Distributed Unbalanced Optimization Design Over Nonidentical Constraints

Qing Huang, Yuan Fan, *Senior Member, IEEE*, and Songsong Cheng, *Member, IEEE*

***Abstract*—This paper addresses distributed constrained optimization problems involving strongly convex global objective functions represented as the sum of individual convex objective functions, and the corresponding constrained set is the intersection of N nonidentical closed convex sets. To solve the problem, we introduce the distributed projected sub-gradient algorithm with a row-stochastic weight matrix over unbalanced digraphs. Moreover, based on the condition that the strong convexity of the global objective function and using a non-increasing step size, we analyze that this algorithm converges to the optimal solution with an $O\left(\frac{1}{T}\right)$ convergence rate, like the centralized counterpart. Finally, we verify the accuracy of the theoretical analysis by examining simulation results.**

***Index Terms*—Distributed optimization, nonidentical constraints, strongly convex, row stochastic, convergence.**

I. Introduction

DISTRIBUTED optimization is a type of optimization problem addressed by collaborating networked multi-agents with the advantages of decentralized communication, network robustness, and privacy protection et al. It possesses crucial research significance and application in power distribution control within communication networks [1], machine learning [2], [3], resource allocation in smart grid [4], [5], and cooperative source location in sensor networks [6]. This field has gained increasing attention in recent decades and yielded many promising results, such as subgradient method [7], [8], EXTRA [9], [10], and gradient tracking [11], [12], [13], [14] for unconstrained optimization and primal-dual [15], [16], subgradient projection [17], [18], dual-averaging [19], [20], [21], and Frank-Wolfe [22], [23] for constrained counterparts.

In some circumstances, the decision variables of the optimization problem are constrained by limited resources. For the contained optimization problem, [24] designed a distributed mirror descent algorithm to obtain the optimal solution with an $O(1/\sqrt{T})$ convergence rate, [25] proposed a randomized gradient-free distributed optimization method that achieved an $O(\ln T/\sqrt{T})$ convergence rate. However, these methods are designed based on the assumption that all local decision variables are constrained by an identical set. In many practices, each agent may constrained by nonidentical sets, which raises more challenges than that of identical counterparts, such as slow convergence rate [26], inexact convergence accuracy [27]. For nonidentical constrained optimization problem, [16] designed a primal-dual method with an $O(1/T)$ convergence rate, [28] developed a zeroth-order algorithm for nonsmooth optimization with an $O(1/\sqrt{T})$ convergence rate. [29] proposed a Fenchel dual gradient for distributed optimization problems over nonidentical constraints with an $O(1/\sqrt{T})$ convergence rate. In [27] modified the subgradient method with an averaging strategy and improved the convergence accuracy. However, all of these methods [16], [24], [27], [28], [29] are proposed based on undirected graphs with doubly stochastic weighted matrix.

In some cases, due to environmental influences or individual energy differences, the network topology is unbalanced, which makes it difficult or impossible to construct a doubly stochastic matrix. To deal with this problem, there are some common strategies in the preceding literature including Surplus [30], [31], Push-Sum [32], [33], and Push-Pull [34], [35], just to name a few. Nevertheless, implementing these methods mentioned above involves the construction of a column stochastic weight matrix as a prerequisite, which requires all agents to possess precise knowledge of their out-degree, denoting the quantity of information transmitted to other agents. However, meeting such requirements is challenging in certain cases, where information dispatched to other agents may encounter partial reception, resulting in data loss. In an alternative scheme, agents have the capacity to autonomously decide the weighting of information received from neighboring agents. This scheme can be guaranteed as long as the summation of per row of the weighted matrix equals 1. Consequently, the feasibility of achieving a row stochastic matrix appears more tractable than the column counterpart [32], [33]. When using the row stochastic matrix to solve problems with nonidentical constraints, a distributed projected subgradient algorithm and its variation are proposed in [36] with an $O(\ln T/\sqrt{T})$ convergence rate, whereas the objective functions are convex. A summary of the main proceeding results for solving distributed optimization is presented in Table I.

Most existing publications mainly concentrate on satisfying the single or several conditions of strongly convex objective functions, nonidentical set constraints, and row stochastic weight matrices. Building upon the insights from the preceding discussion, this paper investigates distributed strongly convex

Manuscript received 25 September 2023; revised 5 March 2024; accepted 5 March 2024. Date of publication 12 March 2024; date of current version 12 June 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62103003 and Grant 61973002, and in part by the Anhui Provincial Natural Science Foundation under Grant 2008085J32. Recommended for acceptance by Dr. Y. Sun. *(Corresponding author: Songsong Cheng.)*

The authors are with the School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China (e-mail: yuanf@ahu.edu.cn; sscheng@ahu.edu.cn).

