/2-q} p_{\min}^2} L_{\max} \|\mathbf{y}_i(t) - \mathbf{u}\|
\end{aligned}$$

$$\begin{aligned}
& \geq \frac{2}{t\gamma_i} (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \quad - \frac{2}{t^{3/2-q} p_{\min}^2} (L_{\max}^2 + \|\mathbf{y}_i(t) - \mathbf{u}\|^2),
\end{aligned} \tag{18}$$

where we use the Cauchy-Schwarz inequality. Then, from (16) and (18), we obtain

$$\begin{aligned}
& \|\mathbf{x}_i(t) - \mathbf{u}\|^2 \\
& \leq \|\mathbf{z}_i(t) - \mathbf{u}\|^2 - \frac{2}{t\gamma_i} (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \quad - \frac{3}{4} \|\mathbf{x}_i(t) - \mathbf{z}_i(t)\|^2 + \frac{2}{t^{3/2-q} p_{\min}^2} \|\mathbf{y}_i(t) - \mathbf{u}\|^2 \\
& \quad + \frac{1}{4\eta} \|\mathbf{z}_i(t) - \mathbf{y}_i(t)\|^2 \\
& \quad + 2\alpha_i(t) \boldsymbol{\epsilon}_i(t)^\top (\mathbf{z}_i(t) - \mathbf{x}_i(t)) \\
& \quad - 2\alpha_i(t) \boldsymbol{\epsilon}_i(t)^\top (\mathbf{z}_i(t) - \mathbf{u}) \\
& \quad + \left(\frac{2}{t^{3/2-q} p_{\min}^2} + \frac{16n^2}{t^2 p_{\min}^2} (1 + \eta) \right) L_{\max}^2.
\end{aligned} \tag{19}$$

Note that $\|\mathbf{z}_i(t) - \mathbf{y}_i(t)\| = \rho(\mathbf{z}_i(t), \mathcal{X})$; we have

$$\begin{aligned}
\|\mathbf{x}_i(t) - \mathbf{z}_i(t)\| & \geq \|P_{\mathcal{X}_i^{\omega_i(t)}} [\mathbf{z}_i(t)] - \mathbf{z}_i(t)\| \\
& = \rho(\mathbf{z}_i(t), \mathcal{X}_i^{\omega_i(t)}).
\end{aligned} \tag{20}$$

Hence, taking conditional expectation in both sides of (19) and using the fact that the mean of stochastic errors $\boldsymbol{\epsilon}_i(t)$ is zero, we obtain that with probability 1 for all $t \geq \hat{t}$ and $i \in J_t$

$$\begin{aligned}
& \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{u}\|^2 \mid \mathcal{F}_{t-1}, I_t, J_t] \\
& \leq \|\mathbf{z}_i(t) - \mathbf{u}\|^2 - \frac{2}{t\gamma_i} (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\
& \quad - \frac{3}{4} \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}_i^{\omega_i(t)}) \mid \mathbf{z}_i(t)] \\
& \quad + \frac{2}{t^{3/2-q} p_{\min}^2} \|\mathbf{y}_i(t) - \mathbf{u}\|^2 + \frac{1}{4\eta} \rho^2(\mathbf{z}_i(t), \mathcal{X}) \\
& \quad + \left(\frac{2}{t^{3/2-q} p_{\min}^2} + \frac{16n^2}{t^2 p_{\min}^2} (1 + \eta) \right) L_{\max}^2.
\end{aligned} \tag{21}$$

By Assumption 3, we have

$$\mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}_i^{\omega_i(t)}) \mid \mathbf{z}_i(t)] \geq \frac{1}{c} \rho^2(\mathbf{z}_i(t), \mathcal{X}), \tag{22}$$

where $c = \max_i c_i$. Let $\eta = c$. For all $t \geq \hat{t}$, we obtain with probability 1

$$\begin{aligned} & \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{u}\|^2 \mid \mathcal{F}_{t-1}, I_t, J_t] \\ & \leq \|\mathbf{z}_i(t) - \mathbf{u}\|^2 - \frac{2}{t\gamma_i} (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\ & \quad - \frac{1}{2c} \rho^2(\mathbf{z}_i(t), \mathcal{X}) + \frac{2}{t^{3/2-q} p_{\min}^2} \|\mathbf{y}_i(t) - \mathbf{u}\|^2 \\ & \quad + \left(\frac{2}{t^{3/2-q} p_{\min}^2} + \frac{16n^2}{t^2 p_{\min}^2} (1+c) \right) L_{\max}^2. \end{aligned} \quad (23)$$

Note that $\mathbf{x}_i(t) = \mathbf{z}_i(t)$ if $i \notin J_t$. Besides, if $i \in J_t$, then the agent i updates its estimate with probability γ_i and does not update its estimate with probability $1 - \gamma_i$. Hence, following from (23), the desired result is obtained completely. $\square$

We next consider the errors $\mathbf{e}_i(t)$ for all $i \in \mathcal{V}$, which are caused by the random projection operation. The errors are defined as follows: $\mathbf{e}_i(t) = \mathbf{x}_i(t) - \mathbf{z}_i(t)$ for all $i \in \mathcal{V}$. Moreover, we have the following lemma.

Lemma 8. *Under Assumptions 1–4, for $t \geq 1$ and all $i \in \{1, \dots, n\}$, the relations $\lim_{t \rightarrow \infty} \rho(\mathbf{z}_i(t), \mathcal{X}) = 0$ and $\sum_{t=1}^{\infty} \mathbb{E}[\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] < \infty$ hold with probability 1. Furthermore, for all $i \in \{1, \dots, n\}$, one has that $\lim_{t \rightarrow \infty} \rho(\mathbf{z}_i(t), \mathcal{X}) = 0$ with probability 1.*

Proof. Let $\mathbf{u} = \mathbf{y}_i(t) = P_{\mathcal{X}}[\mathbf{z}_i(t)]$ in Lemma 7. For all $i \in \{1, \dots, n\}$, we obtain with probability 1

$$\begin{aligned} & \mathbb{E} [\|\mathbf{x}_i(t) - P_{\mathcal{X}}[\mathbf{z}_i(t)]\|^2 \mid \mathcal{F}_{t-1}] \\ & = \left(1 + \frac{2}{t^{3/2-q} p_{\min}^2} \right) \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] \\ & \quad - \frac{\gamma_i}{2c} \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] \\ & \quad + \left(\frac{2}{t^{3/2-q} p_{\min}^2} + \frac{16n^2}{t^2 p_{\min}^2} (1+c) \right) L_{\max}^2, \end{aligned} \quad (24)$$

where $t \geq \hat{t}$ with large enough $\hat{t}$ and $q \in (0, 1/2)$. Since the function $\rho(\mathbf{x}, \mathcal{X})$ is convex [28], we have

$$\sum_{i=1}^n \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] \leq \sum_{j=1}^n \rho^2(\mathbf{x}_j(t-1), \mathcal{X}). \quad (25)$$

Combining (24), (25), and the definition of projection operation, we obtain for all $t \geq \hat{t}$ with probability 1

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - P_{\mathcal{X}}[\mathbf{z}_i(t)]\|^2 \mid \mathcal{F}_{t-1}] \\ & = \left(1 + \frac{2}{t^{3/2-q} p_{\min}^2} \right) \sum_{j=1}^n \rho^2(\mathbf{x}_j(t-1), \mathcal{X}) \end{aligned}$$

$$\begin{aligned} & - \frac{\gamma_{\min}}{2c} \sum_{i=1}^n \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] \\ & + nL_{\max}^2 \left(\frac{2}{t^{3/2-q} p_{\min}^2} + \frac{16n^2}{t^2 p_{\min}^2} (1+c) \right), \end{aligned} \quad (26)$$

where $\gamma_{\min} = \min_i \gamma_i$. Hence, following from the Supermartingale Convergence Theorem, which is stated in [29], we obtain $\sum_{t=1}^{\infty} \mathbb{E}[\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] < \infty$. Further, taking expectation on $\mathcal{F}_t$ in (26), we have that $\sum_{t=1}^{\infty} \mathbb{E}[\rho^2(\mathbf{z}_i(t), \mathcal{X})] < \infty$ for all $i = 1, \dots, n$. By Monotone Convergence Theorem [30], we obtain $i \in \{1, \dots, n\}$; the relations $\lim_{t \rightarrow \infty} \rho(\mathbf{z}_i(t), \mathcal{X}) = 0$.

