

## Research Article

# Distributed Stochastic Subgradient Projection Algorithms Based on Weight-Balancing over Time-Varying Directed Graphs

Junlong Zhu , Ping Xie , Mingchuan Zhang , Ruijuan Zheng ,  
Ling Xing , and Qingtao Wu 

Information Engineering College, Henan University of Science and Technology, Luoyang 471023, China

Correspondence should be addressed to Mingchuan Zhang; zhang\_mch@haust.edu.cn

Received 27 April 2019; Accepted 10 July 2019; Published 20 August 2019

Academic Editor: Átila Bueno

Copyright © 2019 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 graphs, where cost function of each agent is private. Moreover, we assume that the graphs are time-varying and directed. In order to address such problem, a fully decentralized stochastic subgradient projection algorithm is proposed over time-varying directed graphs. However, since the graphs are directed, the weight matrix may not be a doubly stochastic matrix. Therefore, we overcome this difficulty by using weight-balancing technique. By choosing appropriate step-sizes, we show that iterations of all agents asymptotically converge to some optimal solutions. Further, by our analysis, convergence rate of our proposed algorithm is  $O(\ln\Gamma/\Gamma)$  under local strong convexity, where  $\Gamma$  is the number of iterations. In addition, under local convexity, we prove that our proposed algorithm can converge with rate  $O(\ln\Gamma/\sqrt{\Gamma})$ . In addition, we verify the theoretical results through simulations.

## 1. Introduction

In this paper, we focus on distributed constrained optimization problems, which have arisen in many applications. For instance, large-scale machine learning [1–4], resource allocation [5, 6], sensor networks [7–10], and multiagent systems [11]. To address such problems, the designs of distributed optimization algorithms are necessary. Moreover, the goal is to minimize the sum of cost functions of all agents over networks, where each agent only knows its own information and can receive the information from its neighbors.

Distributed optimization algorithms are originally introduced in seminal work [12] and have a vast literature devoted to them in recent years [13–19]. In this literature, the distributed (sub)gradient methods are used over networks. Moreover, the performance limitations and convergence rates of these algorithms are well understood. In addition, the distributed Newton methods and other descent methods are used to solve the distributed optimization problems [20, 21], and their rates of convergence are also analyzed.

However, these methods cited above assumed that information exchange among agents takes place over either fixed

or undirected graphs. Nevertheless, in some communication networks such as mobile sensor networks, the communication between agents is unidirectional because different agents have different interference and noise patterns, the information are broadcasted at different power level, and the communication links between agents are directed in these networks [22–24]. Hence, the directed network topology is a natural assumption. In addition, the time-varying communication network topology is also a valid assumption in wireless networks, where each agent can move or the communication links may be destroyed randomly. For these reasons, we assume that these network topologies can be modeled as time-varying directed graphs. In this paper, we assume that each out-degree is known to each agent at each round, where the assumption cannot be removed [25]. To obtain knowledge of the out-degree, we can bidirectionally exchange “Hello” messages in a single communication process.

In fact, the recent works [22, 23] provide subgradient-push distributed algorithms to minimize the cost function over time-varying directed network in discrete-time. Similarly, [24] proposes a distributed subgradient algorithm, which used weight-balancing in discrete-time. In addition,

[26] considers a continuous-time optimization algorithm over underlying time-varying directed networks. Moreover, the distributed algorithms in [22, 24] converge with rate  $O(\ln \Gamma / \sqrt{\Gamma})$ , whereas in [23], their algorithm converges with rate  $O(\ln \Gamma / \Gamma)$  by assuming that the local cost functions are differentiable and strongly convex. Nevertheless, these works consider the unconstrained optimization problem. In [27], the authors propose the D-DPS algorithm over a directed network with convergence rate  $O(\ln \Gamma / \sqrt{\Gamma})$ . However, we assume that the graph is time-varying and directed in this paper. To overcome the asymmetry caused by directed graphs, we employ the weight-balancing technique in the paper. Hence, a distributed stochastic subgradient projection algorithm is proposed, based on weight-balancing technique. Assume that each local cost function is strongly convex; even if all agents have access to their own noisy subgradient, our proposed algorithm is asymptotically convergent with rate  $O(\ln \Gamma / \Gamma)$ . Besides, our proposed algorithm also asymptotically converges with rate  $O(\ln \Gamma / \sqrt{\Gamma})$  for generally convex cost functions. Compared with the best previous convergence rate of  $O(1/\Gamma)$ , which is achieved in centralized way, where the local functions are strongly convex and differentiable, the variance of noisy subgradient is bounded [28, 29]. Thus, our convergence results are quite close to it. However, we need not assume that local cost functions are differentiable. In addition, we assume that local cost functions do not establish Lipschitz continuous gradients.

Our goal is to design a distributed optimization algorithm and analyze the properties of the proposed algorithm, based on weight-balancing over time-varying directed networks. This work has the following contributions:

- (i) We propose a distributed stochastic subgradient projection algorithm based on weight-balancing over time-varying directed networks. Furthermore, each local cost function is private information for other agents. Hence, each agent only utilizes its own private information. Moreover, noisy subgradient  $\mathbf{s}_i$  of local cost function  $g_i$  is known to each agent  $i$  for  $i = 1, \dots, N$ . In addition, the algorithm is implemented without any centralized control, and every agent need not to know network topology and only requires to know its out-degree at each round.
- (ii) By some standard assumptions, we show that our proposed algorithm asymptotically converges to some optimal solutions.
- (iii) For strongly convex cost functions, we prove that the convergence rate of  $O(\ln \Gamma / \Gamma)$  is achieved. In addition, we also show that the convergence rate is  $O(\ln \Gamma / \sqrt{\Gamma})$  for locally convex cost functions.

*Organization.* In Section 2, the constrained optimization problem is described, we also give some assumptions, and then a distributed stochastic subgradient projection algorithm is proposed over time-varying directed networks. We state the main results of this paper in Section 3. We provide the proofs of the main results in Section 4. In Section 5,

simulations are also presented. The conclusion of the paper is provided in Section 6.

*Notation.* We use lowercase boldface to denote the vectors in  $\mathbb{R}^d$  and use lowercase normal font to denote scalars or vectors, which are not  $d$ -dimensional vectors. For instance,  $\mathbf{u}_i(\tau)$  denotes a vector in  $\mathbb{R}^d$  at agent  $i$  at round  $\tau$ , while the notation  $y_i(\tau)$  is a scalar in  $\mathbb{R}$ . The vector such as  $\mathbf{y}(\tau)$  in  $\mathbb{R}^N$  is obtained by stacking all scalars  $y_i(\tau)$  for  $i = 1, \dots, N$ . In addition, we use the notation “T” to denote transpose operation.  $\|\mathbf{u}\|$  is the Euclidean norm of  $\mathbf{u}$ . Besides, the notation  $\|\mathbf{u}\|_1$  is 1-norm of  $\mathbf{u}$ . The notation  $\mathbf{1}$  denotes a  $N$ -dimensional vector whose all elements are 1, and  $I$  is identity matrix, whose size is  $N \times N$ .  $\mathbb{E}[\cdot]$  denotes the expectation operator. Besides, the notation  $\Pi_{\mathcal{U}}[\mathbf{u}]$  means that a vector  $\mathbf{u}$  is projected onto the constraint set  $\mathcal{U}$ . The notation  $\otimes$  denotes Kronecker product operator.

## 2. Problem Setup, Algorithm Description, and Assumptions

We use a graph to model a network, which consists of  $N$  agents. Moreover, each agent represents a node. Further, we also consider the case that network topology is time-varying and directed. Hence, we use the notation  $\mathcal{G}(\tau) = (\mathcal{V}, \mathcal{E}(\tau))$  to denote network topology at each round  $\tau$ , where  $\mathcal{V}$  denotes the agent set and  $\mathcal{E}(\tau)$  denotes the directed edge set.  $(j, i) \in \mathcal{E}(\tau)$  means that agent  $j$  can send message to agent  $i$  at round  $\tau$ . If two agents can directly exchange information, then we say the agents are neighboring. Furthermore, we use  $\mathcal{N}_j^{\text{out}}(\tau)$  and  $\mathcal{N}_j^{\text{in}}(\tau)$  to represent the set of out-neighbors and the set of in-neighbors of agent  $j$  at round  $\tau$ , respectively. Formally,  $\mathcal{N}_j^{\text{out}}(\tau) = \{k \mid (j, k) \in \mathcal{E}(\tau), k \in \mathcal{V}\}$  and  $\mathcal{N}_j^{\text{in}}(\tau) = \{k \mid (k, j) \in \mathcal{E}(\tau), k \in \mathcal{V}\}$ , respectively. Besides, we denote the out-degree and in-degree by  $d_j^{\text{out}}(\tau) = |\mathcal{N}_j^{\text{out}}(\tau)|$  and  $d_j^{\text{in}}(\tau) = |\mathcal{N}_j^{\text{in}}(\tau)|$ , respectively.

In this paper, the constrained optimization problem is described as follows:

$$\begin{aligned} \underset{\mathbf{u}}{\text{minimize}} \quad & g(\mathbf{u}) = \sum_{i=1}^N g_i(\mathbf{u}) \\ \text{s.t.} \quad & \mathbf{u} \in \mathcal{U}, \end{aligned} \tag{1}$$

where  $g_i : \mathbb{R}^d \rightarrow \mathbb{R}$  denotes local cost function of agent  $i$  and  $\mathcal{U} \subseteq \mathbb{R}^d$  denotes constraint set.

Our goal is to solve problem (1) by cooperative and fully decentralized way over time-varying and directed networks. Further, each local cost function  $g_i$  can be only known to each agent  $i$  and all agents know constraint set  $\mathcal{U}$ . Moreover, each agent can share its own iteration with its out-neighbors.

In this paper, we assume that the network topology is time-varying and directed. Note that the directed graph may cause the asymmetry. Thus, to overcome the asymmetry, we employ the weight-balancing technique in the paper. Following from [24], we give the definition of balancing weights over time-varying directed networks as follows.

*Definition 1* (balancing weights). The weight  $w_{i,\varrho(\tau)}$  of agent  $i$ ,  $i \in \mathcal{V}$ , balances a time-varying directed network at round  $\tau$  if, for any agent  $i$ , the agent weight  $w_{i,\varrho(\tau)}$  satisfies

$$w_{i,\varrho(\tau)} d_i^{\text{out}}(\tau) = \sum_{j \in \mathcal{N}_i^{\text{in}}(\tau)} w_{j,\varrho(\tau)}, \quad (2)$$

where  $\varrho(\tau)$  is the index of relationships of neighbors at time  $\tau$ .

From Definition 1, the total weight incoming from agent  $i$  (which is  $\sum_{j \in \mathcal{N}_i^{\text{in}}(\tau)} w_{j,\varrho(\tau)}$ ) is equal to the total outgoing weight of agent  $i$  (which is  $w_{i,\varrho(\tau)} d_i^{\text{out}}(\tau)$ ) at round  $\tau$  over a time-varying directed network.

In order to solve problem (1) over a time-varying directed graph  $\mathcal{G}(\tau)$ , we first give some standard assumptions as follows.

*Assumption 2.* Assume that the time-varying directed network sequence  $\mathcal{G}(\tau)$  is strongly connected at each round  $\tau$ .

Assumption 2 ensures that each agent can receive information from the other agents at round  $\tau$  in network  $\mathcal{G}(\tau)$ .

*Assumption 3.* Let constraint set  $\mathcal{U} \subseteq \mathbb{R}^d$  be closed and convex. In addition, assuming that local cost function  $g_i(\mathbf{u})$  is convex with  $\sigma_i \geq 0$ , for all  $i \in \mathcal{V}$ , i.e., for  $\mathbf{u}, \mathbf{v} \in \mathcal{U}$ , each local cost function  $g_i$  satisfies

$$g_i(\mathbf{u}) - g_i(\mathbf{v}) \geq \nabla g_i(\mathbf{v})^\top (\mathbf{u} - \mathbf{v}) + \frac{\sigma_i}{2} \|\mathbf{u} - \mathbf{v}\|^2, \quad (3)$$

where  $\nabla g_i(\mathbf{v})$  denotes a (sub)gradient of  $g_i(\mathbf{u})$  at  $\mathbf{u} = \mathbf{v}$ . If  $\sigma_i > 0$ ,  $g_i(\mathbf{v})$  is  $\sigma_i$ -strongly convex. Otherwise,  $g_i(\mathbf{v})$  is convex.

Assumption 3 ensures that the set of subgradients  $\partial g_i$  is nonempty. Besides, the assumption on the subgradient is provided as follows. The similar assumption can be found in [13, 24, 27].

*Assumption 4.* We assume that the subgradient  $\nabla g_i(\mathbf{u})$  of  $g_i(\mathbf{u})$  is uniformly bounded over  $\mathcal{U}$  for  $i \in \mathcal{V}$ . Namely,  $\|\nabla g_i(\mathbf{u})\| \leq L_i$ , for all  $\mathbf{u} \in \mathcal{U}$ .