Digital Object Identifier 10.1109/TNSE.2024.3374765

TABLE I
RELATED WORKS ON SOLVING DISTRIBUTED OPTIMIZATION PROBLEMS

Related works	Objective functions	Constraints	Graphs	Weight matrix	Convergence
[17]	strongly convex	identical	directed	doubly stochastic	$O(1/\sqrt{T})$
[19]	convex	identical	undirected	doubly stochastic	$O(1/\sqrt{T})$
[20], [21]	convex	identical	directed	column stochastic	$O(1/\sqrt{T})$
[23]	convex	identical	balanced	doubly stochastic	$O(1/T)$
[24]	strongly convex	identical	directed	doubly stochastic	$O(1/T)$
[25], [30]	convex	identical	directed	row & column stochastic	$O(\ln T/\sqrt{T})$
[31]	convex	identical	directed	row & column stochastic	$O(1/\sqrt{T})$
[33]	convex	identical	unbalanced	column stochastic	$O(1/\sqrt{T})$
[27]	convex	nonidentical	undirected	doubly stochastic	$O(\ln T/\sqrt{T})$
[16]	convex	nonidentical	undirected	doubly stochastic	Asymptotic
[36]	convex	nonidentical	unbalanced	row stochastic	$O(\ln T/\sqrt{T})$
[29]	strongly convex	nonidentical	undirected	doubly stochastic	$O(1/T)$
[26]	strongly convex	nonidentical	unbalanced	row & column stochastic	$O(1/T)$
[32]	strongly convex	nonidentical	unbalanced	column stochastic	$O(1/T)$
This work	strongly convex	nonidentical	unbalanced	row stochastic	$O(1/T)$

optimization problems by a distributed projected subgradient algorithm in conjunction with the row stochastic matrix, over the framework of nonidentical constraints and unbalanced graphs, and attains the expected convergence rate. The significance of our work manifests in three principal facets, constituting its core contributions:

1) Compared to the more stringent requirements imposed in prior works, which necessitated the use of doubly stochastic matrices in [16], [17], [19], [23], [29] and column stochastic matrices in [10], [21], [25], [26], [32], [33], this paper employs a row stochastic matrix, offering enhanced flexibility and robustness in terms of communication topology. Moreover, although the communication topology equipping doubly or column stochastic matrix achieves an $O(1/T)$ convergence rate for distributed strongly convex optimization in works such as [16], [26], [29], [32], the convergence performance under a row stochastic matrix remains an open problem.
2) In contrast to the considered optimization problem with identical constraints in [17], [19], [20], [21], [23], [24], [25], [30], [31], [33], this paper takes into account the nonidentical constraints and overcomes the challenges of slow or inaccurate convergence, as pointed out in [26], [27].
3) This paper addresses distributed strongly convex optimization problems over unbalanced graphs. Compared to the Lipschitz continuity of each gradient in [13], [15], [16], [26], [33] and strong convexity of each objective function in [13], [15], [26], this work only needs that each objective function itself is Lipschitz continuous and at least one with strong convexity, thereby relaxing the condition of the objective function.

The remainder of this paper is constructed as follows. Section II gives a few relevant preliminaries, such as notations and graph theory. Section III provides the problem formulation, algorithm design, necessary assumptions, and technical lemmas and Section IV analyses the convergence performance of the algorithm. Lastly, Section V presents the simulation results and Section VI concludes this paper.

II. PRELIMINARIES

A. Notations

In this paper, the m-dimensional vectors collection and the $m \times m$ matrices collection of real-valued space are denoted as $\mathbb{R}^m$ and $\mathbb{R}^{m\times m}$, respectively. For a given vector $\zeta \in \mathbb{R}^m$ and matrix $A \in \mathbb{R}^{m\times m}$, $\|\zeta\|$ and $\|A\|$ represent the 2-norm of ζ and the induced 2-norm of A, respectively. For matrix $A := [a_{ij}]$, a_{ij} denotes its element lie in the i-th row and j-th column. Especially, if $A\mathbf{1} = \mathbf{1}$ with $\mathbf{1} = [1, 1, \ldots, 1]^{\mathrm{T}}$ in terms of a non-negative square matrix A, then it is a row stochastic. Moreover, A is column stochastic if $A^{\mathrm{T}}\mathbf{1} = \mathbf{1}$. Additionally, $\partial f(\zeta)$ is the subgradient of function f at ζ. Moreover, regarding arbitrarily given nonempty closed convex set $\Omega \subseteq \mathbb{R}^m$, the projection of vector ζ on Ω is denoted by $P_\Omega(\zeta)$. Let $\mathrm{dist}(\zeta, \Omega) = \|\zeta - P_\Omega(\zeta)\|$ be the distance from ζ to Ω.