Further, for all $i \in J_t$, following from the definition of error $\mathbf{e}_i(t)$, we have

$$\begin{aligned} \|\mathbf{e}_i(t)\| & \leq \|\mathbf{x}_i(t) - \mathbf{y}_i(t)\| + \|\mathbf{y}_i(t) - \mathbf{z}_i(t)\| \\ & = \left\| P_{\mathcal{X}^{\omega_i(t)}} [\mathbf{z}_i(t) - \alpha_i(t) (\nabla \phi_i(\mathbf{z}_i(t)) + \mathbf{e}_i(t))] \right. \\ & \quad \left. - \mathbf{y}_i(t) \right\| + \|\mathbf{y}_i(t) - \mathbf{z}_i(t)\|. \end{aligned} \quad (27)$$

Moreover, note that $\mathbf{y}_i(t) \in \mathcal{X}_i^{\omega_i(t)}$. Hence, we have

$$\begin{aligned} \|\mathbf{e}_i(t)\| & \leq \|\mathbf{z}_i(t) - \alpha_i(t) (\nabla \phi_i(\mathbf{z}_i(t)) + \mathbf{e}_i(t)) - \mathbf{y}_i(t)\| \\ & \quad + \|\mathbf{y}_i(t) - \mathbf{z}_i(t)\| \\ & \leq 2 \|\mathbf{z}_i(t) - \mathbf{y}_i(t)\| \\ & \quad + \alpha_i(t) \|\nabla \phi_i(\mathbf{z}_i(t)) + \mathbf{e}_i(t)\| \\ & = 2\rho(\mathbf{z}_i(t), \mathcal{X}) + \alpha_i(t) \|\nabla \phi_i(\mathbf{z}_i(t)) + \mathbf{e}_i(t)\|, \end{aligned} \quad (28)$$

where we use the nonexpansive property of the projection. Furthermore, following from the inequality $(a+b)^2 \leq 2(a^2 + b^2)$, we obtain

$$\begin{aligned} \|\mathbf{e}_i(t)\|^2 & \leq 8\rho^2(\mathbf{z}_i(t), \mathcal{X}) \\ & \quad + 2\alpha_i^2(t) \|\nabla \phi_i(\mathbf{z}_i(t)) + \mathbf{e}_i(t)\|^2 \\ & \leq 8\rho^2(\mathbf{z}_i(t), \mathcal{X}) \\ & \quad + \frac{8n^2}{t^2 p_{\min}^2} \|\nabla \phi_i(\mathbf{z}_i(t)) + \mathbf{e}_i(t)\|^2, \end{aligned} \quad (29)$$

where the last inequality follows from Lemma 3 in [26]. Taking the conditional expectation on both sides of (29), we have

$$\begin{aligned} & \mathbb{E} [\|\mathbf{e}_i(t)\|^2 \mid \mathcal{F}_{t-1}] \\ & \leq 8\gamma_i \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] \\ & \quad + \frac{8\gamma_i n^2}{t^2 p_{\min}^2} \mathbb{E} [\|\nabla \phi_i(\mathbf{z}_i(t)) + \mathbf{e}_i(t)\|^2 \mid \mathcal{F}_{t-1}] \end{aligned}$$

$$\begin{aligned} &\leq 8\gamma_i \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] \\ &\quad + \frac{8\gamma_i n^2}{t^2 p_{\min}^2} (L_{\max} + \nu)^2. \end{aligned} \quad (30)$$

Following from the fact that $\sum_{t=1}^{\infty} \mathbb{E}[\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] < \infty$ and $\sum_{t=1}^{\infty} (1/t^2) < \infty$, we obtain with probability 1 $\sum_{t=1}^{\infty} \mathbb{E}[\|\mathbf{e}_i(t)\|^2 \mid \mathcal{F}_{t-1}] < \infty$ for all $i = 1, \dots, n$. Further, we also have $\lim_{t \rightarrow \infty} \|\mathbf{e}_i(t)\| = 0$ with probability 1. Therefore, the conclusions of this lemma are proved completely. $\square$

We also establish the following lemma to prove the main results of this paper.

Lemma 9. *Under Assumptions 1–4, let $\alpha_i(t) = 1/\Phi_i(t)$. Assume that $\{\mathbf{z}_i(t)\}$ are generated by algorithm (5). Then, one finds $\sum_{t=1}^{\infty} (1/t) \mathbb{E}[\|\mathbf{z}_i(t) - \bar{\mathbf{z}}(t)\| \mid \mathcal{F}_{t-1}] < \infty$ with probability 1 for $i = 1, \dots, n$, where $\bar{\mathbf{z}}(t) = (1/n) \sum_{i=1}^n \mathbf{z}_i(t)$.*

Proof. We define the vector $\mathbf{v}_\ell(t)$ in $\mathbb{R}^n$ for $\ell = 1, \dots, d$ such that $[\mathbf{v}_\ell(t)]_i = [\mathbf{x}_i(t)]_\ell$ for $i = 1, \dots, n$. According to algorithm (5), we obtain

$$\begin{aligned} \mathbf{v}_\ell(t) &= Q(t) \mathbf{v}_\ell(t-1) + \xi_\ell(t) \\ &\quad \text{for } t \geq 1, \ell = 1, \dots, d, \end{aligned} \quad (31)$$

where $[\xi_\ell(t)]_i = 0$ for $i \notin J_t$, and if $i \in J_t$, then $[\xi_\ell(t)]_i = [P_{\mathcal{G}_i^{w_i(t)}}[\mathbf{z}_i(t) - \alpha_i(t)(\nabla\phi_i(\mathbf{z}_i(t)) + \mathbf{e}_i(t))] - \mathbf{z}_i(t)]_\ell$. Further, note that $[\bar{\mathbf{x}}(t)]_\ell = (1/n)\mathbf{1}^\top \mathbf{v}_\ell(t)$. Thus, by fixing arbitrary index $\ell = 1, \dots, d$, we obtain

$$\begin{aligned} \mathbf{v}_\ell(t) - [\bar{\mathbf{x}}(t)]_\ell \mathbf{1} &= \left(Q(t) - \frac{1}{n} \mathbf{1} \mathbf{1}^\top Q(t) \right) \mathbf{v}_\ell(t-1) \\ &\quad + \left(I - \frac{1}{n} \mathbf{1} \mathbf{1}^\top \right) \xi_\ell(t). \end{aligned} \quad (32)$$