Next, we describe our proposed optimization algorithm which is executed over a time-varying directed network. Assume that  $\mathbf{u}_i(\tau) \in \mathcal{U}$  is the iteration of agent  $i$  at round  $\tau$ . Moreover, the iteration  $\mathbf{u}_i(\tau)$  is updated as follows:

$$\begin{aligned} \mathbf{v}_i(\tau) &= (1 - w_i(\tau) d_i^{\text{out}}(\tau)) \mathbf{u}_i(\tau) \\ &+ \sum_{j \in \mathcal{N}_i^{\text{in}}(\tau)} w_j(\tau) \mathbf{u}_j(\tau), \end{aligned} \quad (4)$$

$$\mathbf{u}_i(\tau + 1) = \Pi_{\mathcal{U}} [\mathbf{v}_i(\tau) - \alpha(\tau) \mathbf{s}_i(\tau)], \quad (5)$$

for all  $i \in \mathcal{V}$ , where  $\alpha(\tau)$  denotes a step-size sequence and we use  $\mathbf{s}_i(\tau)$  to abbreviate the notation  $\mathbf{s}_i(\mathbf{u}_i(\tau))$ , which represents a noisy subgradient of  $g_i(\mathbf{u})$  at  $\mathbf{u} = \mathbf{u}_i(\tau)$ . Following from (4)-(5), each agent  $i$  first linearly fuses from its own estimate and the estimates of in-coming neighbor agents and

updates the estimate in opposite direction of its own noisy subgradient. Finally, a new estimate of agent  $i$  is obtained by projecting the updated estimate onto the constraint set  $\mathcal{U}$ . Moreover, the above update equations can be executed by simple broadcast communication.

From (5), we need to make some assumptions about the noisy subgradient. Specifically, we describe the noisy subgradient  $\mathbf{s}_i(\mathbf{u}_i(\tau))$  as follows:

$$\mathbf{s}_i(\mathbf{u}_i(\tau)) = \nabla g_i(\mathbf{u}_i(\tau)) + \boldsymbol{\epsilon}_i(\tau), \quad (6)$$

where  $\boldsymbol{\epsilon}_i(\tau) \in \mathcal{U}$  denotes a stochastic subgradient error and  $\nabla g_i(\mathbf{u}_i(\tau))$  denotes a subgradient of  $g_i(\mathbf{u})$  at  $\mathbf{u} = \mathbf{u}_i(\tau)$ . Let  $\mathcal{F}_\tau$  denote all the information generated by the distributed stochastic subgradient projection algorithms (4)-(5) for all  $\tau \geq 0$ . Hence, the assumption for stochastic subgradient error  $\boldsymbol{\epsilon}_i(\tau)$  is as follows.

*Assumption 5.* For every  $i$ ,  $i = 1, \dots, N$ , we assume that the stochastic subgradient error  $\boldsymbol{\epsilon}_i(\tau)$  at round  $\tau \geq 1$  is a random variable with  $\mathbb{E}[\boldsymbol{\epsilon}_i(\tau) \mid \mathcal{F}_{\tau-1}] = \mathbf{0}$ . Moreover, assume that  $\boldsymbol{\epsilon}_i(\tau)$  are independent of each other. Assume that the noise-norm  $\|\boldsymbol{\epsilon}_i(\tau)\|$  at round  $\tau \geq 1$  is uniformly bounded, i.e.,  $\mathbb{E}[\|\boldsymbol{\epsilon}_i(\tau)\| \mid \mathcal{F}_{\tau-1}] \leq \nu_i$ , where  $\nu_i$  is a positive scalar.

In our proposed algorithm,  $w_i(\tau)$  is weight of agent  $i$  at round  $\tau$ . As with [24], every agent  $i$  updates its weight  $w_i(\tau)$  over time-varying directed networks as follows:

$$w_i(\tau + 1) = \frac{1}{2} w_i(\tau) + \frac{1}{d_i^{\text{out}}(\tau)} \sum_{j \in \mathcal{N}_i^{\text{in}}(\tau)} \frac{1}{2} w_j(\tau). \quad (7)$$

In this paper, let the optimal set of the proposed algorithms be nonempty, which can be defined as follows:

$$\mathcal{U}^* \triangleq \{\mathbf{u} \mid g(\mathbf{u}) = g^*, \mathbf{u} \in \mathcal{U}\}, \quad (8)$$

where  $g^* \triangleq g(\mathbf{u}^*) \triangleq \min_{\mathbf{u} \in \mathcal{U}} g(\mathbf{u})$ .

First, we formally introduce the constrained optimization problem in this section, and then we also give some valid assumptions. Simultaneously, a distributed stochastic subgradient projection algorithm is proposed to solve constrained optimization problem (1) over time-varying directed networks. The main results of the paper are presented in the next section.

### 3. Main Results

We first present an asymptotic convergence of the distributed stochastic subgradient projection algorithms (4)-(5) with appropriately chosen step-sizes. Specifically, the result is described as follows.

**Theorem 6.** *Under Assumptions 2–5, let the optimal set  $\mathcal{U}^*$  be nonempty. Moreover, for the positive step-size  $\alpha(\tau)$ , which satisfies decay conditions (26), let sequences  $\{\mathbf{u}_i(\tau)\}$ ,  $i = 1, \dots, N$ , be generated by algorithms (4)-(5). Then, each sequence  $\{\mathbf{u}_i(\tau)\}$  converges to some optimal solutions  $\mathbf{u}^*$  in  $\mathcal{U}^*$  with probability 1 for all  $i \in \mathcal{V}$ . Namely, for all  $i \in \mathcal{V}$ ,  $\lim_{\tau \rightarrow \infty} \mathbf{u}_i(\tau) = \mathbf{u}^*$  holds with probability 1.*

By Theorem 6, the iterations asymptotically converge to some optimal solutions over time-varying directed networks. Namely, by our proposed algorithm, we can obtain some optimal solutions with probability 1.

We now state convergence rate of our proposed algorithm. Under different assumptions about local cost functions  $g_i$ , we establish the different convergence rate. To this end, we first introduce the weighted average of the estimate sequence  $\{\mathbf{u}_i(\tau)\}_{\tau \geq 0}$ , which is defined as [23]

$$\hat{\mathbf{u}}_i(\tau) = \frac{\sum_{t=1}^{\tau} (t-1) \mathbf{u}_i(t)}{\tau(\tau-1)/2}, \quad (9)$$

for all  $\tau \geq 2$ . Hence, for all  $\tau \geq 1$ , we have the following recursive relation:

$$\hat{\mathbf{u}}_i(\tau+1) = \frac{\tau \mathbf{u}_i(\tau+1) + S(\tau) \hat{\mathbf{u}}_i(\tau)}{S(\tau+1)}, \quad (10)$$

for all  $i \in \mathcal{V}$ , where  $S(\tau) = \tau(\tau-1)/2$ , for all  $\tau \geq 2$ , and  $\hat{\mathbf{u}}_i(1) = \mathbf{u}_i(0)$ .

**Theorem 7.** Under Assumptions 2-5, let  $\alpha(\tau) = \mu/(\tau+1)$ , where the positive constant  $\mu$  satisfies

$$\mu \frac{\sum_{i=1}^N \sigma_i}{N} \geq 4. \quad (11)$$

Then, we obtain

$$\begin{aligned} & \mathbb{E} \left[ g(\hat{\mathbf{u}}_i(\Gamma)) - g(\mathbf{u}^*) + \sum_{j=1}^N \sigma_j \|\mathbf{u}_j(\Gamma) - \mathbf{u}^*\| \right] \\ & \leq \frac{10LC}{\Gamma} \left( \frac{\lambda}{1-\lambda} \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 \right. \\ & \quad \left. + \frac{\mu}{1-\lambda} \sum_{i=1}^N \sqrt{d}(L_i + \nu_i)(1 + \ln(\Gamma-1)) \right) + \frac{8\mu}{\Gamma} \\ & \quad \cdot \sum_{i=1}^N (L_i + \nu_i)^2, \end{aligned} \quad (12)$$

for all  $i \in \mathcal{V}$ , where  $L_i$ ,  $\nu_i$ , and  $C$  are positive constants,  $\lambda \in (0, 1)$ ,  $L = \sum_{i=1}^N L_i$ ,  $\sigma_i > 0$ , and  $\Gamma$  denotes the number of iterations.

From Theorem 7, we establish the convergence rate when local cost functions are strongly convex. Moreover,  $g(\hat{\mathbf{u}}_i(\Gamma))$  converges to  $g(\mathbf{u}^*)$  for any agent  $i$  with probability 1. Further, following from (12), the numerator has an  $\ln \Gamma$  or a constant and the denominator has a  $\Gamma$ . Therefore, our algorithm converges to some optimal solutions with probability 1 at rate of  $O(\ln \Gamma/\Gamma)$ .

**Theorem 8.** Under Assumptions 2-5, let  $\alpha(\tau) = 1/\sqrt{\tau+1}$  for  $\tau \geq 0$ . Then, for all  $i = 1, \dots, N$ , we obtain that with probability 1

$$\begin{aligned} & \mathbb{E} [g(\tilde{\mathbf{u}}_i(\Gamma)) - g(\mathbf{u}^*)] \\ & \leq \frac{3LC \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1}{(1-\lambda)\sqrt{\Gamma+1}} \\ & \quad + \frac{9LC \ln(\Gamma+1)}{(1-\lambda)\sqrt{\Gamma+1}} \sum_{j=1}^N 2\sqrt{d}(L_j + \nu_j) \\ & \quad + \frac{2(1 + \ln(\Gamma+1))}{\sqrt{\Gamma+1}} \sum_{i=1}^N (L_i + \nu_i)^2 \\ & \quad + \frac{N}{2\sqrt{\Gamma+1}} \|\bar{\mathbf{u}}(0) - \mathbf{u}^*\|^2, \end{aligned} \quad (13)$$

where  $\tilde{\mathbf{u}}_i(\Gamma)$  is defined as (10),  $L_i$ ,  $\nu_i$ , and  $C$  are positive constants,  $\lambda \in (0, 1)$ ,  $L = \sum_{i=1}^N L_i$ , and  $\Gamma$  denotes the number of iterations.

Theorem 8 gives the convergence rate of our proposed algorithm for local convex functions, which is  $O(\ln \Gamma/\sqrt{\Gamma})$ . Further, by Theorems 7 and 8, we can see that the time-varying and directed network topology does not affect the convergence rate.

In this section, we show that our proposed algorithm is asymptotically convergent. For strongly convex functions, we derive a convergence rate  $O(\ln \Gamma/\Gamma)$  for our algorithm over time-varying directed networks. Furthermore, we also present that our algorithm converges with rate  $O(\ln \Gamma/\sqrt{\Gamma})$  under general convexity. In the next section, we will give the detailed proofs of main results.

#### 4. Analysis of Convergence Results

We give relevant proof for our main results, namely, Theorems 6, 7, and 8. For this purpose, a basic iterate relation is established. Then, we use the basic relation to prove our main results.

For the sake of analysis, we first describe the scalar version of (4)-(5), where the variables  $u_i(\tau)$  and  $v_i(\tau)$  are scalar variables, for all  $i \in \mathcal{V}$ . Thus, the estimate  $u_i(\tau)$  of agent  $i$  is updated by

$$\begin{aligned} v_i(\tau) &= (1 - w_i(\tau) d_i^{\text{out}}(\tau)) u_i(\tau) \\ & \quad + \sum_{j \in \mathcal{N}_i^{\text{in}}(\tau)} w_j(\tau) u_j(\tau), \end{aligned} \quad (14)$$

$$u_i(\tau+1) = \Pi_{\mathcal{U}} [v_i(\tau) - \alpha(\tau) s_i(\tau)], \quad (15)$$

for  $\tau \geq 0$  and all  $i \in \mathcal{V}$ , where  $s_i(k) = \nabla g_i(u_i(\tau)) + \epsilon_i(\tau)$  and  $\mathcal{U} \subseteq \mathbb{R}$  denotes a constraint set.