B. Graph Theory

Considering an enumerable agents set $\mathcal{V} = \{1, 2, \ldots, N\}$ together with an edges set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, the information exchange among agents is governed by a digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. For any two agents that are not necessarily completely different, we denote $(i, j) \in \mathcal{E}$ if agent i can accept information from j. An edges sequence $(i_1, i_2), (i_2, i_3), \ldots, (i_{m-1}, i_m)$ is used to define a directed route from node i_m to i_1 in the graph $\mathcal{G}$. A network topology $\mathcal{G}$ is strongly connected if a directed route exists between any two nodes. Define $\mathcal{N}_i^{\mathrm{in}} := \{j \in \mathcal{V} : (i, j) \in \mathcal{E}\}$ and $\mathcal{N}_i^{\mathrm{out}} := \{j \in \mathcal{V} : (j, i) \in \mathcal{E}\}$ as the sets of agent i's in-neighbors and out-neighbors, respectively. The cardinality of $\mathcal{N}_i^{\mathrm{in}}$ is called the in-degree of agent i and denoted by $|\mathcal{N}_i^{\mathrm{in}}|$. Besides, an unbalanced graph means there exists at least one $i \in \mathcal{V}$, $|\mathcal{N}_i^{\mathrm{in}}| \neq |\mathcal{N}_i^{\mathrm{out}}|$.

For the communication topology of this paper, a mild assumption is taken as follows.

Assumption 1: The digraph $\mathcal{G}$ is strongly connected and unbalanced.

For an unbalanced graph, it is not easy to design a doubly stochastic matrix. Alternatively, we formulate the following row stochastic matrix

$$w_{ij} = \frac{a_{ij}}{d_i^{\text{in}}},$$

where $d_i^{\text{in}} = \sum_{j=1}^{n} a_{ij}$ is the in-degree of node i in graph $\mathcal{G}$. Then, $W := [w_{ij}]$ is a row stochastic matrix.

III. Problem Formulation and Algorithm

A. Problem Formulation

We consider a multi-agent system cooperating to minimize the following constrained optimization problem

$$\min_{\zeta\in\Omega} \psi(\zeta) := \sum_{i\in\mathcal{V}} \psi_i(\zeta), \quad \text{s.t.}\, \Omega = \bigcap_{i\in\mathcal{V}} \Omega_i, \tag{1}$$

where $\zeta \in \mathbb{R}^m$ and $\psi_i : \mathbb{R}^m \to \mathbb{R}$ are local decision variable and cost function, respectively. $\Omega_i \subseteq \mathbb{R}^m$ is closed and convex constraint set. The optimal value and the optimal solution set of the problem (1) are denoted as ψ^* and Ω^*, respectively.

Before presenting the algorithm, we take some other helpful assumptions.

Assumption 2: For any $i \in \mathcal{V}$, the feasible sets Ω_i are bounded and the local cost functions $\psi_i(\zeta)$ are convex. Moreover, there exists at least one $j \in \mathcal{V}$ such that ψ_j is μ-strongly convex, which means that the global objective function $\psi(\zeta)$ is μ-strongly convex.

Remark 1: Many existing results, such as [13], [15], [26], assume that all local objective functions are strongly convex to achieve faster convergence. As a substitute, we make a relaxed assumption that at least one local objective function is strongly convex. This relaxation significantly reduces the restrictive assumption on the cost functions to obtain the desired convergence speed.

Assumption 3: For any $i \in \mathcal{V}$, $\psi_i(\zeta)$ is L-Lipschitz continuous, i.e.,

$$|\psi_i(\zeta_1) - \psi_i(\zeta_2)| \le L\|\zeta_1 - \zeta_2\|.$$

Remark 2: Assumption 3 is important for subsequent convergence analysis, indicating that the sub-gradients of all local objective functions are bounded, i.e., $\|\partial\psi_i\| \le L$.

B. Algorithm Design

To optimize the aforementioned problem and inspired from [36], we have the following distributed projected subgradient (DPSG) algorithm:

$$\zeta_i(t+1) = P_{\Omega_i}\left\{\sum_{j\in\mathcal{V}} w_{ij}\zeta_j(t) - \alpha(t)\tfrac{g_i(t)}{y_{ii}(t)}\right\}, \tag{2}$$

$$y_i(t+1) = \sum_{j\in\mathcal{V}} w_{ij} y_j(t), \tag{3}$$

where $g_i(t) \in \partial\psi_i(\sum_{j\in\mathcal{V}} w_{ij}\zeta_j(t))$, $y_{ii}(t)$ is the i-th element of vector $y_i(t)$, $y_i(t) = [y_{i1}, y_{i2}, \ldots, y_{ii}, \cdots, y_{iN}]^{\mathrm{T}} \in \mathbb{R}^{\mathrm{N}}$, $y_i(0) = [0, \ldots, 0, 1, 0, \ldots, 0]^{\mathrm{T}}$ with the i-th entry being 1 and others being 0, and $\alpha(t)$ is the step size with $\sum_{t=0}^{\infty}\alpha(t) = \infty$ and $\sum_{t=0}^{\infty}\alpha^2(t) < \infty$.