Since $Q(t)\mathbf{1} = \mathbf{1}$, $(Q(t) - (1/n)\mathbf{1}\mathbf{1}^\top Q(t))\mathbf{1} = \mathbf{0}$. Thus, we have

$$\begin{aligned} \mathbf{v}_\ell(t) - [\bar{\mathbf{x}}(t)]_\ell \mathbf{1} &= D(t) (\mathbf{v}_\ell(t-1) - [\bar{\mathbf{x}}(t-1)]_\ell \mathbf{1}) \\ &\quad + H \xi_\ell(t), \end{aligned} \quad (33)$$

where $H = I - (1/n)\mathbf{1}\mathbf{1}^\top$ and we use the fact that $(Q(t) - (1/n)\mathbf{1}\mathbf{1}^\top Q(t))[\bar{\mathbf{x}}(t-1)]_\ell \mathbf{1} = \mathbf{0}$. Following from (33), we obtain with probability 1

$$\begin{aligned} &\mathbb{E} [\|\mathbf{v}_\ell(t) - [\bar{\mathbf{x}}(t)]_\ell \mathbf{1}\| \mid \mathcal{F}_{t-1}] \\ &\leq \mathbb{E} [\|H \xi_\ell(t)\| \mid \mathcal{F}_{t-1}] \\ &\quad + \mathbb{E} [\|D(t) (\mathbf{v}_\ell(t-1) - [\bar{\mathbf{x}}(t-1)]_\ell \mathbf{1})\| \mid \mathcal{F}_{t-1}]. \end{aligned} \quad (34)$$

We first compute the second term in right side of (34). According to the property of stochastic matrix $Q(t)$, we have

$$\begin{aligned} &\mathbb{E} [\|D(t) (\mathbf{v}_\ell(t-1) - [\bar{\mathbf{x}}(t-1)]_\ell \mathbf{1})\|^2 \mid \mathcal{F}_{t-1}] \\ &\leq \lambda \|\mathbf{v}_\ell(t-1) - [\bar{\mathbf{x}}(t-1)]_\ell \mathbf{1}\|^2. \end{aligned} \quad (35)$$

By using the inequality $\mathbb{E}[\|x\|] \leq \sqrt{\mathbb{E}[\|x\|^2]}$, we have for $t \geq 1$

$$\begin{aligned} &\mathbb{E} [\|D(t) (\mathbf{v}_\ell(t-1) - [\bar{\mathbf{x}}(t-1)]_\ell \mathbf{1})\| \mid \mathcal{F}_{t-1}] \\ &\leq \sqrt{\lambda} \|\mathbf{v}_\ell(t-1) - [\bar{\mathbf{x}}(t-1)]_\ell \mathbf{1}\|. \end{aligned} \quad (36)$$

In addition, we have $\|H \xi_\ell(t)\|^2 \leq \|\xi_\ell(t)\|^2$ by using $\|H\|^2 = \|I - (1/n)\mathbf{1}\mathbf{1}^\top\|^2 = 1$. Hence, we obtain that

$$\begin{aligned} &\|H \xi_\ell(t)\|^2 \\ &\leq \sum_{i=1}^n \left(\left\| P_{\mathcal{G}_i^{w_i(t)}} [\mathbf{z}_i(t) - \alpha_i(t) (\nabla\phi_i(\mathbf{z}_i(t)) + \mathbf{e}_i(t))] \right. \right. \\ &\quad \left. \left. - \mathbf{z}_i(t) \right\|^2 \right). \end{aligned} \quad (37)$$

Furthermore, from the nonexpansive property and $\alpha_i^2(t) \leq 4n^2/t^2 p_{\min}$ for $t \geq \hat{t}$ with large enough $\hat{t}$, we find

$$\begin{aligned} \|H \xi_\ell(t)\|^2 &\leq \sum_{i=1}^n \left(8\rho^2(\mathbf{z}_i(t), \mathcal{X}) \right. \\ &\quad \left. + \frac{8n^2}{t^2 p_{\min}^2} \|\nabla\phi_i(\mathbf{z}_i(t)) + \mathbf{e}_i(t)\|^2 \right). \end{aligned} \quad (38)$$

Thus, from (38), we have with probability 1

$$\begin{aligned} &\mathbb{E} [\|H \xi_\ell(t)\| \mid \mathcal{F}_{t-1}] \\ &\leq \sqrt{\frac{8n^3}{t^2 p_{\min}^2} (L_{\max} + \nu)^2 + 8 \sum_{i=1}^n \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}]} \end{aligned} \quad (39)$$

Hence, from inequalities (36) and (39), we obtain

$$\begin{aligned} &\mathbb{E} [\|\mathbf{v}_\ell(t) - [\bar{\mathbf{x}}(t)]_\ell \mathbf{1}\| \mid \mathcal{F}_{t-1}] \\ &\leq \sqrt{\lambda} \|\mathbf{v}_\ell(t-1) - [\bar{\mathbf{x}}(t-1)]_\ell \mathbf{1}\| \\ &\quad + \sqrt{\frac{8n^3}{t^2 p_{\min}^2} (L_{\max} + \nu)^2 + 8 \sum_{i=1}^n \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}]} \end{aligned} \quad (40)$$

Further, we obtain for $t \geq \hat{t}$

$$\begin{aligned} &\frac{1}{t} \mathbb{E} [\|\mathbf{v}_\ell(t) - [\bar{\mathbf{x}}(t)]_\ell \mathbf{1}\| \mid \mathcal{F}_{t-1}] \\ &\leq \frac{1}{t-1} \|\mathbf{v}_\ell(t-1) - [\bar{\mathbf{x}}(t-1)]_\ell \mathbf{1}\| \\ &\quad - \frac{1 - \sqrt{\lambda}}{t} \|\mathbf{v}_\ell(t-1) - [\bar{\mathbf{x}}(t-1)]_\ell \mathbf{1}\| + \frac{1}{t} \\ &\quad \cdot \sqrt{\frac{8n^3}{t^2 p_{\min}^2} (L_{\max} + \nu)^2 + 8 \sum_{i=1}^n \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}]} \end{aligned} \quad (41)$$

According to Lemma 8, the first term in (41) is convergent with probability 1. Moreover, since $1 - \sqrt{\lambda} > 0$ and then

following from Supermartingale Convergence Theorem [29], we have

$$\sum_{i=1}^n \frac{1}{t} \mathbb{E} [\|v_\ell(t) - [\bar{x}(t)]_\ell \mathbf{1}\| \mid \mathcal{F}_{t-1}] < \infty. \quad (42)$$

From the definition of $v_\ell(t)$, we obtain for $i \in \mathcal{V}$

$$\sum_{i=1}^n \frac{1}{t} \mathbb{E} [\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\| \mid \mathcal{F}_{t-1}] < \infty. \quad (43)$$

From the convexity of norm and the definition of $\bar{\mathbf{z}}(t)$, we have

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E} [\|\mathbf{z}_i(t) - \bar{\mathbf{z}}(t)\| \mid \mathcal{F}_{t-1}] \\ & \leq \sum_{j=1}^n \|\mathbf{x}_j(t-1) - \bar{\mathbf{x}}(t-1)\|. \end{aligned} \quad (44)$$

Thus, following from (43) and (44), we can see that

$$\sum_{i=1}^n \frac{1}{t} \mathbb{E} [\|\mathbf{z}_i(t) - \bar{\mathbf{z}}(t)\| \mid \mathcal{F}_{t-1}] < \infty \quad (45)$$

for all $i \in \mathcal{V}$ with probability 1. Therefore, the lemma is proved completely. $\square$

By using Lemmas 7, 8, and 9, we start to prove Theorem 5.