To facilitate analysis, we can rewrite algorithms (14)-(15) in a compact form, namely,

$$v(\tau) = W(\tau) u(\tau), \quad (16)$$

$$u(\tau + 1) = \Pi_{\mathcal{U}} [v(\tau) - \alpha(\tau) s(\tau)], \quad (17)$$

where the matrix  $W(\tau)$  is defined as follows: for any  $i$  and  $j \in \mathcal{N}_i^{\text{in}}(\tau)$ ,  $[W(\tau)]_{ij} = w_j(\tau)$ ,  $[W(\tau)]_{ii} = 1 - w_i(\tau) d_i^{\text{out}}(\tau)$  at each round  $\tau$ . Moreover, we also define the following variables  $u(\tau)$  and  $s(\tau)$ :

$$u(\tau) \triangleq [u_1(\tau), \dots, u_N(\tau)]^\top, \quad (18)$$

$$s(\tau) \triangleq [s_1(\tau), \dots, s_N(\tau)]^\top.$$

To analyze our proposed algorithm, we rewrite algorithms (16)-(17) as follows:

$$v(\tau) = W(\tau) u(\tau), \quad (19)$$

$$u(\tau + 1) = v(\tau) - \alpha(\tau) s(\tau) + \delta(\tau),$$

$$\begin{aligned} \delta(\tau) &= \Pi_{\mathcal{U}} [v(\tau) - \alpha(\tau) g(\tau)] \\ &\quad - (v(\tau) - \alpha(\tau) s(\tau)). \end{aligned} \quad (20)$$

Hence, let  $r(\tau) = \delta(\tau) - \alpha(\tau)s(\tau)$ , which can be referred as the perturbation. Then, following from (19)-(20), we have

$$\begin{aligned} u(\tau + 1) &= [W(\tau)W(\tau - 1) \cdots W(0)] u(0) \\ &\quad + \sum_{t=0}^{\tau-1} [W(\tau) \cdots W(t+1)] r(t) + r(\tau) \end{aligned} \quad (21)$$

for any  $t$  and  $\tau$  with  $\tau \geq t$ . Furthermore, we introduce the following matrix:

$$\Delta(\tau : t) \triangleq W(\tau) \cdots W(t) \quad (22)$$

for any  $t$  and  $\tau$  with  $\tau \geq t$ , and  $\Delta(\tau, \tau) = W(\tau)$ . Hence, from (21), we have

$$u(\tau + 1) = \Delta(\tau : 0) u(0) + \sum_{t=0}^{\tau} \Delta(\tau : t+1) r(t) \quad (23)$$

for any  $t$  and  $\tau$  with  $\tau \geq t$ . Moreover, let  $\Delta(\tau : \tau + 1) = I$  for convenience.

Moreover, since  $W(\tau)$  is a sequence of column stochastic, which is proved in Lemma 9, we have  $\mathbf{1}^\top W(\tau) = \mathbf{1}^\top$ . Hence, following from (23), we have

$$\mathbf{1}^\top u(\tau + 1) = \mathbf{1}^\top u(0) + \sum_{t=0}^{\tau} \mathbf{1}^\top r(t) \quad \text{for all } \tau \geq 0. \quad (24)$$

Since the matrices  $W(\tau)$  and  $\Delta(\tau : t)$  are crucial in convergence analysis of our proposed algorithm, we first provide some properties of these matrices.

**Lemma 9.** *We assume that Assumption 2 holds. Then, for any  $\tau \geq 0$ , the matrix  $W(\tau)$  is column stochastic. Moreover, for all*

*$i, j \in \mathcal{V}$ , there exist some positive constants  $C$  and  $\lambda \in (0, 1)$ , and the matrix  $\Delta(\tau : t)$  satisfies*

$$\left| [\Delta(\tau : t)]_{ij} - \frac{1}{N} \right| \leq C \lambda^{\tau-t+1} \quad (25)$$

for any  $t$  and  $\tau$  with  $\tau \geq t$ .

According to [24] (the proofs of Lemma 3 and Proposition 1 in [24]), we immediately obtain this lemma.

In addition, we also give the properties of the projection operator  $\Pi_{\mathcal{U}}[\cdot]$ . Following from [14], we have the following.

**Lemma 10.** *Assume that constraint set  $\mathcal{U}$  is closed and convex in  $\mathbb{R}^d$ . Furthermore, we assume that  $\mathcal{U}$  is nonempty. Thus, we conclude that for any  $\mathbf{u} \in \mathbb{R}^d$ ,*

- (1)  $(\Pi_{\mathcal{U}}[\mathbf{u}] - \mathbf{u})^\top (\mathbf{v} - \Pi_{\mathcal{U}}[\mathbf{u}]) \geq 0$ , for all  $\mathbf{v} \in \mathcal{U}$
- (2)  $\|\Pi_{\mathcal{U}}[\mathbf{u}] - \Pi_{\mathcal{U}}[\mathbf{v}]\| \leq \|\mathbf{u} - \mathbf{v}\|$  for any  $\mathbf{v} \in \mathbb{R}^d$
- (3)  $(\Pi_{\mathcal{U}}[\mathbf{u}] - \mathbf{u})^\top (\mathbf{u} - \mathbf{v}) \leq -\|\Pi_{\mathcal{U}}[\mathbf{u}] - \mathbf{u}\|^2$ , for all  $\mathbf{v} \in \mathcal{U}$
- (4)  $\|\Pi_{\mathcal{U}}[\mathbf{u}] - \mathbf{v}\|^2 + \|\Pi_{\mathcal{U}}[\mathbf{u}] - \mathbf{u}\|^2 \leq \|\mathbf{u} - \mathbf{v}\|^2$ , for all  $\mathbf{v} \in \mathcal{U}$ .

Besides, we present some auxiliary results as follows, which are used to prove the relative conclusions.

**Lemma 11** (see [30]). *We assume that  $\rho_n$  is a scalar for all positive integer  $n$ .*

- (i) If  $\lim_{n \rightarrow \infty} \rho_n = \rho$ , then we have  $\lim_{n \rightarrow \infty} \sum_{m=0}^n \theta^{n-m} \rho_m = \rho / (1 - \theta)$  for any  $\theta \in (0, 1)$ .
- (ii) If  $\sum_{n=0}^{\infty} \rho_n < \infty$  and  $\rho_n \geq 0$ , then, for  $\theta \in (0, 1)$ , we have  $\sum_{n=0}^{\infty} (\sum_{m=0}^n \theta^{n-m} \rho_m) < \infty$ .

In order to obtain a convergence result of algorithms (16)-(17), we first establish the following lemma.

**Lemma 12.** *We assume that Assumptions 2-3 hold. Let estimate sequence  $\{u_i(\tau)\}$ ,  $i \in \mathcal{V}$ , be generated by (16)-(17). Moreover, for  $\tau > t \geq 1$ , the positive sequence  $\{\alpha(\tau)\}$  satisfies the following decay conditions:*

$$\begin{aligned} \sum_{\tau=1}^{\infty} \alpha(\tau) &= \infty, \\ \sum_{\tau=1}^{\infty} \alpha^2(\tau) &< \infty, \end{aligned} \quad (26)$$

$$\alpha(\tau) \leq \alpha(t).$$

- (a) For  $i = 1, \dots, N$ , we conclude that

$$\begin{aligned} \left| u_i(\tau) - \frac{\mathbf{1}^\top u(\tau)}{N} \right| \\ \leq C \left( \lambda^\tau \|u(0)\|_1 + \sum_{t=0}^{\tau-1} \lambda^{\tau-t-1} \|r(t)\|_1 \right). \end{aligned} \quad (27)$$

- (b) If  $\lim_{\tau \rightarrow \infty} r_i(\tau) = 0$  for  $i = 1, \dots, N$ , we obtain

$$\lim_{\tau \rightarrow \infty} \left| u_i(\tau) - \frac{\mathbf{1}^\top u(\tau)}{N} \right| = 0. \quad (28)$$

(c) If  $\sum_{\tau=1}^{\infty} \alpha(\tau) |r_i(\tau)| < \infty$ , then, we obtain

$$\sum_{\tau=1}^{\infty} \alpha(\tau) \left| u_i(\tau) - \frac{\mathbf{1}^\top \mathbf{u}(\tau)}{N} \right| < \infty, \quad (29)$$

for all  $i = 1, \dots, N$ .

*Proof.* (a) From (23) and (24), we have

$$\begin{aligned} \left| u_i(\tau) - \frac{\mathbf{1}^\top \mathbf{u}(\tau)}{N} \right| &= \left| [\Delta(\tau-1:0) \mathbf{u}(0)]_i \right. \\ &\quad \left. - \sum_{t=0}^{\tau-1} [\Delta(\tau-1:t+1) r(t)]_i - \frac{1}{N} \mathbf{1}^\top \mathbf{u}(0) \right. \\ &\quad \left. + \frac{1}{N} \sum_{t=0}^{\tau-1} \mathbf{1}^\top r(t) \right| \leq \max_j \left| [\Delta(\tau-1:0)]_{ij} - \frac{1}{N} \right| \\ &\quad \times \|\mathbf{u}(0)\|_1 \\ &\quad + \sum_{t=0}^{\tau-1} \|r(t)\|_1 \max_j \left| [\Delta(\tau-1:t+1)]_{ij} - \frac{1}{N} \right| \\ &\leq C \lambda^\tau \|\mathbf{u}(0)\|_1 + \sum_{t=0}^{\tau-1} C \lambda^{\tau-t-1} \|r(t)\|_1, \end{aligned} \quad (30)$$

where we use Hölder's inequality to obtain the first inequality, and we also obtain the last inequality by Lemma 9. Hence, this conclusion of part (a) is obtained.

(b) Since  $\lambda \in (0, 1)$ , following from preceding relation (a) and letting  $\tau \rightarrow \infty$ , thus

$$\lim_{\tau \rightarrow \infty} \left| u_i(\tau) - \frac{\mathbf{1}^\top \mathbf{u}(\tau)}{N} \right| \leq \lim_{\tau \rightarrow \infty} \sum_{t=0}^{\tau-1} C \lambda^{\tau-t-1} \|r(t)\|_1. \quad (31)$$

Since  $r_i(\tau) \rightarrow 0$  for  $i \in \mathcal{V}$ , so  $\|r(\tau)\|_1 \rightarrow 0$  as  $\tau \rightarrow \infty$ . Hence, by Lemma 11(i), we obtain

$$\lim_{\tau \rightarrow \infty} \sum_{t=0}^{\tau-1} C \lambda^{\tau-t-1} \|r(t)\|_1 = 0. \quad (32)$$

Thus, we conclude the statement of (b).

(c) By the decay conditions of  $\alpha(\tau)$ , we obtain that

$$\begin{aligned} \alpha(\tau) \left| u_i(\tau) - \frac{\mathbf{1}^\top \mathbf{u}(\tau)}{N} \right| \\ \leq C \left( \alpha(1) \lambda^\tau \|\mathbf{u}(0)\|_1 + \sum_{t=0}^{\tau-1} \alpha(t) \lambda^{\tau-t-1} \|r(t)\|_1 \right). \end{aligned} \quad (33)$$

Since  $\lambda \in (0, 1)$ , we can see that  $\sum_{\tau=0}^{\infty} \lambda^\tau$  is finite. Hence, following from Lemma 11(ii), the sum  $\sum_{\tau=1}^{\infty} \sum_{t=0}^{\tau-1} \lambda^{\tau-t-1} \alpha(t) \|r(t)\|_1$  is finite. Hence, the relation in part (c) is obtained.  $\square$

Informally, the distributed stochastic subgradient algorithm based on weight-balancing ensures that all  $u_i(\tau)$  track the running average  $\mathbf{1}^\top \mathbf{u}(\tau)/N$  with a geometric rate  $\lambda$ .

Following from the similar arguments as (19)-(20), we rewrite algorithms (4)-(5) as

$$\mathbf{v}(\tau) = \mathcal{W}(\tau) \mathbf{u}(\tau), \quad (34)$$

$$\mathbf{u}(\tau+1) = \mathbf{v}(\tau) - \alpha(\tau) \mathbf{s}(\tau) + \boldsymbol{\delta}(\tau),$$

$$\begin{aligned} \boldsymbol{\delta}_i(\tau) &= \Pi_{\mathcal{U}} [\mathbf{v}_i(\tau) - \alpha(\tau) \mathbf{s}_i(\tau)] \\ &\quad - (\mathbf{v}_i(\tau) - \alpha(\tau) \mathbf{s}_i(\tau)), \end{aligned} \quad (35)$$

where  $\mathcal{W}(\tau) = W(\tau) \otimes I_d$  and

$$\begin{aligned} \mathbf{u}(\tau) &\triangleq [\mathbf{u}_1(\tau)^\top, \dots, \mathbf{u}_N(\tau)^\top]^\top, \\ \mathbf{v}(\tau) &\triangleq [\mathbf{v}_1(\tau)^\top, \dots, \mathbf{v}_N(\tau)^\top]^\top, \\ \mathbf{s}(\tau) &\triangleq [\mathbf{s}_1(\tau)^\top, \dots, \mathbf{s}_N(\tau)^\top]^\top, \\ \boldsymbol{\delta}(\tau) &\triangleq [\boldsymbol{\delta}_1(\tau)^\top, \dots, \boldsymbol{\delta}_N(\tau)^\top]^\top. \end{aligned} \quad (36)$$

Moreover, let  $\mathbf{r}_i(\tau) = \boldsymbol{\delta}_i(\tau) - \alpha(\tau) \mathbf{s}_i(\tau)$ , for all  $i = 1, \dots, N$ , which is referred as perturbation vector. Hence, we introduce the following vector:

$$\mathbf{r}(\tau) \triangleq [\mathbf{r}_1(\tau)^\top, \dots, \mathbf{r}_N(\tau)^\top]^\top. \quad (37)$$

Note that the vectors  $\mathbf{u}(\tau)$ ,  $\mathbf{v}(\tau)$ ,  $\mathbf{s}(\tau)$ ,  $\boldsymbol{\delta}(\tau)$ , and  $\mathbf{r}(\tau)$  are in  $\mathbb{R}^{Nd}$ . Besides, the following corollary is obtained immediately from Lemma 12(a).