Remark 3: This algorithm only requires the weight matrix to be a row stochastic matrix. Compared with column stochastic weight matrices [20], [21], [25], [26], [30], [31], [32], [33], each node does not need to know its out-degree and can independently assign weights to the received neighbors information. Specifically, when communication links in the network are broken, the column stochastic matrix fails due to nodes being unable to adjust their outgoing weights. Instead, the agent only needs to assign weights to the received neighbors' information to construct a row stochastic matrix, which enhances the robustness of the communication topology, making the row stochastic matrix more practical.

Lemma 1: [37] Let Assumption 1 hold. Then for any $i \in \mathcal{V}$ and $\lambda \in (|\lambda_2(W)|, 1)$ with $\lambda_2(W)$ is the second largest eigenvalue of weight matrix W, there exist positive constants C and η satisfying

$$|y_{ii}(t) - \sigma_i| \le C\lambda^t, \quad \tfrac{1}{\eta} \le y_{ii}(t) \le 1,$$

where σ is the left normalized Perron eigenvector of W.

Remark 4: The iteration of y_i actually provides each agent with an estimate the Perron eigenvector $\sigma = [\sigma_1, \ldots, \sigma_N]^{\mathrm{T}}$ with $\mathbf{1}^{\mathrm{T}}\sigma = 1$.

Lemma 2: [32] The closed convex constraint sets Ω_i are regular if there exists a constant $R > 0$ such that

$$\text{dist}(\zeta, \Omega) \le R \max_{i\in\mathcal{V}} \text{dist}(\zeta, \Omega_i), \quad \forall \zeta \in \mathbb{R}^m.$$

Remark 5: Lemma 2 satisfies naturally as long as the intersection set Ω has a nonempty interior. It can be seen that it holds for any $R \ge 1$ when Ω_i are identical. Actually, $R = 1$ is an effective regularity constant. Some special cases include the real space ($\Omega_i = \mathbb{R}^m$), the halfspace ($\Omega_i = \{\zeta | u^\top\zeta \le v\}$), and the hyperplane ($\Omega_i = \{\zeta | u^\top\zeta = v\}$).

For the projection operator, we have the following technical lemma.

Lemma 3 ([38]): If $\Omega_i \subseteq \mathbb{R}^m$ is a closed and convex set, then $\|P_{\Omega_i}(\zeta) - \varpi^*\|^2 \le \|\zeta - \varpi^*\|^2 - \|P_{\Omega_i}(\zeta) - \zeta\|^2, \forall\zeta \in \mathbb{R}^m, \varpi^* \in \Omega_i$.

Moreover, for the convenience of the discussion hereinafter, we define

$$\phi_i(t) = \zeta_i(t+1) - \left[\sum_{j=1}^{n} w_{ij}\zeta_j(t) - \alpha(t)\tfrac{g_i(t)}{y_{ii}(t)}\right],$$

$$\sigma_{\mathrm{m}} = \min_{i\in\mathcal{V}} \sigma_i, \quad a = \tfrac{2CL(\eta+RN^2)}{(1-\lambda)b}, \quad b = \sqrt{\tfrac{\sigma_{\mathrm{m}}}{N}},$$

$$A_1 = C\sum_{j\in\mathcal{V}} \|\zeta_j(0)\|, \qquad \varphi(t) = \sum_{i\in\mathcal{V}} \|\phi_i(t)\|^2,$$

$$A_2 = (2L\eta + 2\,N^2LR)NCL\eta, \quad \eta(t) = \sum_{s=0}^{t-1} \lambda^{t-s-1}\varphi(s),$$

$$A_3 = (2L\eta + 2\,N^2LR)A_1, \qquad \bar{\zeta}(t) = \sum_{j\in\mathcal{V}} \sigma_j\zeta_j(t),$$

$$A_4 = L^2\eta^2 + NCL\eta + \tfrac{a^2}{2}, \qquad \beta(t) = \sum_{i\in\mathcal{V}} \|\phi_i(t)\|,$$

$$\delta(t) = \alpha(t) \sum_{0\le s\le t-1} \lambda^{t-1-s}\beta(s), \;\; \delta(0) = 0.$$

Based on the above definitions, we have the following technical lemmas.