Proof of Theorem 5. From Lemma 7, we obtain

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{u}\|^2 \mid \mathcal{F}_{t-1}] \\ & \leq \left(1 + \frac{2}{t^{3/2-q} p_{\min}^2}\right) \sum_{j=1}^n \|\mathbf{x}_j(t-1) - \mathbf{u}\|^2 \\ & \quad - \frac{2}{t} \sum_{i=1}^n \mathbb{E} [\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u}) \mid \mathcal{F}_{t-1}] \\ & \quad + nL_{\max}^2 \left(\frac{2}{t^{3/2-q} p_{\min}^2} + \frac{16n^2}{t^2 p_{\min}^2} (1+c) \right). \end{aligned} \quad (46)$$

Letting $\bar{\mathbf{y}}(t) \triangleq (1/n) \sum_{i=1}^n \mathbf{y}_i(t)$, then $\bar{\mathbf{y}}(t) \in \mathcal{X}$. Therefore, we have

$$\begin{aligned} & \sum_{i=1}^n (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\ & = \sum_{i=1}^n (\phi_i(\mathbf{y}_i(t)) - \phi_i(\bar{\mathbf{y}}(t))) + (\phi(\bar{\mathbf{y}}(t)) - \phi(\mathbf{u})). \end{aligned} \quad (47)$$

By the convexity of each function ϕ_i and the boundedness of each subgradient norm $\|\nabla \phi_i\|$, we have

$$\begin{aligned} & \sum_{i=1}^n (\phi_i(\mathbf{y}_i(t)) - \phi_i(\bar{\mathbf{y}}(t))) \\ & \geq \sum_{i=1}^n \nabla \phi_i(\bar{\mathbf{y}}(t)) (\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)) \end{aligned}$$

$$\geq - \sum_{i=1}^n \|\nabla \phi_i(\bar{\mathbf{y}}(t))\| \|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|$$

$$\geq -L_{\max} \sum_{i=1}^n \|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|. \quad (48)$$

Furthermore, according to the convexity of squared-norm, we obtain

$$\begin{aligned} \|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\| &= \left\| \frac{1}{n} \sum_{k=1}^n (\mathbf{y}_i(t) - \mathbf{y}_k(t)) \right\| \\ &\leq \frac{1}{n} \sum_{k=1}^n \|\mathbf{y}_i(t) - \mathbf{y}_k(t)\| \\ &\leq \frac{1}{n} \sum_{k=1}^n \|\mathbf{z}_i(t) - \mathbf{z}_k(t)\|, \end{aligned} \quad (49)$$

where the last inequality uses the nonexpansive property.

Let $\bar{\mathbf{z}}(t) = (1/n) \sum_{i=1}^n \mathbf{z}_i(t)$. Moreover, we also have

$$\|\mathbf{z}_i(t) - \mathbf{z}_k(t)\| \leq \|\mathbf{z}_i(t) - \bar{\mathbf{z}}(t)\| + \|\mathbf{z}_k(t) - \bar{\mathbf{z}}(t)\|. \quad (50)$$

Hence, combining inequalities (49) and (50) and then summing over i from 1 to n , we have

$$\sum_{i=1}^n \|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\| \leq 2 \sum_{i=1}^n \|\mathbf{z}_i(t) - \bar{\mathbf{z}}(t)\|. \quad (51)$$

Thus, from (47), (48), and (51), we obtain

$$\begin{aligned} & \sum_{i=1}^n (\phi_i(\mathbf{y}_i(t)) - \phi_i(\mathbf{u})) \\ & \geq -2L_{\max} \sum_{i=1}^n \|\mathbf{z}_i(t) - \bar{\mathbf{z}}(t)\| + (\phi(\bar{\mathbf{y}}(t)) - \phi(\mathbf{u})). \end{aligned} \quad (52)$$

Let $\mathbf{u} = \mathbf{x}^*$. Following from inequalities (46) and (52), we have with probability 1

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2 \mid \mathcal{F}_{t-1}] \\ & \leq \left(1 + \frac{2}{t^{3/2-q} p_{\min}^2}\right) \sum_{j=1}^n \|\mathbf{x}_j(t-1) - \mathbf{x}^*\|^2 \\ & \quad - \frac{2}{t} \mathbb{E} [\phi(\bar{\mathbf{y}}(t)) - \phi^* \mid \mathcal{F}_{t-1}] \\ & \quad + \frac{4L_{\max}}{t} \sum_{i=1}^n \mathbb{E} [\|\mathbf{z}_i(t) - \bar{\mathbf{z}}(t)\| \mid \mathcal{F}_{t-1}] \\ & \quad + nL_{\max}^2 \left(\frac{2}{t^{3/2-q} p_{\min}^2} + \frac{16n^2}{t^2 p_{\min}^2} (1+c) \right). \end{aligned} \quad (53)$$

Moreover, note that $\phi(\bar{\mathbf{y}}(t)) - \phi^* \geq 0$. Hence, by using Supermartingale Convergence Theorem [29] and Lemma 9,

for all $i \in \mathcal{V}$, we find that $\{\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2\}$ is convergent with probability 1. Moreover, $\sum_{t=1}^{\infty} (1/t)(\phi(\bar{\mathbf{y}}(t)) - \phi^*) < \infty$ hold with probability 1. Further, noting that $\sum_{t=1}^{\infty} (1/t) = \infty$, we have with probability 1

$$\liminf_{t \rightarrow \infty} (\phi(\bar{\mathbf{y}}(t)) - \phi^*) = 0. \quad (54)$$

Hence, from the definition of $\mathbf{y}_i(t)$ and Lemma 8, we obtain with probability 1

$$\lim_{t \rightarrow \infty} \|\mathbf{z}_i(t) - \mathbf{y}_i(t)\| = 0 \quad \forall i \in J_t. \quad (55)$$

Since $\{\|\mathbf{x}_i(t) - \mathbf{x}^*\|\}$ is convergent with probability 1, $\{\|\mathbf{z}_i(t) - \mathbf{x}^*\|\}$ and $\{\|\mathbf{y}_i(t) - \mathbf{x}^*\|\}$ are also convergent by using (5) and (55). Therefore, $\{\|\bar{\mathbf{z}}(t) - \mathbf{x}^*\|\}$ and $\{\|\bar{\mathbf{y}}(t) - \mathbf{x}^*\|\}$ are convergent with probability 1. According to (54), we obtain with probability 1 $\lim_{t \rightarrow \infty} \bar{\mathbf{y}}(t) = \mathbf{x}^*$. We further have $\lim_{t \rightarrow \infty} \|\bar{\mathbf{z}}(t) - \bar{\mathbf{y}}(t)\| \leq (1/n) \sum_{k=1}^n \lim_{t \rightarrow \infty} \|\mathbf{z}_k(t) - \mathbf{y}_k(t)\| = 0$. Hence, we obtain with probability 1 $\lim_{t \rightarrow \infty} \bar{\mathbf{z}}(t) = \mathbf{x}^*$. Following from Lemma 9, we have with probability 1 $\liminf_{t \rightarrow \infty} \|\mathbf{z}_i(t) - \bar{\mathbf{z}}(t)\| = 0$. Therefore, we obtain $\lim_{t \rightarrow \infty} \|\mathbf{z}_i(t) - \mathbf{x}^*\| = 0$ with probability 1. Moreover, according to Lemma 9, the statement of this theorem is obtained completely. $\square$

In this section, we have given a proof of Theorem 5. From this theorem, we can see that the proposed algorithm is almost sure convergent. In next section, we will analyze the error bounds of the proposed algorithm.

5. Error Bounds Analysis

In this section, we give the proof of Theorem 6, where we assume that the cost function ϕ_i is strongly convex with constant δ_i and the stepsize $\alpha_i(t)$ is constant for all $i \in \mathcal{V}$; that is, $\alpha_i(t) = \alpha_i$. To prove Theorem 6, we first establish a basic iterate relation for a constant stepsize as follows.