**Corollary 13.** *For every agent  $i$ ,  $i \in \mathcal{V}$ , the vector variables  $\mathbf{u}_i(\tau)$  and  $\mathbf{s}_i(\tau)$  in  $\mathbb{R}^d$  are given in (4)-(5); then, we have*

$$\begin{aligned} \left\| \mathbf{u}_i(\tau) - \frac{\sum_{j=1}^N \mathbf{u}_j(\tau)}{N} \right\| \\ \leq C \left( \lambda^\tau \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 + \sum_{t=0}^{\tau-1} \lambda^{\tau-t-1} \sum_{j=1}^N \|\mathbf{r}_j(t)\|_1 \right). \end{aligned} \quad (38)$$

For any vector, 1-norm is greater than or equal to standard Euclidean norm. Therefore, from Lemma 12(a), which is applied to each coordinate of  $\mathbb{R}^d$ , Corollary 13 holds immediately.

In Lemma 12 and Corollary 13, the perturbation term is a crucial role. Therefore, we bound the perturbation term in the following lemma.

**Lemma 14.** *Under Assumptions 2-5, the perturbations vectors  $\mathbf{r}_i(\tau) \in \mathbb{R}^d$ , for all  $i \in \mathcal{V}$ , satisfy*

$$\mathbb{E} [\|\mathbf{r}_i(\tau)\|_1] \leq 2\sqrt{d} \alpha(\tau) (L_i + \nu_i). \quad (39)$$

*Proof.* Following from Lemma 10(4), we have

$$\begin{aligned} \|\mathbf{u}_i(\tau+1) - \mathbf{v}_i(\tau)\|^2 \\ \leq \|\mathbf{v}_i(\tau) - \alpha(\tau) \mathbf{s}_i(\tau) - \mathbf{v}_i(\tau)\|^2 \\ - \|\mathbf{u}_i(\tau+1) - (\mathbf{v}_i(\tau) - \alpha(\tau) \mathbf{s}_i(\tau))\|^2. \end{aligned} \quad (40)$$

Furthermore, by Assumptions 4 and 5, we have

$$\begin{aligned} & \mathbb{E} \left[ \|\mathbf{u}_i(\tau+1) - \mathbf{v}_i(\tau)\|^2 \mid \mathcal{F}_\tau \right] \\ & \leq \alpha^2(\tau) (L_i + \nu_i)^2 - \mathbb{E} \left[ \|\boldsymbol{\delta}_i(\tau)\|^2 \right], \end{aligned} \quad (41)$$

for all  $i \in \mathcal{V}$ . Hence, we obtain with probability 1

$$\mathbb{E} \left[ \|\boldsymbol{\delta}_i(\tau)\|^2 \right] \leq \alpha^2(\tau) (L_i + \nu_i)^2. \quad (42)$$

Further, following from the inequality  $\mathbb{E}[\|\mathbf{u}\|] \leq \sqrt{\mathbb{E}[\|\mathbf{u}\|^2]}$ , then

$$\mathbb{E} \left[ \|\boldsymbol{\delta}_i(\tau)\| \right] \leq \alpha(\tau) (L_i + \nu_i). \quad (43)$$

According to the definition of perturbations vector  $\mathbf{r}_i(\tau)$ , we have

$$\begin{aligned} \mathbb{E} \left[ \|\mathbf{r}_i(\tau)\|_1 \right] & \leq \sqrt{d} \mathbb{E} \left[ \|\mathbf{r}_i(\tau)\| \right] \\ & \leq \sqrt{d} \left( \mathbb{E} \left[ \|\boldsymbol{\delta}_i(\tau)\| \right] + \alpha(\tau) \mathbb{E} \left[ \|\mathbf{s}_i(\tau)\| \right] \right) \\ & \leq 2\sqrt{d}\alpha(\tau) (L_i + \nu_i). \end{aligned} \quad (44)$$

Therefore, this lemma is proved completely.  $\square$

To establish the convergence properties of our algorithm, an auxiliary variable is first defined, for all  $\tau \geq 0$ , as follows:

$$\bar{\mathbf{u}}(\tau) = \frac{1}{N} \sum_{i=1}^N \mathbf{u}_i(\tau). \quad (45)$$

We now establish the recurrence relation for  $\bar{\mathbf{u}}(\tau)$ . Hence, for all  $i \in \mathcal{V}$  and  $\ell \in \mathcal{D}$ , where  $\mathcal{D} = \{1, \dots, d\}$ , we introduce a vector  $\mathbf{u}^\ell(\tau) \in \mathbb{R}^N$  such that  $[u^\ell(\tau)]_i = [\mathbf{u}_i(\tau)]_\ell$ . Following from algorithms (4)-(5), we obtain for  $\ell \in \mathcal{D}$

$$\mathbf{u}^\ell(\tau+1) = W(\tau) \mathbf{u}^\ell(\tau) + \mathbf{r}^\ell(\tau), \quad (46)$$

where  $\mathbf{r}^\ell(\tau)$  is a vector in  $\mathbb{R}^N$ , which is defined as  $[r^\ell(\tau)]_i = [\mathbf{r}_i(\tau)]_\ell$ , for all  $i \in \mathcal{V}$  and  $\ell \in \mathcal{D}$ .

Since the matrix  $W(\tau)$  is column stochastic, we have, for all  $\ell \in \mathcal{D}$ ,

$$\frac{1}{N} \sum_{i=1}^N [u^\ell(\tau+1)]_i = \frac{1}{N} \sum_{i=1}^N [u^\ell(\tau)]_i + \frac{1}{N} \sum_{i=1}^N [r^\ell(\tau)]_i. \quad (47)$$

We follow from the definition of  $[u^\ell(\tau)]_i$ ; then,

$$\bar{\mathbf{u}}(\tau+1) = \bar{\mathbf{u}}(\tau) + \frac{1}{N} \sum_{i=1}^N \mathbf{r}_i(\tau). \quad (48)$$

We also establish the following lemma, which is crucial in our proofs.

**Lemma 15** (basic iteration relation). *Let the estimate sequence  $\{\mathbf{u}_i(\tau)\}$ ,  $i \in \mathcal{V}$ , be generated by (4)-(5). Under Assumptions*

2-5, for all vectors  $\mathbf{w} \in \mathcal{U}$  and  $\tau \geq 0$ , then we obtain with probability 1

$$\begin{aligned} & \mathbb{E} \left[ \|\bar{\mathbf{u}}(\tau+1) - \mathbf{w}\|^2 \mid \mathcal{F}_\tau \right] \\ & \leq \|\bar{\mathbf{u}}(\tau) - \mathbf{w}\|^2 + \frac{4\alpha(\tau)}{N} \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ & \quad + \frac{2}{N} \sum_{i=1}^N \boldsymbol{\delta}_i(\tau)^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}) \\ & \quad + \frac{4\alpha^2(\tau)}{N} \sum_{i=1}^N (L_i + \nu_i)^2 \\ & \quad - \frac{2\alpha(\tau)}{N} (g(\bar{\mathbf{u}}(\tau)) - g(\mathbf{w})) \\ & \quad - \frac{\alpha(\tau)}{N} \sum_{i=1}^N \sigma_i \|\mathbf{u}_i(\tau) - \mathbf{w}\|^2, \end{aligned} \quad (49)$$

where  $L_i$  and  $\nu_i$  are defined as Assumptions 2-3, respectively. The constant  $\sigma_i$  is defined in (3).

*Proof.* By (48), we have

$$\begin{aligned} \|\bar{\mathbf{u}}(\tau+1) - \mathbf{w}\|^2 & = \left\| \bar{\mathbf{u}}(\tau) + \frac{1}{N} \sum_{i=1}^N \mathbf{r}_i(\tau) - \mathbf{w} \right\|^2 \\ & = \|\bar{\mathbf{u}}(\tau) - \mathbf{w}\|^2 \\ & \quad + \frac{2}{N} \sum_{i=1}^N \mathbf{r}_i(\tau)^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}) \\ & \quad + \frac{1}{N^2} \left\| \sum_{i=1}^N \mathbf{r}_i(\tau) \right\|^2 \\ & = \|\bar{\mathbf{u}}(\tau) - \mathbf{w}\|^2 \\ & \quad - \frac{2\alpha(\tau)}{N} \sum_{i=1}^N \mathbf{s}_i(\tau)^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}) \\ & \quad + \frac{2}{N} \sum_{i=1}^N \boldsymbol{\delta}_i(\tau)^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}) \\ & \quad + \frac{1}{N^2} \left\| \sum_{i=1}^N \mathbf{r}_i(\tau) \right\|^2. \end{aligned} \quad (50)$$

By taking expectation on the both sides of (50) with respect to  $\mathcal{F}_\tau$ , furthermore, we follow from (6) and the fact

$$\mathbb{E} [\boldsymbol{\epsilon}_i(\tau) \mid \mathcal{F}_\tau] = 0, \quad (51)$$

and we have that

$$\begin{aligned}
& \mathbb{E} \left[ \|\bar{\mathbf{u}}(\tau+1) - \mathbf{w}\|^2 \mid \mathcal{F}_\tau \right] \\
& \leq \|\bar{\mathbf{u}}(\tau) - \mathbf{w}\|^2 + \frac{2}{N} \sum_{i=1}^N \boldsymbol{\delta}_i(\tau)^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}) \\
& \quad + \frac{1}{N^2} \mathbb{E} \left[ \left\| \sum_{i=1}^N \mathbf{r}_i(\tau) \right\|^2 \right] \\
& \quad - \frac{2\alpha(\tau)}{N} \sum_{i=1}^N \nabla g_i(\mathbf{u}_i(\tau))^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}).
\end{aligned} \tag{52}$$

Next, we bound the term  $\mathbb{E}[\|\sum_{i=1}^N \mathbf{r}_i(\tau)\|^2]$  in (52) as follows:

$$\mathbb{E} \left[ \left\| \sum_{i=1}^N \mathbf{r}_i(\tau) \right\|^2 \right] \leq N \sum_{i=1}^N \mathbb{E} \left[ \|\mathbf{r}_i(\tau)\|^2 \right], \tag{53}$$

where we use the inequality  $(\sum_{i=1}^N a_i)^2 \leq N \sum_{i=1}^N a_i^2$  to obtain the above relation. In order to obtain the upper bound of  $\mathbb{E}[\|\sum_{i=1}^N \mathbf{r}_i(\tau)\|^2]$ , we need to bound the term  $\mathbb{E}[\|\mathbf{r}_i(\tau)\|^2]$  in (53). Following from the similar arguments as Lemma 14, we have

$$\mathbb{E} \left[ \|\mathbf{r}_i(\tau)\|^2 \right] \leq 4\alpha^2(\tau) (L_i + \nu_i)^2, \tag{54}$$

for all  $i \in \mathcal{V}$ . Hence, combining (53) and (54), we obtain

$$\mathbb{E} \left[ \left\| \sum_{i=1}^N \mathbf{r}_i(\tau) \right\|^2 \right] \leq 4N\alpha^2(\tau) \sum_{i=1}^N (L_i + \nu_i)^2. \tag{55}$$

Substituting (55) into (52), we have

$$\begin{aligned}
& \mathbb{E} \left[ \|\bar{\mathbf{u}}(\tau+1) - \mathbf{w}\|^2 \mid \mathcal{F}_\tau \right] \\
& \leq \|\bar{\mathbf{u}}(\tau) - \mathbf{w}\|^2 + \frac{2}{N} \sum_{i=1}^N \boldsymbol{\delta}_i(\tau)^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}) \\
& \quad + \frac{4\alpha^2(\tau)}{N} \sum_{i=1}^N (L_i + \nu_i)^2 \\
& \quad - \frac{2\alpha(\tau)}{N} \sum_{i=1}^N \nabla g_i(\mathbf{u}_i(\tau))^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}).
\end{aligned} \tag{56}$$

We next estimate the term  $\nabla g_i(\mathbf{u}_i(\tau))^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w})$  in (56). By (3) in Assumption 3, we have that

$$\begin{aligned}
& \nabla g_i(\mathbf{u}_i(\tau))^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}) \\
& = \nabla g_i(\mathbf{u}_i(\tau))^\top (\bar{\mathbf{u}}(\tau) - \mathbf{u}_i(\tau)) \\
& \quad + \nabla g_i(\mathbf{u}_i(\tau))^\top (\mathbf{u}_i(\tau) - \mathbf{w}) \\
& \geq \nabla g_i(\mathbf{u}_i(\tau))^\top (\bar{\mathbf{u}}(\tau) - \mathbf{u}_i(\tau)) + g_i(\mathbf{u}_i(\tau)) \\
& \quad - g_i(\mathbf{w}) + \frac{\sigma_i}{2} \|\mathbf{u}_i(\tau) - \mathbf{w}\|^2.
\end{aligned} \tag{57}$$