Lemma 4 ([36]): Supposing Assumptions 1–3 hold, we have

a) For any $i \in \mathcal{V}$,

$$\|\zeta_i(t) - \bar{\zeta}(t)\| \le A_1\lambda^t + \sum_{s=0}^{t-1}\lambda^{t-1-s}\left(NC \times L\eta\alpha(s) + C\beta(s)\right). \quad (4)$$

b) If $\alpha(t)$ is nonincreasing then

$$\delta(t+1) \le \lambda\delta(t) + \alpha(t)\beta(t). \quad (5)$$

Lemma 5 ([32]): If sequence $\{\theta(t)\}$ is strictly positive, then

$$\left(\sum_{s=0}^{t-1}\lambda^{t-1-s}\theta(s)\right)^2 \le \tfrac{1}{1-\lambda}\sum_{s=0}^{t-1}\lambda^{t-1-s}\theta^2(s), \quad \forall t\ge 1. \quad (6)$$

IV. Convergence Results

In this section, we provide three primary theorems to show the convergence performance of the DPSG. Firstly, we give a pivotal theorem that serves as a basis for the convergence analysis hereinafter.

Theorem 1: Posit Assumptions 1–3 hold. If $\{\alpha(t)\}$ is nonincreasing, then for any $\varpi^* \in \Omega^*$

$$\begin{aligned}
&\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1)-\varpi^*\|^2 + ab\delta(t+1)\\
&\le (1-\mu\alpha(t)+NCL\eta\lambda^{2t})\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\varpi^*\|^2 + ab\delta(t)\\
&\quad - 2\alpha(t)\left(\psi(\varpi(t))-\psi(\varpi^*)\right) + A_3\alpha(t)\lambda^t + A_4\alpha^2(t)\\
&\quad + A_2\alpha(t)\sum_{s=0}^{t-1}\lambda^{t-1-s}\alpha(s) - \tfrac{\sigma_{\rm m}}{2}\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2, \qquad (7)
\end{aligned}$$

where $\varpi(t) = P_\Omega(\bar{\zeta}(t))$.

Proof: Let $u_i(t) = \sum_{j=1}^N w_{ij}\zeta_j(t)$ and $\varpi^* \in \Omega$. Then

$$\begin{aligned}
&\|\zeta_i(t+1)-\varpi^*\|^2\\
&= \|u_i(t)-\varpi^*-\alpha(t)\tfrac{g_i(t)}{y_{ii}(t)}+\phi_i(t)\|^2\\
&\le \|u_i(t)-\varpi^*-\alpha(t)\tfrac{g_i(t)}{y_{ii}(t)}\|^2 - \|\phi_i(t)\|^2\\
&= \|u_i(t)-\varpi^*\|^2 + \tfrac{2\alpha(t)}{y_{ii}(t)}g_i(t)^{\rm T}(\varpi^*-u_i(t))\\
&\quad + \tfrac{\alpha^2(t)}{y_{ii}^2(t)}\|g_i(t)\|^2 - \|\phi_i(t)\|^2, \qquad (8)
\end{aligned}$$

where the inequality follows from Lemma 3. Note that

$$\begin{aligned}
&g_i(t)^{\rm T}(\varpi^*-u_i(t))\\
&\le \psi_i(\varpi^*) - \psi_i(u_i(t)) - \tfrac{\mu}{2}\|\varpi^*-u_i(t)\|^2\\
&\le \psi_i(\varpi^*) - \psi_i(\bar{\zeta}(t)) + L\sum_{j=1}^N w_{ij}\|\zeta_j(t)-\bar{\zeta}(t)\|\\
&\quad - \tfrac{\mu}{2}\|\varpi^*-u_i(t)\|^2, \qquad (9)
\end{aligned}$$

where the first inequality holds because of Assumption 2 and the last one follows from Assumption 3 as well as the triangle inequality. In accordance with (8) and (9), we have

$$\begin{aligned}
&\|\zeta_i(t+1)-\varpi^*\|^2 + \|\phi_i(t)\|^2\\
&\le \left(1-\tfrac{\mu\alpha(t)}{y_{ii}(t)}\right)\|u_i(t)-\varpi^*\|^2 + \tfrac{2\alpha(t)}{y_{ii}(t)}\left(\psi_i(\varpi^*)\right.\\
&\quad \left.-\psi_i(\bar{\zeta}(t))\right) + 2L\tfrac{\alpha(t)}{y_{ii}(t)}\sum_{j=1}^N w_{ij}\|\zeta_j(t)-\bar{\zeta}(t)\|\\
&\quad + \alpha^2(t)L^2\eta^2. \qquad (10)
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1)-\varpi^*\|^2\\
&\le \sum_{i\in\mathcal{V}}\sigma_i\left(1-\tfrac{\mu\alpha(t)}{y_{ii}(t)}\right)\|u_i(t)-\varpi^*\|^2\\
&\quad + 2\alpha(t)\sum_{i\in\mathcal{V}}\tfrac{\sigma_i}{y_{ii}(t)}\left(\psi_i(\varpi^*)-\psi_i(\bar{\zeta}(t))\right)\\
&\quad + 2L\sum_{i\in\mathcal{V}}\sigma_i\tfrac{\alpha(t)}{y_{ii}(t)}\sum_{j=1}^N w_{ij}\|\zeta_j(t)-\bar{\zeta}(t)\|\\
&\quad + \sum_{i\in\mathcal{V}}\sigma_i\alpha^2(t)L^2\eta^2 - \sum_{i\in\mathcal{V}}\sigma_i\|\phi_i(t)\|^2. \qquad (11)
\end{aligned}$$