Lemma 10. *Under Assumptions 1–4, let the estimate sequence $\{\mathbf{x}_i(t)\}$, $i \in \mathcal{V}$, be generated by algorithm (5). Moreover, $\alpha_i(t) = \alpha_i$ for all $i = 1, \dots, n$, assuming that each function $\phi_i(\mathbf{x})$ is δ_i -strongly convex, where the constant δ_i satisfies $0 < 2\alpha_i\delta_i < 1$. Then, with probability 1, the following relation holds for all $i \in J_t$ and $t \geq 1$:*

$$\begin{aligned} & \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{u}\|^2 \mid \mathcal{F}_{t-1}, I_t, J_t] \\ & \leq (1 - 2\delta_i\alpha_i) \|\mathbf{z}_i(t) - \mathbf{u}\|^2 \\ & \quad - 2\alpha_i \nabla \phi_i(\mathbf{u})^\top (\mathbf{y}_i(t) - \mathbf{u}) + 4(1+c)\alpha_i^2 L_{\max}^2. \end{aligned} \quad (56)$$

Proof. Since $\mathbf{x}_i(t) \in \mathcal{X}_i^{\omega_i(t)}$, we have

$$\begin{aligned} & \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{u}\|^2 \mid \mathcal{F}_{t-1}, I_t, J_t] \\ & \leq \|\mathbf{z}_i(t) - \mathbf{u}\|^2 - \frac{3}{4} \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}_i^{\omega_i(t)}) \mid \mathbf{z}_i(t)] \\ & \quad - 2\alpha_i \nabla \phi_i(\mathbf{z}_i(t))^\top (\mathbf{z}_i(t) - \mathbf{u}) + 4\alpha_i^2 L_{\max}^2. \end{aligned} \quad (57)$$

Moreover, we also have

$$\begin{aligned} & \nabla \phi_i(\mathbf{z}_i(t))^\top (\mathbf{z}_i(t) - \mathbf{u}) \\ & \geq \nabla \phi_i(\mathbf{u})^\top (\mathbf{z}_i(t) - \mathbf{u}) + \delta_i \|\mathbf{z}_i(t) - \mathbf{u}\|^2, \end{aligned} \quad (58)$$

where the inequality follows from the fact that the function ϕ_i is strongly convex. Further, by using Assumption 2, we obtain

$$\begin{aligned} & \nabla \phi_i(\mathbf{u})^\top (\mathbf{z}_i(t) - \mathbf{u}) \geq \nabla \phi_i(\mathbf{u})^\top (\mathbf{y}_i(t) - \mathbf{u}) \\ & \quad - L_{\max} \|\mathbf{z}_i(t) - \mathbf{y}_i(t)\|. \end{aligned} \quad (59)$$

Combining (57), (58), and (59), we obtain

$$\begin{aligned} & \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{u}\|^2 \mid \mathcal{F}_{t-1}, I_t, J_t] \\ & \leq (1 - 2\delta_i\alpha_i) \|\mathbf{z}_i(t) - \mathbf{u}\|^2 \\ & \quad - \frac{3}{4} \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}_i^{\omega_i(t)}) \mid \mathbf{z}_i(t)] + 4\alpha_i^2 L_{\max}^2 \\ & \quad - 2\alpha_i \nabla \phi_i(\mathbf{u})^\top (\mathbf{y}_i(t) - \mathbf{u}) \\ & \quad + 2\alpha_i L_{\max} \rho(\mathbf{z}_i(t), \mathcal{X}) \\ & \leq (1 - 2\delta_i\alpha_i) \|\mathbf{z}_i(t) - \mathbf{u}\|^2 \\ & \quad - 2\alpha_i \nabla \phi_i(\mathbf{u})^\top (\mathbf{y}_i(t) - \mathbf{u}) \\ & \quad - \frac{3}{4} \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}_i^{\omega_i(t)}) \mid \mathbf{z}_i(t)] + 4\alpha_i^2 L_{\max}^2 \\ & \quad + 4c\alpha_i^2 L_{\max}^2 + \frac{1}{4c} \rho^2(\mathbf{z}_i(t), \mathcal{X}). \end{aligned} \quad (60)$$

By Assumption 3, we see that

$$\rho^2(\mathbf{z}_i(t), \mathcal{X}) \leq c \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}_i^{\omega_i(t)}) \mid \mathcal{F}_{t-1}, I_t, J_t] \quad (61)$$

Following from inequality (61), we have

$$\begin{aligned} & - \frac{3}{4} \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X}_i^{\omega_i(t)}) \mid \mathbf{z}_i(t)] + \frac{1}{4c} \rho^2(\mathbf{z}_i(t), \mathcal{X}) \\ & \leq - \frac{1}{2c} \rho^2(\mathbf{z}_i(t), \mathcal{X}) \leq 0. \end{aligned} \quad (62)$$

From (60) and (62), the conclusion of this lemma is obtained. $\square$

Further, we also establish the following lemma.

Lemma 11. *Under Assumptions 1–4, let the estimate sequence $\{\mathbf{x}_i(t)\}$, $i \in \mathcal{V}$, be generated by algorithm (5). Let $\alpha_i(t) = \alpha_i$ and assume that each function ϕ_i is δ_i -strongly convex, where the constant δ_i satisfies $0 < 2\alpha_i\delta_i < 1$. Then, we have with probability 1*

$$\begin{aligned} & (a) \\ & \limsup_{t \rightarrow \infty} \sum_{i=1}^n \mathbb{E} [\rho^2(\mathbf{x}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] \\ & \leq \frac{2(1+c)nL_{\max}^2 \max_i \{\gamma_i \alpha_i^2\}}{\min_i \{\gamma_i \delta_i \alpha_i\}}. \end{aligned} \quad (63)$$

(b)

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|^2] \\ & \leq \frac{2\alpha_{\max}^2 nd}{(1 - \sqrt{\lambda})^2} \left((L_{\max} + \nu)^2 + \frac{8(1+c)\gamma_{\max} L_{\max}^2}{\min_i \{\gamma_i \delta_i \alpha_i\}} \right), \end{aligned} \quad (64)$$

where d is the dimension of vector $\mathbf{x}_i(t)$, $\alpha_{\max} = \max_i \alpha_i$, and $\gamma_{\max} = \max_i \gamma_i$.

Proof. Letting $\mathbf{u} = \mathbf{y}_i(t) = P_{\mathcal{X}}[\mathbf{z}_i(t)]$ in Lemma 10 and then following from the fact that $\rho(\mathbf{x}_i(t), \mathcal{X}) \leq \|\mathbf{x}_i(t) - \mathbf{y}_i(t)\|$, for $t \geq 1$ and all $i \in J_t$, we have

$$\begin{aligned} & \mathbb{E} [\rho^2(\mathbf{x}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}, I_t, J_t] \\ & \leq (1 - 2\delta_i \alpha_i) \rho^2(\mathbf{z}_i(t), \mathcal{X}) + 4(1+c) \alpha_i^2 L_{\max}^2. \end{aligned} \quad (65)$$

Note that the communication mode employs asynchronous broadcast protocol. Hence, we obtain with probability 1

$$\begin{aligned} & \mathbb{E} [\rho^2(\mathbf{x}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] \\ & \leq (1 - 2\gamma_i \delta_i \alpha_i) \rho^2(\mathbf{z}_i(t), \mathcal{X}) + 4(1+c) \gamma_i \alpha_i^2 L_{\max}^2. \end{aligned} \quad (66)$$

Summing (66) over i from 1 to n and then following from (25), we obtain

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E} [\rho^2(\mathbf{x}_i(t), \mathcal{X}) \mid \mathcal{F}_{t-1}] \\ & \leq 4(1+c) n L_{\max}^2 \max_i \{\gamma_i \alpha_i^2\} \\ & \quad + \left(1 - 2 \min_i \{\gamma_i \delta_i \alpha_i\}\right) \sum_{j=1}^n \rho^2(\mathbf{x}_j(t-1), \mathcal{X}). \end{aligned} \quad (67)$$

Therefore, from Lemma 6 in [26], the desired result is obtained.