Further, we also obtain

$$\nabla g_i(\mathbf{u}_i(\tau))^\top (\bar{\mathbf{u}}(\tau) - \mathbf{u}_i(\tau)) \geq -L_i \|\bar{\mathbf{u}}(\tau) - \mathbf{u}_i(\tau)\|, \tag{58}$$

where we use Cauchy-Schwarz inequality to obtain the last inequality. Moreover, the term  $g_i(\mathbf{u}_i(\tau)) - g_i(\mathbf{w})$  in (57) can be rewritten as

$$\begin{aligned}
g_i(\mathbf{u}_i(\tau)) - g_i(\mathbf{w}) & = g_i(\bar{\mathbf{u}}(\tau)) - g_i(\mathbf{w}) + g_i(\mathbf{u}_i(\tau)) \\
& \quad - g_i(\bar{\mathbf{u}}(\tau)).
\end{aligned} \tag{59}$$

By using Assumptions 3 and 4, we have that

$$\begin{aligned}
g_i(\mathbf{u}_i(\tau)) - g_i(\mathbf{w}) & \geq g_i(\bar{\mathbf{u}}(\tau)) - g_i(\mathbf{w}) \\
& \quad + \nabla g_i(\bar{\mathbf{u}}(\tau))^\top (\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)) \\
& \geq -L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| + g_i(\bar{\mathbf{u}}(\tau)) \\
& \quad - g_i(\mathbf{w}).
\end{aligned} \tag{60}$$

Hence, combining (57), (58), and (60), we obtain

$$\begin{aligned}
& \nabla g_i(\mathbf{u}_i(\tau))^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}) \\
& \geq g_i(\bar{\mathbf{u}}(\tau)) - g_i(\mathbf{w}) + \frac{\sigma_i}{2} \|\mathbf{u}_i(\tau) - \mathbf{w}\|^2 \\
& \quad - 2L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\|.
\end{aligned} \tag{61}$$

Substituting (61) into (56), we have

$$\begin{aligned}
& \mathbb{E} \left[ \|\bar{\mathbf{u}}(\tau+1) - \mathbf{w}\|^2 \mid \mathcal{F}_\tau \right] \\
& \leq \|\bar{\mathbf{u}}(\tau) - \mathbf{w}\|^2 + \frac{4\alpha(\tau)}{N} \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\
& \quad + \frac{2}{N} \sum_{i=1}^N \boldsymbol{\delta}_i(\tau)^\top (\bar{\mathbf{u}}(\tau) - \mathbf{w}) \\
& \quad + \frac{4\alpha^2(\tau)}{N} \sum_{i=1}^N (L_i + \nu_i)^2 \\
& \quad - \frac{2\alpha(\tau)}{N} \sum_{i=1}^N (g_i(\bar{\mathbf{u}}(\tau)) - g_i(\mathbf{w})) \\
& \quad - \frac{2\alpha(\tau)}{N} \sum_{i=1}^N \frac{\sigma_i}{2} \|\mathbf{u}_i(\tau) - \mathbf{w}\|^2.
\end{aligned} \tag{62}$$

Following from the definition of the global cost function  $g(\mathbf{u})$ , and then using (62), the lemma is completely proved.  $\square$

To prove Theorem 6, we employ the following lemma, which is given by [31].

**Lemma 16.** *We assume that  $\{v_\tau\}$ ,  $\{u_\tau\}$ ,  $\{a_\tau\}$ , and  $\{b_\tau\}$  are random sequences. Moreover, we further suppose that elements of these sequences are nonnegative random variables and with probability 1 satisfy the following relation:*

$$\mathbb{E} [v_{\tau+1} \mid \mathcal{F}_\tau] \leq (1 + a_\tau) v_\tau - u_\tau + b_\tau \tag{63}$$

for  $\tau \geq 0$ . Further, if  $\sum_{\tau=0}^{\infty} a_{\tau} < \infty$  and  $\sum_{\tau=0}^{\infty} b_{\tau} < \infty$  hold, then,  $\{v_{\tau}\}$  asymptotically converges to  $v$ , where  $v$  is a nonnegative random variable, i.e.,  $\lim_{\tau \rightarrow \infty} v_{\tau} = v$ . Moreover, we also have  $\sum_{\tau=0}^{\infty} u_{\tau} < \infty$  with probability 1.

Using the conclusion of Lemma 16, we have the following lemma.

**Lemma 17.** Consider an optimization problem  $\min_{u \in \mathbb{R}^n} h(u)$ . Assume that the set  $\mathcal{S}^*$  be nonempty, where  $\mathcal{S}^*$  is the optimal set. Let the function  $h : \mathbb{R}^n \rightarrow \mathbb{R}$  be continuous. Suppose that  $\{u_{\tau}\}$ ,  $\{a_{\tau}\}$ , and  $\{b_{\tau}\}$  are random sequences with nonnegative elements. Moreover, these sequences satisfy the following relation:

$$\mathbb{E} [\|u_{\tau+1} - u\|^2 \mid \mathcal{F}_{\tau}] \leq (1 + a_{\tau}) \|u_{\tau} - u\|^2 - \kappa_{\tau} (h(u_{\tau}) - h(u)) + b_{\tau} \quad (64)$$

for all  $u \in \mathcal{S}^*$  and  $\tau \geq 0$  with probability 1, where  $\kappa_{\tau} \geq 0$  for all  $\tau \geq 0$ . Further, we assume that  $\sum_{\tau=0}^{\infty} b_{\tau} < \infty$  and  $\sum_{\tau=0}^{\infty} \kappa_{\tau} = \infty$  hold. Then, the random sequence  $\{u_{\tau}\}$  converges to some optimal solutions  $u^* \in \mathcal{S}^*$  with probability 1.

*Proof.* Let  $u = u^*$  and  $h^* = \min_{u \in \mathbb{R}^n} h(u)$ ; then, we have

$$\begin{aligned} \mathbb{E} [\|u_{\tau+1} - u^*\|^2 \mid \mathcal{F}_{\tau}] \\ \leq (1 + a_{\tau}) \|u_{\tau} - u^*\|^2 - \kappa_{\tau} (h(u_{\tau}) - h^*) + b_{\tau}. \end{aligned} \quad (65)$$

Thus, we satisfy the conditions of Lemma 16. Hence, the sequence  $\{\|u_{\tau+1} - u^*\|^2\}$  is convergent, and we have

$$\sum_{\tau=0}^{\infty} \kappa_{\tau} (h(u_{\tau}) - h^*) < \infty. \quad (66)$$

Since  $\sum_{\tau=0}^{\infty} \kappa_{\tau} = \infty$ , with probability 1 we have

$$\liminf_{\tau \rightarrow \infty} h(u_{\tau}) = h^*. \quad (67)$$

Since the sequence  $\{\|u_{\tau+1} - u^*\|^2\}$  is convergent for  $\tau \geq 0$ , the sequence  $\{u_{\tau}\}$  is bounded. Therefore, there exists a convergent subsequence  $\{u_{\tau_j}\}$  of the sequence  $\{u_{\tau}\}$  that converges to some  $\bar{u}$ . Moreover, the subsequence  $\{u_{\tau_j}\}$  satisfies

$$\lim_{j \rightarrow \infty} h(u_{\tau_j}) = \liminf_{\tau \rightarrow \infty} h(u_{\tau}) = h^*. \quad (68)$$

By using continuity of  $h$ , we have

$$\lim_{j \rightarrow \infty} h(u_{\tau_j}) = h(\bar{u}). \quad (69)$$

Following from (68), we can see that the random sequence  $\{u_{\tau}\}$  asymptotically converges to some  $\bar{u} \in \mathcal{S}^*$ . Therefore, the random sequence  $\{u_{\tau}\}$  converges to  $u^*$  by setting  $u^* = \bar{u}$ .  $\square$

We now start to prove Theorem 6 by using the conclusions of Lemmas 15 and 17.

*Proof of Theorem 6.* Since  $\alpha(\tau) \rightarrow 0$  as  $\tau \rightarrow \infty$ , we obtain

$$\lim_{\tau \rightarrow \infty} \left\| \left[ u^{\ell}(\tau + 1) \right]_i - \frac{\mathbf{1}^{\top} u^{\ell}(\tau)}{N} \right\| = 0, \quad (70)$$

which follows from Lemma 10(2). Therefore, using (70), we have

$$\lim_{\tau \rightarrow \infty} \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| = 0. \quad (71)$$

We also have

$$\mathbb{E} [\|r^{\ell}(\tau)\|_i] \leq \mathbb{E} [\|r^{\ell}(\tau)\|_{\infty}] \leq \max_i \mathbb{E} [\|\mathbf{r}_i(\tau)\|], \quad (72)$$

where Assumptions 2 and 3 are used. Thus, following from the decay conditions of  $\alpha(\tau)$ , we have

$$\sum_{\tau=1}^{\infty} \alpha(\tau) \|\mathbf{r}_i(\tau)\| \leq \max_i 2(L_i + \nu_i) \sum_{\tau=1}^{\infty} \alpha^2(\tau) < \infty \quad (73)$$

for  $i \in \mathcal{V}$ . Following from Lemma 10(3), we obtain

$$\sum_{\tau=1}^{\infty} \alpha(\tau) \left\| \left[ u^{\ell}(\tau) \right]_i - \frac{\mathbf{1}^{\top} u^{\ell}(\tau)}{N} \right\| < \infty. \quad (74)$$

Hence, we obtain

$$\sum_{\tau=1}^{\infty} \alpha(\tau) \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| < \infty. \quad (75)$$

Let  $\mathbf{w} = \mathbf{u}^*$  for  $\mathbf{u}^* \in \mathcal{U}^*$  in Lemma 15, and we have

$$\begin{aligned} \mathbb{E} [\|\bar{\mathbf{u}}(\tau + 1) - \mathbf{u}^*\|^2 \mid \mathcal{F}_{\tau}] \\ \leq \|\bar{\mathbf{u}}(\tau) - \mathbf{u}^*\|^2 + \frac{4\alpha(\tau)}{N} \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ + \frac{2}{N} \sum_{i=1}^N \delta_i(\tau)^{\top} (\bar{\mathbf{u}}(\tau) - \mathbf{u}^*) \\ + \frac{4\alpha^2(\tau)}{N} \sum_{i=1}^N (L_i + \nu_i)^2 \\ - \frac{2\alpha(\tau)}{N} (g(\bar{\mathbf{u}}(\tau)) - g(\mathbf{u}^*)). \end{aligned} \quad (76)$$

According to the conclusion of Lemma 10(1), we obtain

$$\delta_i(\tau)^{\top} (\bar{\mathbf{u}}(\tau) - \mathbf{u}^*) \leq 0. \quad (77)$$

Hence, we have with probability one

$$\begin{aligned} \mathbb{E} [\|\bar{\mathbf{u}}(\tau + 1) - \mathbf{u}^*\|^2 \mid \mathcal{F}_{\tau}] \\ \leq \|\bar{\mathbf{u}}(\tau) - \mathbf{u}^*\|^2 + \frac{4\alpha(\tau)}{N} \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ + \frac{4\alpha^2(\tau)}{N} \sum_{i=1}^N (L_i + \nu_i)^2 \\ - \frac{2\alpha(\tau)}{N} (g(\bar{\mathbf{u}}(\tau)) - g(\mathbf{u}^*)). \end{aligned} \quad (78)$$

By using (75), we have

$$\sum_{\tau=1}^{\infty} \frac{4\alpha(\tau)}{N} \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| < \infty. \quad (79)$$

Since  $\sum_{\tau=1}^{\infty} \alpha(\tau) = \infty$  and  $\sum_{\tau=1}^{\infty} \alpha^2(\tau) < \infty$  hold by the assumption of  $\alpha(\tau)$ , the conditions of Lemma 17 hold. Following from Lemma 17, we can see that  $\{\bar{\mathbf{u}}(\tau)\}$  asymptotically converges to  $\mathbf{u}^*$ . Further, following from (71),  $\{\mathbf{u}_i(\tau)\}$  also asymptotically converges to  $\mathbf{u}^*$  for all  $i \in \mathcal{V}$ , i.e.,

$$\lim_{\tau \rightarrow \infty} \mathbf{u}_i(\tau) = \mathbf{u}^* \quad (80)$$

with the notion that probability 1 holds. Therefore, Theorem 6 is proved completely.  $\square$

Now, we establish the following lemma, which is an important relation for the proof of the Theorem 7.