For the term $\sum_{i\in\mathcal{V}}\frac{\sigma_i}{y_{ii}(t)}(\psi_i(\varpi^*)-\psi_i(\bar{\zeta}(t)))$, we have

$$\begin{aligned}
&\sum_{i\in\mathcal{V}}\tfrac{\sigma_i}{y_{ii}(t)}\left(\psi_i(\varpi^*)-\psi_i(\bar{\zeta}(t))\right)\\
&= \sum_{i\in\mathcal{V}}\left(1+\tfrac{\sigma_i-y_{ii}(t)}{y_{ii}(t)}\right)\left(\psi_i(\varpi^*)-\psi_i(\bar{\zeta}(t))\right)\\
&\le \psi(\varpi^*)-\psi(\bar{\zeta}(t)) + \sum_{i\in\mathcal{V}}\tfrac{|y_{ii}(t)-\sigma_i|}{y_{ii}(t)}|\psi_i(\bar{\zeta}(t))-\psi_i(\varpi^*)|\\
&\le \psi(\varpi^*)-\psi(\bar{\zeta}(t)) + NCL\eta\lambda^t\|\bar{\zeta}(t)-\varpi^*\|, \qquad (12)
\end{aligned}$$

where the last inequality follows from Lemma 1 and Assumption 3. For the term $\sum_{i\in\mathcal{V}}\frac{\sigma_i}{y_{ii}(t)}\sum_{j=1}^N w_{ij}\|\zeta_j(t)-\bar{\zeta}(t)\|$, by using $\sigma^{\rm T}W = \sigma^{\rm T}$ and Lemma 1, we have

$$\sum_{i\in\mathcal{V}}\tfrac{\sigma_i}{y_{ii}(t)}\sum_{j=1}^N w_{ij}\|\zeta_j(t)-\bar{\zeta}(t)\| \le \eta\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\bar{\zeta}(t)\|. \quad (13)$$

Substituting (12) and (13) into (11),

$$\begin{aligned}
&\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1)-\varpi^*\|^2\\
&\le \sum_{i\in\mathcal{V}}\sigma_i\left(1-\tfrac{\mu\alpha(t)}{y_{ii}(t)}\right)\|u_i(t)-\varpi^*\|^2 + 2\alpha(t)\left(\psi(\varpi^*)\right.\\
&\quad \left.-\psi(\bar{\zeta}(t))\right) + 2\alpha(t)NCL\eta\lambda^t\|\bar{\zeta}(t)-\varpi^*\|\\
&\quad + \alpha^2(t)L^2\eta^2 - \sum_{i\in\mathcal{V}}\sigma_i\|\phi_i(t)\|^2\\
&\quad + 2L\eta\alpha(t)\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\bar{\zeta}(t)\|. \qquad (14)
\end{aligned}$$

Considering the third term on the right side of (14) and leveraging the property of the convex function,

$$\begin{aligned}&2\alpha(t)\lambda^t\|\bar{\zeta}(t)-\varpi^*\|\\&\leq \alpha^2(t)+\lambda^{2t}\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\varpi^*\|^2.\end{aligned}\tag{15}$$

Substituting (15) into (14) yields

$$\begin{aligned}&\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1)-\varpi^*\|^2\\&\leq\sum_{i\in\mathcal{V}}\sigma_i\left(1-\tfrac{\mu\alpha(t)}{y_{ii}(t)}\right)\|u_i(t)-\varpi^*\|^2+(NCL\eta\\&\quad+L^2\eta^2)\alpha^2(t)+2\alpha(t)\left(\psi(\varpi^*)-\psi(\bar{\zeta}(t))\right)\\&\quad-\sum_{i\in\mathcal{V}}\sigma_i\|\phi_i(t)\|^2+2L\eta\alpha(t)\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\bar{\zeta}(t)\|\\&\quad+NCL\eta\lambda^{2t}\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\varpi^*\|^2.\end{aligned}\tag{16}$$