(b) Summing (34) over ℓ from 1 to d , we obtain

$$\begin{aligned} & \sum_{\ell=1}^d \mathbb{E} [\|\mathbf{v}_{\ell}(t) - [\bar{\mathbf{x}}(t)]_{\ell} \mathbf{1}\|^2] \\ & \leq \left(\sqrt{\sum_{\ell=1}^d \mathbb{E} [\|D(t)(\mathbf{v}_{\ell}(t-1) - [\bar{\mathbf{x}}(t-1)]_{\ell} \mathbf{1})\|^2]} \right. \\ & \quad \left. + \sqrt{\sum_{\ell=1}^d \mathbb{E} [\|H\xi_{\ell}(t)\|^2]} \right)^2, \end{aligned} \quad (68)$$

where the inequality follows from Hölder's inequality. From the definitions of $D(t)$ and H and then from (35), we have

$$\begin{aligned} & \sum_{\ell=1}^d \mathbb{E} [\|D(t)(\mathbf{v}_{\ell}(t-1) - [\bar{\mathbf{x}}(t-1)]_{\ell} \mathbf{1})\|^2] \\ & \leq \lambda \sum_{\ell=1}^d \|\mathbf{v}_{\ell}(t-1) - [\bar{\mathbf{x}}(t-1)]_{\ell} \mathbf{1}\|^2. \end{aligned} \quad (69)$$

In addition, we also have

$$\begin{aligned} & \sum_{\ell=1}^d \mathbb{E} [\|H\xi_{\ell}(t)\|^2] \leq 8d \sum_{i=1}^n \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X})] \\ & \quad + 2\alpha_{\max}^2 nd (L_{\max} + \nu)^2. \end{aligned} \quad (70)$$

Letting $\kappa(t) = \sqrt{\sum_{\ell=1}^d \mathbb{E} [\|\mathbf{v}_{\ell}(t) - [\bar{\mathbf{x}}(t)]_{\ell} \mathbf{1}\|^2]}$ and then following from (68), (69), and (70), we obtain

$$\begin{aligned} & \kappa(t) = \sqrt{\lambda} \kappa(t-1) \\ & \quad + \sqrt{2\alpha_{\max}^2 nd (L_{\max} + \nu)^2 + 8d \sum_{i=1}^n \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X})]}. \end{aligned} \quad (71)$$

Following from Lemma 6 in [26], we have

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \kappa(t) \leq \frac{1}{1 - \sqrt{\lambda}} \\ & \quad \times \limsup_{t \rightarrow \infty} \sqrt{2\alpha_{\max}^2 nd (L_{\max} + \nu)^2 + 8d \sum_{i=1}^n \mathbb{E} [\rho^2(\mathbf{z}_i(t), \mathcal{X})]}. \end{aligned} \quad (72)$$

Therefore, we have

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \kappa^2(t) \\ & \leq \frac{2\alpha_{\max}^2 nd (L_{\max} + \nu)^2}{(1 - \sqrt{\lambda})^2} \\ & \quad + \limsup_{t \rightarrow \infty} \frac{8d \sum_{j=1}^n \mathbb{E} [\rho^2(\mathbf{x}_j(t-1), \mathcal{X})]}{(1 - \sqrt{\lambda})^2}. \end{aligned} \quad (73)$$

Besides, we have $\kappa^2(t) = \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \bar{\mathbf{x}}(t)\|^2]$. Therefore, by inequality (73) and part (a) of this lemma, part (b) is obtained. $\square$

We now start to prove Theorem 6.

Proof of Theorem 6. (a) Let $\mathbf{u} = \mathbf{x}^*$ in Lemma 10. Moreover, we have

$$\begin{aligned} & \nabla \phi_i(\mathbf{x}^*)^\top (\mathbf{y}_i(t) - \mathbf{x}^*) \\ & \geq \nabla \phi_i(\mathbf{x}^*)^\top (\bar{\mathbf{y}}(t) - \mathbf{x}^*) - L_{\max} \|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|. \end{aligned} \quad (74)$$

Thus, from Lemma 10 and (74), for all $i \in J_t$, we obtain

$$\begin{aligned} & \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2 \mid \mathcal{F}_{t-1}, I_t, J_t] \\ & \leq (1 - 2\delta_i \alpha_i) \|\mathbf{z}_i(t) - \mathbf{x}^*\|^2 \\ & \quad - 2\alpha_i \mathbb{E} [\nabla \phi_i(\mathbf{x}^*)^\top (\bar{\mathbf{y}}(t) - \mathbf{x}^*) \mid \mathcal{F}_{t-1}] \\ & \quad + 4(1+c) \alpha_i^2 L_{\max}^2 + 2\alpha_i L_{\max} \|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|, \end{aligned} \quad (75)$$

where we use the fact that the communication mode employs asynchronous broadcast protocol. Further, we obtain

$$\begin{aligned} & \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2 \mid \mathcal{F}_{t-1}] \\ & \leq (1 - 2\gamma_i\delta_i\alpha_i) \mathbb{E} [\|\mathbf{z}_i(t) - \mathbf{x}^*\|^2 \mid \mathcal{F}_{t-1}] \\ & \quad - 2 \min_i \{\gamma_i\alpha_i\} \mathbb{E} [\nabla\phi_i(\mathbf{x}^*)^\top (\bar{\mathbf{y}}(t) - \mathbf{x}^*) \mid \mathcal{F}_{t-1}] \\ & \quad + 2\Delta_{\gamma\alpha} \mathbb{E} [\|\nabla\phi_i(\mathbf{x}^*)\| \|\bar{\mathbf{y}}(t) - \mathbf{x}^*\| \mid \mathcal{F}_{t-1}] \\ & \quad + 2\gamma_i\alpha_i L_{\max} \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\| \mid \mathcal{F}_{t-1}] \\ & \quad + 4(1+c) \gamma_i\alpha_i^2 L_{\max}^2, \end{aligned} \quad (76)$$

where $\Delta_{\gamma\alpha} = \max_i \gamma_i\alpha_i - \min_i \gamma_i\alpha_i$. Moreover, we also have

$$\|\nabla\phi_i(\mathbf{x}^*)\| \|\bar{\mathbf{y}}(t) - \mathbf{x}^*\| \leq L_{\max} L_{\mathcal{X}}, \quad (77)$$

where $L_{\mathcal{X}} = \max_{\mathbf{x}, \mathbf{y}} \|\mathbf{x} - \mathbf{y}\|$. Hence, summing (76) over i , we have

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2 \mid \mathcal{F}_{t-1}] \leq \left(1 - 2 \min_i \{\gamma_i\delta_i\alpha_i\}\right) \\ & \quad \cdot \sum_{j=1}^n \mathbb{E} [\|\mathbf{x}_j(t-1) - \mathbf{x}^*\|^2 \mid \mathcal{F}_{t-1}] \\ & \quad + 2\gamma_{\max}\alpha_{\max} L_{\max} \sum_{i=1}^n \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\| \mid \mathcal{F}_{t-1}] \\ & \quad + 2n\Delta_{\gamma\alpha} L_{\max} L_{\mathcal{X}} + 4(1+c) n\gamma_{\max}\alpha_{\max}^2 L_{\max}^2. \end{aligned} \quad (78)$$

Thus, from (78), we have

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2] \\ & \leq \frac{\Delta_{\gamma\alpha} L_{\max} L_{\mathcal{X}} + 2(1+c) \gamma_{\max}\alpha_{\max}^2 L_{\max}^2}{\min_i \{\gamma_i\delta_i\alpha_i\}} \\ & \quad + \frac{\gamma_{\max}\alpha_{\max} L_{\max}}{n \min_i \{\gamma_i\delta_i\alpha_i\}} \limsup_{t \rightarrow \infty} \sum_{i=1}^n \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|]. \end{aligned} \quad (79)$$