**Lemma 18.** *Let the estimate sequence  $\{\mathbf{u}_i(\tau)\}$ ,  $i \in \mathcal{V}$ , be generated by (4)-(5), and  $\alpha(\tau) = \mu/(\tau + 1)$ , where the constant  $\mu$  satisfies (52). Under Assumptions 2-5, then we obtain with probability 1*

$$\begin{aligned} \mathbb{E} \left[ \sum_{\tau=1}^{\Gamma-1} \left\| \mathbf{u}_i(\tau) - \frac{\sum_{j=1}^N \mathbf{u}_j(\tau)}{N} \right\| \right] &\leq C \left( \frac{\lambda}{1-\lambda} \right. \\ &\cdot \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 \\ &\left. + \frac{\mu}{1-\lambda} \sum_{i=1}^N 2\sqrt{d}(L_i + \nu_i)(1 + \ln(\Gamma - 1)) \right) \end{aligned} \quad (81)$$

for all  $i \in \mathcal{V}$  and  $\Gamma \geq 2$ .

*Proof.* Following from Corollary 13, we obtain

$$\begin{aligned} \sum_{\tau=1}^{\Gamma-1} \left\| \mathbf{u}_i(\tau) - \frac{\sum_{j=1}^N \mathbf{u}_j(\tau)}{N} \right\| \\ \leq C \sum_{\tau=1}^{\Gamma-1} \lambda^\tau \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 + C \sum_{\tau=1}^{\Gamma-1} \sum_{t=0}^{\tau-1} \lambda^{\tau-t-1} \sum_{j=1}^N \|\mathbf{r}_j(t)\|_1. \end{aligned} \quad (82)$$

Using (82) and the result of Lemma 14, we obtain

$$\begin{aligned} \mathbb{E} \left[ \sum_{\tau=1}^{\Gamma-1} \left\| \mathbf{u}_i(\tau) - \frac{\sum_{j=1}^N \mathbf{u}_j(\tau)}{N} \right\| \right] \\ \leq C \left( \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 \sum_{\tau=1}^{\Gamma-1} \lambda^\tau \right. \\ \left. + \sum_{i=1}^N 2\sqrt{d}(L_i + \nu_i) \sum_{\tau=1}^{\Gamma-1} \sum_{t=0}^{\tau-1} \alpha(t) \lambda^{\tau-t-1} \right). \end{aligned} \quad (83)$$

For  $\lambda \in (0, 1)$ , we have  $\sum_{\tau=1}^{\Gamma-1} \lambda^\tau \leq \lambda/(1-\lambda)$ . Hence, the first term in (83) is bounded as follows:

$$\sum_{\tau=1}^{\Gamma-1} \lambda^\tau \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 \leq \frac{\lambda}{1-\lambda} \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1. \quad (84)$$

Next, we estimate the second term in (83). From the expression  $\alpha(\tau) = \mu/(\tau + 1)$ , we have

$$\begin{aligned} \sum_{\tau=1}^{\Gamma-1} \sum_{t=0}^{\tau-1} \alpha(t) \lambda^{\tau-t-1} &= \sum_{\tau=1}^{\Gamma-1} \sum_{t=0}^{\tau-1} \frac{\mu}{t+1} \lambda^{\tau-t-1} = \sum_{\tau=1}^{\Gamma-1} \sum_{t=1}^{\tau} \frac{\mu}{t} \lambda^{\tau-t} \\ &= \sum_{t=1}^{\Gamma-1} \sum_{\tau=t}^{\Gamma-1} \frac{\mu}{t} \lambda^{\tau-t} \leq \frac{\mu}{1-\lambda} \sum_{t=1}^{\Gamma-1} \frac{1}{t} \\ &= \frac{\mu}{1-\lambda} \left( 1 + \sum_{t=2}^{\Gamma-1} \frac{1}{t} \right) \\ &\leq \frac{\mu}{1-\lambda} \left( 1 + \int_1^{\Gamma-1} \frac{1}{t} dt \right) \\ &= \frac{\mu}{1-\lambda} (1 + \ln(\Gamma - 1)) \end{aligned} \quad (85)$$

for  $\Gamma \geq 2$ . Therefore, combining (83), (84), and (85), the result of this lemma is obtained.  $\square$

Next, we start to prove Theorem 7 by using the results of Lemmas 15 and 18.

*Proof of Theorem 7.* We assume that  $\sigma_i$  is a positive constant for all  $i \in \mathcal{V}$ . Hence, for each agent  $i \in \mathcal{V}$ , local cost function  $g_i$  is a  $\sigma_i$ -strongly convex function. Moreover, the function  $g(\mathbf{u}) = \sum_{i=1}^N g_i(\mathbf{u})$  has a unique global minimization solution  $\mathbf{u}^*$ . Hence, let  $\mathbf{w} = \mathbf{u}^*$  in Lemma 15, and we have

$$\begin{aligned} \mathbb{E} \left[ \|\bar{\mathbf{u}}(\tau + 1) - \mathbf{u}^*\|^2 \mid \mathcal{F}_\tau \right] \\ \leq \|\bar{\mathbf{u}}(\tau) - \mathbf{u}^*\|^2 + \frac{4\alpha(\tau)}{N} \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ + \frac{4\alpha^2(\tau)}{N} \sum_{i=1}^N (L_i + \nu_i)^2 \\ - \frac{2\alpha(\tau)}{N} (g(\bar{\mathbf{u}}(\tau)) - g(\mathbf{u}^*)) \\ - \frac{\alpha(\tau)}{N} \sum_{i=1}^N \sigma_i \|\mathbf{u}_i(\tau) - \mathbf{u}^*\|^2, \end{aligned} \quad (86)$$

where we employ inequality  $\delta_i(\tau)^\top (\bar{\mathbf{u}}(\tau) - \mathbf{u}^*) \leq 0$ . Next, we first bound the term  $g(\bar{\mathbf{u}}(\tau)) - g(\mathbf{u}^*)$  in (86). By using the fact that the function  $g_i$  is  $\sigma_i$ -strongly convex and  $\nabla g_i(\mathbf{u}^*) = 0$ , we have

$$g(\bar{\mathbf{u}}(\tau)) - g(\mathbf{u}^*) \geq \frac{1}{2} \left( \sum_{i=1}^N \sigma_i \right) \|\bar{\mathbf{u}}(\tau) - \mathbf{u}^*\|^2. \quad (87)$$

In addition, note that  $g$  is the strongly convex function. Hence, by using the boundedness of subgradient  $\nabla g(\mathbf{u})$ , we also have that

$$\begin{aligned} g(\bar{\mathbf{u}}(\tau)) - g(\mathbf{u}^*) &= g(\bar{\mathbf{u}}(\tau)) - g(\mathbf{u}_i(\tau)) + g(\mathbf{u}_i(\tau)) \\ &\quad - g(\mathbf{u}^*) \\ &\geq -L \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| + g(\mathbf{u}_i(\tau)) \\ &\quad - g(\mathbf{u}^*), \end{aligned} \quad (88)$$

where  $L \triangleq \sum_{i=1}^N L_i$ . Therefore, following from (87) and (88), we obtain

$$\begin{aligned} 2(g(\bar{\mathbf{u}}(\tau)) - g(\mathbf{u}^*)) &\geq \frac{1}{2} \left( \sum_{i=1}^N \sigma_i \right) \|\bar{\mathbf{u}}(\tau) - \mathbf{u}^*\|^2 \\ &\quad - L \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ &\quad + g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*). \end{aligned} \quad (89)$$

Therefore, substituting (89) into (86), we obtain with probability 1

$$\begin{aligned} &\mathbb{E} \left[ \|\bar{\mathbf{u}}(\tau+1) - \mathbf{u}^*\|^2 \mid \mathcal{F}_\tau \right] \\ &\leq \|\bar{\mathbf{u}}(\tau) - \mathbf{u}^*\|^2 + \frac{\alpha(\tau)}{N} L \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ &\quad - \frac{\alpha(\tau)}{N} \sum_{i=1}^N \sigma_i \|\mathbf{u}_i(\tau) - \mathbf{u}^*\|^2 \\ &\quad + \frac{4\alpha(\tau)}{N} \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ &\quad + \frac{4\alpha^2(\tau)}{N} \sum_{i=1}^N (L_i + \nu_i)^2 \\ &\quad - \frac{\alpha(\tau)}{2N} \left( \sum_{i=1}^N \sigma_i \right) \|\bar{\mathbf{u}}(\tau) - \mathbf{u}^*\|^2 \\ &\quad - \frac{\alpha(\tau)}{N} (g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*)). \end{aligned} \quad (90)$$

Following from the expression of  $\alpha(\tau) = \mu/(\tau+1)$ , where  $\mu$  is given by (11), we have that

$$\begin{aligned} &\mathbb{E} \left[ \|\bar{\mathbf{u}}(\tau+1) - \mathbf{u}^*\|^2 \mid \mathcal{F}_\tau \right] \leq \left( 1 - \frac{2}{\tau+1} \right) \|\bar{\mathbf{u}}(\tau) \\ &\quad - \mathbf{u}^*\|^2 - \frac{\mu}{N(\tau+1)} \left( g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*) \right) \\ &\quad + \sum_{i=1}^N \sigma_i \|\mathbf{u}_i(\tau) - \mathbf{u}^*\|^2 \Big) + \frac{\mu L}{N(\tau+1)} \|\mathbf{u}_i(\tau) \end{aligned}$$

$$\begin{aligned} &- \bar{\mathbf{u}}(\tau)\| + \frac{4\mu^2}{N(\tau+1)^2} \sum_{i=1}^N (L_i + \nu_i)^2 + \frac{4\mu}{N(\tau+1)} \\ &\quad \cdot \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\|. \end{aligned} \quad (91)$$

Multiplying the both sides of (91) by  $\tau(\tau+1)$ , we have

$$\begin{aligned} &(\tau+1)\tau \mathbb{E} \left[ \|\bar{\mathbf{u}}(\tau+1) - \mathbf{u}^*\|^2 \mid \mathcal{F}_\tau \right] \\ &\leq \frac{4\mu^2 \tau \sum_{i=1}^N (L_i + \nu_i)^2}{N(\tau+1)} \\ &\quad - \frac{\mu\tau}{N} \left( g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*) + \sum_{i=1}^N \sigma_i \|\mathbf{u}_i(\tau) - \mathbf{u}^*\|^2 \right) \\ &\quad + \frac{\mu L \tau}{N} \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ &\quad + \frac{4\mu\tau}{N} \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ &\quad + \tau(\tau-1) \|\bar{\mathbf{u}}(\tau) - \mathbf{u}^*\|^2. \end{aligned} \quad (92)$$

By iterating the expression of (92), we obtain

$$\begin{aligned} &\Gamma(\Gamma-1) \mathbb{E} \left[ \|\bar{\mathbf{u}}(\Gamma) - \mathbf{u}^*\|^2 \right] \\ &\leq -\frac{\mu}{N} \sum_{\tau=1}^{\Gamma-1} \tau \mathbb{E} [g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*)] \\ &\quad + \frac{\mu L}{N} \sum_{\tau=1}^{\Gamma-1} \tau \mathbb{E} [\|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\|] \\ &\quad - \frac{\mu}{N} \sum_{\tau=1}^{\Gamma-1} \tau \sum_{i=1}^N \sigma_i \mathbb{E} [\|\mathbf{u}_i(\tau) - \mathbf{u}^*\|^2] \\ &\quad + \frac{4\mu}{N} \sum_{\tau=1}^{\Gamma-1} \tau \sum_{i=1}^N L_i \mathbb{E} [\|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\|] \\ &\quad + \frac{4\mu^2}{N} \sum_{i=1}^N (L_i + \nu_i)^2 \sum_{\tau=1}^{\Gamma-1} \frac{\tau}{\tau+1} \end{aligned} \quad (93)$$

for all  $i \in \mathcal{V}$  and  $\Gamma \geq 2$ . Due to  $\tau \leq \Gamma-1$ , so the relations

$$\begin{aligned} &\sum_{\tau=1}^{\Gamma-1} \tau \mathbb{E} [\|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\|] \\ &\leq (\Gamma-1) \sum_{\tau=1}^{\Gamma-1} \mathbb{E} [\|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\|] \end{aligned} \quad (94)$$

and

$$\begin{aligned} & \sum_{\tau=1}^{\Gamma-1} \tau \sum_{i=1}^N L_i \mathbb{E} [\|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\|] \\ & \leq (\Gamma-1) \sum_{i=1}^N L_i \sum_{\tau=1}^{\Gamma-1} \mathbb{E} [\|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\|] \end{aligned} \quad (95)$$

hold.