For the term $\sum_{i\in\mathcal{V}}\sigma_i(1-\frac{\mu\alpha(t)}{y_{ii}(t)})\|u_i(t)-\varpi^*\|^2$ in (16),

$$\begin{aligned}&\sum_{i\in\mathcal{V}}\sigma_i\left(1-\tfrac{\mu\alpha(t)}{y_{ii}(t)}\right)\|u_i(t)-\varpi^*\|^2\\&\quad\leq(1-\mu\alpha(t))\sum_{i\in\mathcal{V}}\sigma_i\|\sum_{j\in\mathcal{V}}w_{ij}(\zeta_j(t)-\varpi^*)\|^2\\&\quad\leq(1-\mu\alpha(t))\sum_{j\in\mathcal{V}}\sum_{i\in\mathcal{V}}\sigma_iw_{ij}\|\zeta_j(t)-\varpi^*\|^2\\&\quad=(1-\mu\alpha(t))\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\varpi^*\|^2,\end{aligned}\tag{17}$$

where the first inequality holds due to $y_{ii}(t)\leq 1$, the second one follows from the convexity of the norm function, and the equality is by the row stochastic matrix W. Consequently,

$$\begin{aligned}&\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1)-\varpi^*\|^2\\&\leq(1-\mu\alpha(t)+NCL\eta\lambda^{2t})\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\varpi^*\|^2\\&\quad+2\alpha(t)\left(\psi(\varpi^*)-\psi(\bar{\zeta}(t))\right)+(NCL\eta+L^2\eta^2)\alpha^2(t)\\&\quad+2L\eta\alpha(t)\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\bar{\zeta}(t)\|-\sum_{i\in\mathcal{V}}\sigma_i\|\phi_i(t)\|^2.\end{aligned}\tag{18}$$

Moreover,

$$\begin{aligned}\|\varpi(t)-\bar{\zeta}(t)\|&=\mathrm{dist}(\bar{\zeta}(t),\Omega)\\&\leq R\max_{i\in\mathcal{V}}\mathrm{dist}(\bar{\zeta}(t),\Omega_i)\\&\leq R\sum_{i\in\mathcal{V}}\mathrm{dist}(\bar{\zeta}(t),\Omega_i)\\&\leq R\sum_{i\in\mathcal{V}}\|\zeta_i(t)-\bar{\zeta}(t)\|,\end{aligned}\tag{19}$$

where the first inequality follows from Lemma 2. Then, by utilizing (19),

$$\begin{aligned}&\psi(\varpi^*)-\psi(\bar{\zeta}(t))\\&\leq\psi(\varpi^*)-\psi(\varpi(t))+NLR\sum_{i\in\mathcal{V}}\|\zeta_i(t)-\bar{\zeta}(t)\|.\end{aligned}\tag{20}$$

Substituting (20) into (18) and adding $ab\delta(t+1)$ to both sides of (18), we have

$$\begin{aligned}&\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1)-\varpi^*\|^2+ab\delta(t+1)\\&\leq(1-\mu\alpha(t)+NCL\eta\lambda^{2t})\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\varpi^*\|^2\\&\quad-2\alpha(t)(\psi(\varpi(t))-\psi(\varpi^*))+2\alpha(t)NLR\\&\quad\times\sum_{i\in\mathcal{V}}\|\zeta_i(t)-\bar{\zeta}(t)\|+(NCL\eta+L^2\eta^2)\alpha^2(t)\\&\quad+2L\eta\alpha(t)\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\bar{\zeta}(t)\|-\sum_{i\in\mathcal{V}}\sigma_i\|\phi_i(t)\|^2\\&\quad+ab\delta(t+1).\end{aligned}\tag{21}$$

According to the facts $a=\frac{2CL(\eta+RN^2)}{(1-\lambda)b}$, $b=\sqrt{\frac{\sigma_m}{N}}$, Lemma 4, and (21), we have

$$\begin{aligned}&\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1)-\varpi^*\|^2+ab\delta(t+1)\\&\leq(1-\mu\alpha(t)+NCL\eta\lambda^{2t})\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t)-\varpi^*\|^2\\&\quad+ab\alpha(t)\beta(t)-2\alpha(t)\left(\psi(\varpi(t))-\psi(\varpi^*)\right)\\&\quad+2A_1N^2LR\alpha(t)\lambda^t+ab\delta(t)+2NCL^2\eta(N^2R\\&\quad+\eta)\alpha(t)\sum_{s=0}^{t-1}\lambda^{t-1-s}\alpha(s)+2L\eta A_1\alpha(t)\lambda^t\\&\quad+(NCL\eta+L^2\eta^2)\alpha^2(t)-\sum_{i\in\mathcal{V}}\sigma_i\|\phi_i(t)\|^2.\end{aligned}\tag{22}$$

Furthermore, substituting $2ab\alpha(t)\beta(t)\leq a^2\alpha^2(t)+b^2N\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2$ into (22), it yields (7). This completes the proof. $\square$

Remark 6: Theorem 1 unveils an essential relationship regarding the evolution of the agents' states. Particularly, we leverage the strong convexity of the local objective functions, which provides a tighter upper bound than that of [36] and is beneficial to achieve a faster convergence rate than the counterpart of [36].