Besides, we also obtain

$$\begin{aligned} & \sum_{i=1}^n \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|] \leq \sqrt{n \sum_{i=1}^n \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|^2]} \\ & \leq \sqrt{n \sum_{i=1}^n \mathbb{E} [\|\mathbf{z}_i(t) - \bar{\mathbf{z}}(t)\|^2]} \\ & \leq \sqrt{n \sum_{j=1}^n \mathbb{E} [\|\mathbf{x}_j(t-1) - \bar{\mathbf{x}}(t-1)\|^2]}. \end{aligned} \quad (80)$$

Hence, combining (80) and Lemma 11(b), we have

$$\begin{aligned} & \limsup_{t \rightarrow \infty} \sum_{i=1}^n \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|] \\ & \leq \frac{\alpha_{\max} n}{1 - \sqrt{\lambda}} \sqrt{2d \left((L_{\max} + \nu)^2 + \frac{8(1+c) \gamma_{\max} L_{\max}^2}{\min_i \{\gamma_i\delta_i\alpha_i\}} \right)}. \end{aligned} \quad (81)$$

Therefore, the conclusion of part (a) is proved.

(b) From (76) and (77), we have

$$\begin{aligned} & \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2 \mid \mathcal{F}_{t-1}] \\ & \leq (1 - 2\gamma_i\delta_i\alpha_i) \mathbb{E} [\|\mathbf{z}_i(t) - \mathbf{x}^*\|^2 \mid \mathcal{F}_{t-1}] \\ & \quad - 2 \min_i \{\gamma_i\alpha_i\} (\phi_i(\bar{\mathbf{y}}(t)) - \phi_i(\mathbf{x}^*)) \\ & \quad + 2\gamma_{\max}\alpha_{\max} L_{\max} \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\| \mid \mathcal{F}_{t-1}] \\ & \quad + 2\Delta_{\gamma\alpha} L_{\max} L_{\mathcal{X}} + 4(1+c) \gamma_{\max}\alpha_{\max}^2 L_{\max}^2. \end{aligned} \quad (82)$$

Due to

$$\sum_{i=1}^n \mathbb{E} [\|\mathbf{z}_i(t) - \mathbf{x}^*\|^2 \mid \mathcal{F}_{t-1}] \leq \sum_{j=1}^n \|\mathbf{x}_j(t-1) - \mathbf{x}^*\|^2, \quad (83)$$

by using the above relation and by some algebraic operations, we obtain with probability 1

$$\begin{aligned} & \mathbb{E} [\phi(\mathbf{x}_i(t-1)) - \phi^*] \leq \frac{1 - 2\gamma_i\delta_i\alpha_i}{2 \min_i \{\gamma_i\alpha_i\}} \\ & \quad \times \left(\sum_{j=1}^n \mathbb{E} [\|\mathbf{x}_j(t-1) - \mathbf{x}^*\|^2] \right. \\ & \quad \left. - \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{x}^*\|^2] \right) + \frac{\gamma_{\max}\alpha_{\max} L_{\max}}{\min_i \{\gamma_i\alpha_i\}} \\ & \quad \cdot \sum_{i=1}^n \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|] \\ & \quad + \frac{n\Delta_{\gamma\alpha} L_{\max} L_{\mathcal{X}} + 2n(1+c) \gamma_{\max}\alpha_{\max}^2 L_{\max}^2}{\min_i \{\gamma_i\alpha_i\}} \\ & \quad + \frac{L_{\max}}{\min_i \{\gamma_i\alpha_i\}} \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t-1) - \bar{\mathbf{y}}(t-1)\|], \end{aligned} \quad (84)$$

where we use the following relation:

$$\begin{aligned} & \phi(\mathbf{x}_i(t-1)) - \phi^* \\ & \leq \phi(\bar{\mathbf{y}}(t-1)) - \phi^* \\ & \quad + L_{\max} \sum_{j=1}^n \|\mathbf{x}_j(t-1) - \bar{\mathbf{y}}(t-1)\|. \end{aligned} \quad (85)$$

Hence, summing (84) over t from 1 to T and then taking the average, we have

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^T \mathbb{E} [\phi(\mathbf{x}_i(t-1)) - \phi^*] \\
& \leq \frac{1 - 2\gamma_i \delta_i \alpha_i}{2T \min_i \{\gamma_i \alpha_i\}} \sum_{j=1}^n \mathbb{E} [\|\mathbf{x}_j(0) - \mathbf{x}^*\|^2] \\
& \quad + \frac{\gamma_{\max} \alpha_{\max} L_{\max}}{\min_i \{\gamma_i \alpha_i\}} \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^n \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|] \\
& \quad + \frac{n\Delta_{\gamma\alpha} L_{\max} L_{\mathcal{X}} + 2n(1+c) \gamma_{\max} \alpha_{\max}^2 L_{\max}^2}{\min_i \{\gamma_i \alpha_i\}} \\
& \quad + \frac{L_{\max}}{\min_i \{\gamma_i \alpha_i\}} \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t-1) - \bar{\mathbf{y}}(t-1)\|].
\end{aligned} \tag{86}$$

Letting $T \rightarrow \infty$, we obtain

$$\begin{aligned}
& \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} [\phi(\mathbf{x}_i(t-1)) - \phi^*] \\
& \leq \frac{\gamma_{\max} \alpha_{\max} L_{\max}}{\min_i \{\gamma_i \alpha_i\}} \limsup_{t \rightarrow \infty} \sum_{i=1}^n \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|] \\
& \quad + \frac{L_{\max}}{\min_i \{\gamma_i \alpha_i\}} \limsup_{t \rightarrow \infty} \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t-1) - \bar{\mathbf{y}}(t-1)\|] \\
& \quad + \frac{n\Delta_{\gamma\alpha} L_{\max} L_{\mathcal{X}} + 2n(1+c) \gamma_{\max} \alpha_{\max}^2 L_{\max}^2}{\min_i \{\gamma_i \alpha_i\}},
\end{aligned} \tag{87}$$

where the inequality follows from inequality $\limsup_{T \rightarrow \infty} (1/T) \sum_{t=1}^T \theta_t = \limsup_{t \rightarrow \infty} \theta_t$. Next, we compute the term $\limsup_{t \rightarrow \infty} \sum_{i=1}^n \mathbb{E} [\|\mathbf{y}_i(t) - \bar{\mathbf{y}}(t)\|]$. From triangle inequality, we have

$$\begin{aligned}
& \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \bar{\mathbf{y}}(t)\|] \\
& \leq \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \mathbf{z}_i(t)\| + \|\mathbf{z}_i(t) - \bar{\mathbf{y}}(t)\|].
\end{aligned} \tag{88}$$

Further, from the definition of $\bar{\mathbf{y}}(t)$ and then by the convexity of norm, we obtain

$$\begin{aligned}
& \sum_{i=1}^n \mathbb{E} [\|\mathbf{z}_i(t) - \bar{\mathbf{y}}(t)\|] \leq \sum_{i=1}^n \mathbb{E} [\|\mathbf{z}_i(t) - \mathbf{y}_i(t)\|] \\
& = \sum_{i=1}^n \mathbb{E} [\rho(\mathbf{z}_i(t), \mathcal{X})].
\end{aligned} \tag{89}$$

Hence, according to Lemma 8, we have for all $i \in \mathcal{V}$

$$\limsup_{t \rightarrow \infty} \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_i(t) - \bar{\mathbf{y}}(t)\|] = 0. \tag{90}$$