Hence, following from the conclusion of Lemma 18, we obtain

$$\begin{aligned} & \Gamma(\Gamma-1) \mathbb{E} [\|\bar{\mathbf{u}}(\Gamma) - \mathbf{u}^*\|^2] \leq -\frac{\mu}{N} \\ & \cdot \sum_{\tau=1}^{\Gamma-1} \tau \mathbb{E} [g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*)] - \frac{\mu}{N} \\ & \cdot \sum_{\tau=1}^{\Gamma-1} \tau \sum_{i=1}^N \sigma_i \mathbb{E} [\|\mathbf{u}_i(\tau) - \mathbf{u}^*\|^2] + \frac{\mu L(\Gamma-1)}{N} \\ & \cdot C \left( \frac{\lambda}{1-\lambda} \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 \right. \\ & \left. + \frac{\mu}{1-\lambda} \sum_{i=1}^N 2\sqrt{d}(L_i + \nu_i)(1 + \ln(\Gamma-1)) \right) \\ & + \frac{4\mu(\Gamma-1)}{N} \sum_{i=1}^N L_i C \left( \frac{\lambda}{1-\lambda} \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 \right. \\ & \left. + \frac{\mu}{1-\lambda} \sum_{i=1}^N \sqrt{d}(L_i + \nu_i)(1 + \ln(\Gamma-1)) \right) + \frac{4\mu^2}{N} \\ & \cdot \sum_{i=1}^N (L_i + \nu_i)^2 \sum_{\tau=1}^{\Gamma-1} \frac{\tau}{\tau+1}. \end{aligned} \quad (96)$$

Dividing both sides of (96) by  $\Gamma(\Gamma-1)$ , then we have that

$$\begin{aligned} & \frac{\mu}{N\Gamma(\Gamma-1)} \sum_{\tau=1}^{\Gamma-1} \tau \left[ \mathbb{E} [g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*)] \right. \\ & \left. + \sum_{i=1}^N \sigma_i \mathbb{E} [\|\mathbf{u}_i(\tau) - \mathbf{u}^*\|^2] \right] \leq \frac{\mu LC}{N\Gamma} \left( \frac{\lambda}{1-\lambda} \right. \\ & \left. \cdot \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 + \frac{\mu}{1-\lambda} \right. \\ & \left. \cdot \sum_{i=1}^N \sqrt{d}(L_i + \nu_i)(1 + \ln(\Gamma-1)) \right) + \frac{4\mu C}{N\Gamma} \sum_{i=1}^N L_i \end{aligned}$$

$$\begin{aligned} & \cdot \left( \frac{\lambda}{1-\lambda} \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 + \frac{\mu}{1-\lambda} \right. \\ & \left. \cdot \sum_{i=1}^N \sqrt{d}(L_i + \nu_i)(1 + \ln(\Gamma-1)) \right) + \frac{4\mu^2}{N\Gamma} \sum_{i=1}^N (L_i \\ & + \nu_i)^2, \end{aligned} \quad (97)$$

where we use the relation  $\sum_{\tau=1}^{\Gamma-1} (\tau/(\tau+1)) \leq \Gamma-1$  for  $\Gamma \geq 2$ . Hence, by a few algebraic operators, we obtain

$$\begin{aligned} & \frac{1}{\Gamma(\Gamma-1)} \sum_{\tau=1}^{\Gamma-1} \tau \left[ \mathbb{E} [g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*)] \right. \\ & \left. + \sum_{i=1}^N \sigma_i \mathbb{E} [\|\mathbf{u}_i(\tau) - \mathbf{u}^*\|^2] \right] \leq \frac{5LC}{\Gamma} \left( \frac{\lambda}{1-\lambda} \right. \\ & \left. \cdot \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 + \frac{\mu}{1-\lambda} \right. \\ & \left. \cdot \sum_{i=1}^N \sqrt{d}(L_i + \nu_i)(1 + \ln(\Gamma-1)) \right) + \frac{4\mu}{\Gamma} \sum_{i=1}^N (L_i \\ & + \nu_i)^2. \end{aligned} \quad (98)$$

Due to convexity of the function  $g$ , using the definition of  $\hat{\mathbf{u}}_i(\Gamma)$  and convexity of norm function, we obtain

$$\begin{aligned} & g(\hat{\mathbf{u}}_i(\Gamma)) - g(\mathbf{u}^*) + \sum_{j=1}^N \sigma_j \|\mathbf{u}_j(\Gamma) - \mathbf{u}^*\| \\ & \leq \frac{\sum_{\tau=1}^{\Gamma-1} \tau (g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*) + \sum_{j=1}^N \sigma_j \|\mathbf{u}_j(\tau) - \mathbf{u}^*\|)}{\Gamma(\Gamma-1)/2}. \end{aligned} \quad (99)$$

Taking expectation in both sides of (99), then

$$\begin{aligned} & \mathbb{E} [g(\hat{\mathbf{u}}_i(\Gamma)) - g(\mathbf{u}^*)] + \sum_{j=1}^N \sigma_j \mathbb{E} [\|\mathbf{u}_j(\Gamma) - \mathbf{u}^*\|] \\ & \leq \frac{2 \sum_{\tau=1}^{\Gamma-1} \tau (\mathbb{E} [g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*)] + \sum_{j=1}^N \sigma_j \mathbb{E} [\|\mathbf{u}_j(\tau) - \mathbf{u}^*\|])}{\Gamma(\Gamma-1)}. \end{aligned} \quad (100)$$

Therefore, substituting (98) into (100) and by a few algebraic operators, we derive the desired result.  $\square$

Next, we give the proof of Theorem 8, where we assume that the functions  $g_i$  are generally convex for all  $i \in \mathcal{V}$ .

*Proof of Theorem 8.* Let  $\sigma_i = 0$  for all  $i = 1, \dots, N$  in Assumption 3. Moreover, we introduce the following variable, which is the ergodic average of the estimates  $\mathbf{u}_i(\tau)$ :

$$\bar{\mathbf{u}}_i(\Gamma) = \frac{1}{\sum_{\tau=0}^{\Gamma} \alpha(\tau)} \sum_{\tau=0}^{\Gamma} \alpha(\tau) \mathbf{u}_i(\tau). \quad (101)$$

According to (86), we have

$$\begin{aligned} & \mathbb{E} \left[ \|\bar{\mathbf{u}}(\tau+1) - \mathbf{u}^*\|^2 \mid \mathcal{F}_\tau \right] \\ & \leq \|\bar{\mathbf{u}}(\tau) - \mathbf{u}^*\|^2 + \frac{4\alpha(\tau)}{N} \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ & \quad + \frac{4\alpha^2(\tau)}{N} \sum_{i=1}^N (L_i + \nu_i)^2 \\ & \quad - \frac{2\alpha(\tau)}{N} \mathbb{E} [g(\bar{\mathbf{u}}(\tau)) - g(\mathbf{u}^*)]. \end{aligned} \quad (102)$$

Combining (88) with (102), and after some elementary algebra operations, we obtain

$$\begin{aligned} & \frac{2\alpha(\tau)}{N} \mathbb{E} [g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*)] \\ & \leq \|\bar{\mathbf{u}}(\tau) - \mathbf{u}^*\|^2 + \frac{4\alpha^2(\tau)}{N} \sum_{i=1}^N (L_i + \nu_i)^2 \\ & \quad + \frac{2\alpha(\tau)}{N} L \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ & \quad + \frac{4\alpha(\tau)}{N} \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ & \quad - \mathbb{E} \left[ \|\bar{\mathbf{u}}(\tau+1) - \mathbf{u}^*\|^2 \mid \mathcal{F}_\tau \right]. \end{aligned} \quad (103)$$

By summing the preceding inequality and dividing by  $(2/N)\Phi(\Gamma)$ , where  $\Phi(\Gamma) = \sum_{\tau=0}^{\Gamma} \alpha(\tau)$ , we obtain

$$\begin{aligned} & \frac{\sum_{\tau=0}^{\Gamma} \alpha(\tau) \mathbb{E} [g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*)]}{\Phi(\Gamma)} \\ & \leq \frac{N}{2\Phi(\Gamma)} \|\bar{\mathbf{u}}(0) - \mathbf{u}^*\|^2 \\ & \quad + \frac{L}{\Phi(\Gamma)} \sum_{\tau=0}^{\Gamma} \alpha(\tau) \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ & \quad + \frac{2}{\Phi(\Gamma)} \sum_{\tau=0}^{\Gamma} \alpha(\tau) \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ & \quad + \frac{2}{\Phi(\Gamma)} \sum_{\tau=0}^{\Gamma} \alpha^2(\tau) \sum_{i=1}^N (L_i + \nu_i)^2. \end{aligned} \quad (104)$$

Since  $\alpha(\tau) = 1/\sqrt{\tau+1}$ , we have

$$\begin{aligned} \Phi(\Gamma) & = \sum_{\tau=0}^{\Gamma} \frac{1}{\sqrt{\tau+1}} \geq \int_0^{\Gamma+1} \frac{d\gamma}{\sqrt{\gamma+1}} = 2(\sqrt{\Gamma+2} - 1) \\ & \geq \sqrt{\Gamma+1} \end{aligned} \quad (105)$$

for all  $\Gamma \geq 1$ . Moreover, we also have

$$\begin{aligned} \sum_{\tau=0}^{\Gamma} \alpha^2(\tau) & = \sum_{\tau=0}^{\Gamma} \frac{1}{\tau+1} = \sum_{\gamma=1}^{\Gamma+1} \frac{1}{\gamma} \leq 1 + \int_1^{\Gamma+1} \frac{d\gamma}{\gamma} \\ & = 1 + \ln(\Gamma+1). \end{aligned} \quad (106)$$

In addition, using the result of Corollary 13, we have

$$\begin{aligned} \sum_{\tau=0}^{\Gamma} \alpha(\tau) \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| & \leq \sum_{\tau=0}^{\Gamma} \alpha(\tau) \left( C\lambda^\tau \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 \right. \\ & \quad \left. + \sum_{j=1}^N 2\sqrt{d}(L_j + \nu_j) \sum_{t=0}^{\tau-1} C\lambda^{\tau-t-1} \alpha(t) \right), \end{aligned} \quad (107)$$

where the inequality follows from Lemma 14. Since  $\alpha(\tau) = 1/\sqrt{\tau+1} \leq 1$ , so we obtain

$$\sum_{\tau=0}^{\Gamma} \alpha(\tau) \lambda^\tau \leq \frac{1}{1-\lambda}. \quad (108)$$

Further, we also have

$$\begin{aligned} \sum_{\tau=0}^{\Gamma} \sum_{t=0}^{\tau-1} \lambda^{\tau-t-1} \alpha(t) \alpha(\tau) & = 1 + \sum_{\vartheta=1}^{\Gamma} \lambda^{\vartheta-1} \sum_{t=1}^{\Gamma-\vartheta+1} \frac{1}{\sqrt{t}\sqrt{\vartheta+t}} \\ & \leq 1 + \sum_{\vartheta=1}^{\Gamma} \lambda^{\vartheta-1} \int_0^{\Gamma-\vartheta+1} \frac{d\gamma}{\sqrt{\gamma}\sqrt{\gamma+\vartheta}} \\ & \leq 1 + 2 \sum_{\vartheta=1}^{\Gamma} \lambda^{\vartheta-1} \ln(\sqrt{\Gamma-\vartheta+1} + \sqrt{\Gamma+1}) \\ & \leq 1 + \frac{2}{1-\lambda} \left( \ln 2 + \frac{1}{2} \ln(\Gamma+1) \right) \\ & \leq \frac{4}{1-\lambda} \ln(\Gamma+1). \end{aligned} \quad (109)$$

Combining (107), (108), and (109), we obtain

$$\begin{aligned} \sum_{\tau=0}^{\Gamma} \alpha(\tau) \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| & \leq \frac{C}{1-\lambda} \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 \\ & \quad + \frac{4C \ln(\Gamma+1)}{1-\lambda} \sum_{j=1}^N 2\sqrt{d}(L_j + \nu_j). \end{aligned} \quad (110)$$

Thus, following from the preceding relation, we obtain

$$\begin{aligned} \sum_{\tau=0}^{\Gamma} \alpha(\tau) \sum_{i=1}^N L_i \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| & = \sum_{i=1}^N L_i \sum_{\tau=0}^{\Gamma} \alpha(\tau) \|\mathbf{u}_i(\tau) - \bar{\mathbf{u}}(\tau)\| \\ & \leq \frac{LC}{1-\lambda} \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1 \\ & \quad + \frac{4LC \ln(\Gamma+1)}{1-\lambda} \sum_{j=1}^N 2\sqrt{d}(L_j + \nu_j). \end{aligned} \quad (111)$$

Substituting (105), (106), (110), and (111) into (104), we have

$$\begin{aligned}
& \frac{\sum_{\tau=0}^{\Gamma} \alpha(\tau) \mathbb{E} [g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*)]}{\Phi(\Gamma)} \\
& \leq \frac{3LC \sum_{j=1}^N \|\mathbf{u}_j(0)\|_1}{(1-\lambda)\sqrt{\Gamma+1}} \\
& \quad + \frac{9LC \ln(\Gamma+1)}{(1-\lambda)\sqrt{\Gamma+1}} \sum_{j=1}^N 2\sqrt{d}(L_j + \nu_j) \quad (112) \\
& \quad + \frac{2(1 + \ln(\Gamma+1))}{\sqrt{\Gamma+1}} \sum_{i=1}^N (L_i + \nu_i)^2 \\
& \quad + \frac{N}{2\sqrt{\Gamma+1}} \|\bar{\mathbf{u}}(0) - \mathbf{u}^*\|^2.
\end{aligned}$$

Besides, by convexity of function  $g$ , and then following from the definition of  $\tilde{\mathbf{u}}_i(\Gamma)$  in (101), then we conclude that

$$\begin{aligned}
& \mathbb{E} [g(\tilde{\mathbf{u}}_i(\Gamma)) - g(\mathbf{u}^*)] \\
& \leq \frac{\sum_{\tau=0}^{\Gamma} \alpha(\tau) \mathbb{E} [g(\mathbf{u}_i(\tau)) - g(\mathbf{u}^*)]}{\Phi(\Gamma)} \quad (113)
\end{aligned}$$

for all  $i \in \mathcal{V}$  and  $\Gamma \geq 1$ . Therefore, combining (112) with (113), the conclusion of this theorem is obtained completely.  $\square$

In this section, we prove our main results, which are presented in Section 3. If step-sizes  $\alpha(\tau)$  are positive and satisfy decay conditions (26), our proposed algorithm is asymptotically convergent by appropriately choosing step-sizes  $\alpha(\tau)$ . Further, our algorithm converges with rate  $O(\ln \Gamma/\Gamma)$  under strong convexity and with rate  $O(\ln \Gamma/\sqrt{\Gamma})$  under general convexity.