Following from (81) and (90), with probability 1 we have for all $i \in \mathcal{V}$ and $t \geq 1$

$$\begin{aligned}
& \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E} [\phi(\mathbf{x}_i(t-1)) - \phi^*] \\
& \leq \frac{n\Delta_{\gamma\alpha} L_{\max} L_{\mathcal{X}} + 2n(1+c) \gamma_{\max} \alpha_{\max}^2 L_{\max}^2}{\min_i \{\gamma_i \alpha_i\}} \\
& \quad + \frac{n\gamma_{\max} \alpha_{\max}^2 L_{\max}}{(1 - \sqrt{\lambda}) \min_i \{\gamma_i \alpha_i\}} \\
& \quad \times \sqrt{2d \left((L_{\max} + \nu)^2 + \frac{8(1+c) \gamma_{\max} L_{\max}^2}{\min_i \{\gamma_i \sigma_i \alpha_i\}} \right)}.
\end{aligned} \tag{91}$$

Therefore, by the convexity of global cost function ϕ , part (b) is obtained. $\square$

6. Conclusion

In this paper, we have considered a constrained distributed optimization problem. Moreover, the global cost function is a sum of local convex function and each agent only knows its own cost function and has access to noisy subgradient of its own objective function. However, each agent does not know constraint set in advance. To solve this problem, we proposed a distributed optimization algorithm over a time-varying network. The algorithm employs random projection method and the communication protocol uses broadcast communication protocol in an asynchronous way. We have showed that the proposed algorithm is convergent with probability 1 by choosing suitable stepsizes. We have also established two asymptotic error bounds with appropriately chosen stepsizes.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Basic Research Program of China (973) under Grant 2013CB329102, in part by the National Natural Science Foundation of China (NSFC) under Grant nos. 61522103, 61372112, 61370221, U1404611, and U1604155, in part by the Program for Science & Technology Innovation Talents in the University of Henan Province under Grant no. 16HASTIT035, in part by Henan Science and Technology Innovation Outstanding Talent under Grant no. 164200510007, and in part by Foundation of Henan Educational Committee under Grant no. 16A510005.

References

- [1] V. Lesser, C. L. Ortiz, and M. Tambe, "Distributed sensor networks: introduction to a multiagent perspective," in *Distributed*

- Sensor Networks*, vol. 9 of *Multiagent Systems, Artificial Societies, and Simulated Organizations*, pp. 1–8, Springer US, Boston, MA, 2003.
- [2] M. Rabbat and R. Nowak, “Distributed optimization in sensor networks,” in *Proceedings of the the third international symposium*, pp. 20–27, Berkeley, California, CA, USA, 2004.
 - [3] S. Kar and J. M. Moura, “Distributed consensus algorithms in sensor networks: quantized data and random link failures,” *IEEE Transactions on Signal Processing*, vol. 58, no. 3, part 1, pp. 1383–1400, 2010.
 - [4] S. Kar, J. M. Moura, and K. Ramanan, “Distributed parameter estimation in sensor networks: nonlinear observation models and imperfect communication,” *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.
 - [5] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, Prediction*, Springer-Verlag, New York, NY, USA, 2001.
 - [6] J. L. R. Bekkerman and M. Bilenko, *Scaling up Machine Learning: Parallel and Distributed Approaches*, Cambridge Univ. Press, Cambridge, U.K., 2011.
 - [7] R. L. Cavalcante, I. Yamada, and B. Mulgrew, “An adaptive projected subgradient approach to learning in diffusion networks,” *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2762–2774, 2009.
 - [8] B. Johansson, *On Distributed Optimization in Networked Systems [Ph.D. thesis]*, Royal Insitute of Technology, Stockholm, Sweden, 2008.
 - [9] A. Beck, A. Nedić, A. Ozdaglar, and M. Teboulle, “An O1/k gradient method for network resource allocation problems,” *IEEE Transactions on Control of Network Systems*, vol. 1, no. 1, pp. 64–73, 2014.
 - [10] T.-H. Chang, A. Nedić, and A. Scaglione, “Distributed constrained optimization by consensus-based primal-dual perturbation method,” *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, vol. 59, no. 6, pp. 1524–1538, 2014.
 - [11] Z. Jiang, S. Sarkar, and K. Mukherjee, “On distributed optimization using generalized gossip,” in *Proceedings of the 54th IEEE Conference on Decision and Control, CDC 2015*, pp. 2667–2672, Japan, 2015.
 - [12] A. Nedić and A. Olshevsky, “Distributed optimization over time-varying directed graphs,” *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.
 - [13] A. Makhdoomi and A. Ozdaglar, “Graph balancing for distributed subgradient methods over directed graphs,” in *Proceedings of the 54th IEEE Conference on Decision and Control, CDC 2015*, pp. 1364–1371, Japan, 2015.
 - [14] D. Varagnolo, F. Zanella, A. Cenedese, G. Pillonetto, and L. Schenato, “Newton-Raphson consensus for distributed convex optimization,” *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, vol. 61, no. 4, pp. 994–1009, 2016.
 - [15] A. Nedić and A. Olshevsky, “Stochastic gradient-push for strongly convex functions on time-varying directed graphs,” *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, vol. 61, no. 12, pp. 3936–3947, 2016.
 - [16] A. Nedić, A. Ozdaglar, and A. Parrilo, “Constrained consensus and optimization in multi-agent networks,” *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.
 - [17] K. Srivastava and A. Nedić, “Distributed asynchronous constrained stochastic optimization,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 772–790, 2011.
 - [18] I. Lobel, A. Ozdaglar, and D. Feijer, “Distributed multi-agent optimization with state-dependent communication,” *Mathematical Programming*, vol. 129, no. 2, Ser. B, pp. 255–284, 2011.
 - [19] A. Nedić, “Random algorithms for convex minimization problems,” *Mathematical Programming*, vol. 129, no. 2, Ser. B, pp. 225–253, 2011.
 - [20] S. Lee and A. Nedić, “Distributed random projection algorithm for convex optimization,” *IEEE Journal on Selected Topics in Signal Processing*, vol. 7, no. 2, pp. 221–229, 2013.
 - [21] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, “Randomized gossip algorithms,” *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.
 - [22] T. C. Aysal, M. E. Yildiz, A. Sarwate, and A. Scaglione, “Broadcast gossip algorithms for consensus,” *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2748–2761, 2009.
 - [23] S. Lee and A. Nedić, “Asynchronous gossip-based random projection algorithms over networks,” *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, vol. 61, no. 4, pp. 953–968, 2016.
 - [24] F. Fagnani and S. Zampieri, “Randomized consensus algorithms over large scale networks,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 634–649, 2008.
 - [25] Y. Yang and R. S. Blum, “Broadcast-based consensus with non-zero-mean stochastic perturbations,” *IEEE Transactions on Information Theory*, vol. 59, no. 6, pp. 3971–3989, 2013.
 - [26] A. Nedić, “Asynchronous broadcast-based convex optimization over a network,” *Institute of Electrical and Electronics Engineers. Transactions on Automatic Control*, vol. 56, no. 6, pp. 1337–1351, 2011.
 - [27] L. Gubin, B. Polyak, and E. Raik, “The method of projections for finding the common point of convex sets,” *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 6, pp. 1211–1228, 1967.
 - [28] D. Bertsekas, A. Nedić, and A. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, Belmont, Mass, MA, USA, 2003.
 - [29] B. T. Polyak, *Introduction to Optimization*, Optimization Software, New York, NY, USA, 1987.
 - [30] H. Royden, *Real Analysis*, Prentice Hall, Englewood Cliffs, NJ, USA, 3rd edition, 1998.

Hindawi

Submit your manuscripts at
<https://www.hindawi.com>