## 5. Simulations

In this section, we consider the multiclass classification problem in machine learning. In the problem, each agent  $i \in \mathcal{V}$  generates a decision matrix  $U_i(t) = [\mathbf{u}_1^T; \dots; \mathbf{u}_c^T]^T \in \mathbb{R}^{c \times d}$  through a data example  $\mathbf{e}_i(t) \in \mathbb{R}^d$  that is one of the classes  $\mathcal{C} = \{1, \dots, c\}$ . The decision is used to predict the class label via  $\arg \max_{\ell \in \mathcal{C}} \mathbf{u}_\ell^T \mathbf{e}_i(t)$ . The loss function of each agent  $i \in \mathcal{V}$  is defined as

$$\begin{aligned}
& g_i(U_i(t)) \\
& = \ln \left( 1 + \sum_{\ell \neq y_i(t)} \exp(\mathbf{u}_\ell^T \mathbf{e}_i(t) - \mathbf{u}_{y_i(t)}^T \mathbf{e}_i(t)) \right), \quad (114)
\end{aligned}$$

where  $y_i(t)$  denotes the true class label. In this problem, the domain  $\mathcal{U}$  is given by  $\mathcal{U} = \{U \in \mathbb{R}^{c \times d} \mid \|U\|_{\text{tr}} \leq \zeta\}$ , in which  $\|U\|_{\text{tr}}$  represents the nuclear norm of  $U$ .

In our simulations, we set  $N = 10$ , the step-size is set as  $\alpha(\tau) = 1/\sqrt{\tau}$ . Moreover, we use a multiclass dataset, which is called news20 (<http://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/>). This dataset has 62,061 features, 20

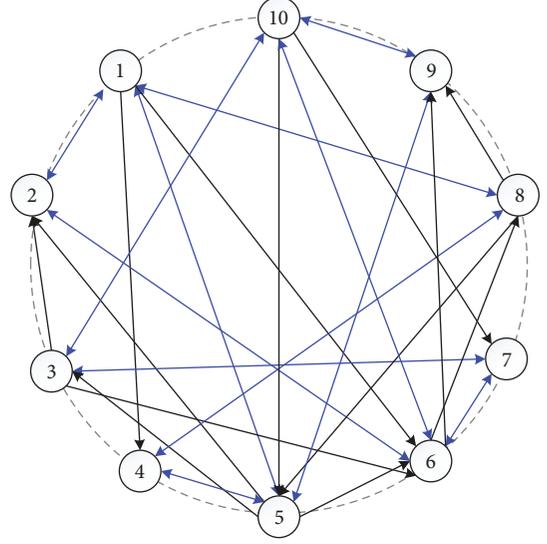


FIGURE 1: The strongly connected time-varying graph.

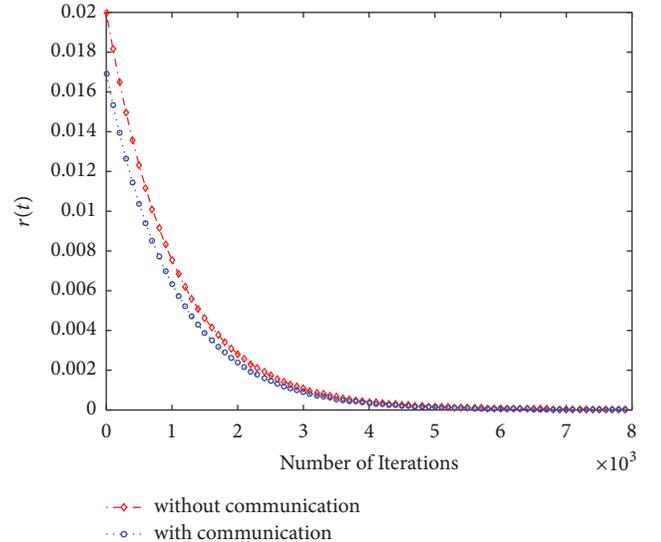


FIGURE 2: The performance of our algorithm with communication and without communication.

classes, and 15,935 instances. The testing size of news20 is 3,993. In addition, the time-varying graph at round  $\tau$  is shown in Figure 1. We use the following quantity to measure the performance of the algorithms:

$$r(t) = \frac{1}{N} \sum_{i=1}^N (g_i(\mathbf{u}_i(t)) - g_i(\mathbf{u}^*)). \quad (115)$$

In the first simulation experiment, we compare the performances of the algorithm with communication and without communication. As shown in Figure 2, we can see that the performance with communication is better than that without communication.

In the second simulation experiment, we compare the convergence performances with D-DPS [27]. As shown in

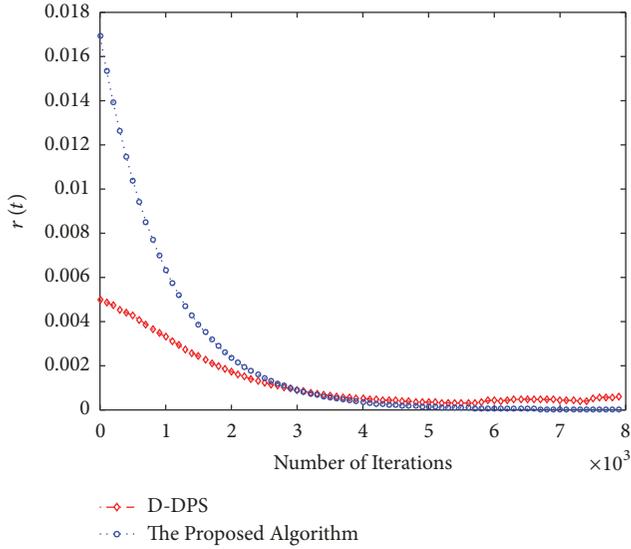


FIGURE 3: Comparison of D-DPS and the proposed algorithm.

Figure 3, for the similar setting, these algorithms have similar convergence rates. However, our algorithm considers the noisy subgradient. Moreover, the cost functions may be non-differentiable and we consider the case that the graph is time-varying. Therefore, our algorithm has better applicability to real environments.

## 6. Conclusion

We have proposed a fully decentralized stochastic subgradient projection algorithm to solve distributed constrained optimization problem over time-varying directed networks. We employ weight-balancing technique to overcome the influence of directed graphs and assume that every agent knows its own out-degree at each time epoch in our algorithm. Moreover, all agents have access to their own noisy subgradient. We have proved that our proposed algorithm is asymptotically convergent with suitable chosen step-sizes  $\alpha(\tau)$  over time-varying directed networks. Furthermore, for local strongly convex functions, we have proved that our algorithm converges with the rate  $O(\ln \Gamma/\Gamma)$ . Meanwhile, for generally convex functions, we have also showed that the convergence rate  $O(\ln \Gamma/\sqrt{\Gamma})$  is achieved. Additionally, the performance of the proposed algorithm has also been evaluated by simulations.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

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

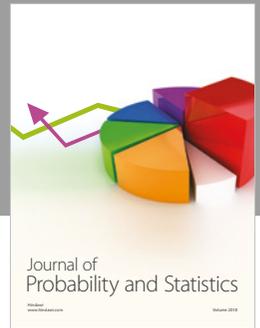
## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (NSFC), under Grants no. 61801171, no. 61871430, no. U1604155, and no. 61602155, in part by the basic research projects in the University of Henan Province, under Grant no. 19zx010, and in part by the Science and Technology Development Programs of Henan Province, under Grant no. 192102210284.

## References

- [1] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, Prediction*, Springer-Verlag, New York, NY, USA, 2001.
- [2] J. L. R. Bekkerman and M. Bilenko, *Scaling up Machine Learning: Parallel and Distributed Approaches*, Cambridge University Press, Cambridge, UK, 2011.
- [3] 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.
- [4] D. Belomestny, A. Kolodko, and J. Schoenmakers, "Regression methods for stochastic control problems and their convergence analysis," *SIAM Journal on Control and Optimization*, vol. 48, no. 5, pp. 3562–3588, 2010.
- [5] B. Johansson, *On Distributed Optimization in Networked Systems [Ph.D. dissertation]*, Royal Institute of Technology, Stockholm, Sweden, 2008.
- [6] 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.
- [7] V. Lesser, C. Ortiz, and M. Tambe, *Distributed Sensor Networks: A Multiagent Perspective*, Kluwer Academic Publishers, Norwell, MA, USA, 2003.
- [8] M. Rabbat and R. Nowak, "Distributed optimization in sensor networks," in *Proceedings of ACM IPSN*, pp. 20–27, Berkeley, CA, USA, 2004.
- [9] 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.
- [10] S. Kar, J. Moura, and K. Ramanan, "Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.
- [11] R. Olfati-Saber, J. A. Fax, and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.
- [12] J. N. Tsitsiklis, *Problems in Decentralized Decision Making and Computation [Ph. D. Dissertation]*, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, UK, 1984.
- [13] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.
- [14] Y. A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

- [15] I. Lobel, A. Ozdaglar, and D. Feijer, "Distributed multi-agent optimization with state-dependent communication," *Mathematical Programming*, vol. 129, no. 2, pp. 255–284, 2011.
- [16] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4289–4305, 2012.
- [17] H. Wang, X. Liao, T. Huang, and C. Li, "Cooperative distributed optimization in multi-agent networks with delays," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 2, pp. 363–369, 2014.
- [18] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: an exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.
- [19] G. Qu and N. Li, "Harnessing smoothness to accelerate distributed optimization," *IEEE Transactions on Control of Network Systems*, vol. 5, no. 3, pp. 1245–1260, 2018.
- [20] A. Mokhtari, Q. Ling, and A. Ribeiro, "Network newton distributed optimization methods," *IEEE Transactions on Signal Processing*, vol. 65, no. 1, pp. 146–161, 2017.
- [21] I. Necoara, "Distributed and parallel random coordinate descent methods for huge convex programming over networks," in *Proceedings of the IEEE Conference on Decision and Control*, pp. 425–430, 2015.
- [22] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.
- [23] 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.
- [24] A. Makhdoumi 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.
- [25] J. M. Hendrickx and J. N. Tsitsiklis, "Fundamental limitations for anonymous distributed systems with broadcast communications," in *Proceedings of the 53rd Annual Allerton Conference on Communication, Control, and Computing, Allerton 2015*, pp. 9–16, USA, October 2015.
- [26] B. Touri and B. Ghahsifard, "Continuous-time distributed convex optimization on time-varying directed networks," in *Proceedings of the 54th IEEE Conference on Decision and Control, CDC 2015*, pp. 724–729, Japan, December 2015.
- [27] C. Xi and U. A. Khan, "Distributed subgradient projection algorithm over directed graphs," *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 62, no. 8, pp. 3986–3992, 2017.
- [28] A. Agarwal, P. L. Bartlett, P. Ravikumar, and M. J. Wainwright, "Information-theoretic lower bounds on the oracle complexity of stochastic convex optimization," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 58, no. 5, pp. 3235–3249, 2012.
- [29] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*, John Wiley, New York, NY, USA, 1983.
- [30] S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, 2010.
- [31] B. T. Polyak, *Introduction to Optimization*, Optimization Software, New York, NY, USA, 1987.



**Hindawi**

Submit your manuscripts at  
[www.hindawi.com](http://www.hindawi.com)