Lemma 6: If $\alpha(t)=\frac{G}{\mu t}$ and $G\geq\frac{-NCL\eta}{2e\ln\lambda}$, then there holds $NCL\eta\lambda^{2t}-\mu\alpha(t)\leq 0, \forall t>0$.

Proof: To make the inequality $NCL\eta\lambda^{2t}-\mu\alpha(t)\leq 0$ hold, we need $NCL\eta\lambda^{2t}\leq\mu\alpha(t)=\frac{G}{t}$, namely, $G\geq NCL\eta t\lambda^{2t}$. Let $h(t)=t\lambda^{2t}$, then $h'(t)=\lambda^{2t}(1+2t\ln\lambda)$,

$$h'(t)\begin{cases}>0 & \text{for}\quad 0<t<-1/2\ln\lambda,\\=0 & \text{for}\quad t=-1/2\ln\lambda,\\<0 & \text{for}\quad t>-1/2\ln\lambda.\end{cases}$$

Thus, when $0<t<-1/2\ln\lambda$, $h(t)$ monotonically increases, while when $t>-1/2\ln\lambda$, $h(t)$ monotonically decreases. $h(t)$ reaches its maximum value at $t=-1/2\ln\lambda$ and $h(\frac{-1}{2\ln\lambda})=\frac{-1}{2\ln\lambda}\cdot\lambda^{\frac{-1}{\ln\lambda}}=\frac{-1}{2\ln\lambda}\cdot e^{\ln\lambda^{-1/\ln\lambda}}=\frac{-1}{2\ln\lambda}\cdot e^{\frac{-1}{\ln\lambda}\cdot\ln\lambda}=\frac{-1}{2e\ln\lambda}$. So $G\geq\frac{-NCL\eta}{2e\ln\lambda}$. $\square$

Next, we are prepared to give the convergence performance of the DPSG algorithm.

Theorem 2: Suppose Assumptions 1–3 hold. If positive decaying step size $\alpha(t) = \frac{G}{\mu t}, t > 0$, then $\zeta_i(t)$ produced by the DPSG algorithm converges to an optimal solution, i.e., for any $i \in \mathcal{V}$, $\lim_{t\to\infty} \zeta_i(t) = \varpi^*$ with ϖ^* being an optimal solution.

Proof: For the convenience of analysis, we define

$$e(t) = \sum_{i\in\mathcal{V}} \sigma_i \|\zeta_i(t) - \varpi^*\|^2 + ab\delta(t),$$

$$q(t) = A_3\alpha(t)\lambda^t + A_2\alpha(t)\sum_{s=0}^{t-1}\lambda^{t-1-s}\alpha(s) + A_4\alpha^2(t),$$

$$l(t) = 2\alpha(t)\left(\psi(\varpi(t)) - \psi(\varpi^*)\right) + \tfrac{\sigma_{\mathrm{m}}}{2}\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2.$$

Then, we rewrite (7) as follows

$$\begin{aligned} e(t+1) &\leq e(t) - l(t) + q(t) + (NCL\eta\lambda^{2t} - \mu\alpha(t)) \\ &\quad \times \sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi^*\|^2) \\ &\leq e(t) - l(t) + q(t), \end{aligned} \tag{23}$$

where the second inequality follows from Lemma 6. Rearranging and summing the terms from $t = 1$ to T, we have

$$\begin{aligned} \sum_{t=1}^{T} l(t) &\leq \sum_{t=1}^{T}(e(t) - e(t+1)) + \sum_{t=1}^{T} q(t) \\ &\leq e(1) + \sum_{t=1}^{T} q(t). \end{aligned} \tag{24}$$

Besides, it follows from the triangle inequality that $A_3\alpha(t)\lambda^t \leq \alpha^2(t) + \frac{A_3^2}{4}\lambda^{2t}$, then we have $\sum_{t=0}^{\infty} A_3\alpha(t)\lambda^t \leq \sum_{t=0}^{\infty}\alpha^2(t) + \frac{A_3^2}{4}\sum_{t=0}^{\infty}\lambda^{2t} < \infty$. Since $\{\alpha(t)\}$ is a nonincreasing sequence, using the fact $A_2\alpha(t)\sum_{s=0}^{t-1}\lambda^{t-1-s}\alpha(s) \leq A_2\sum_{s=0}^{t-1}\lambda^{t-1-s}\alpha^2(s)$, we have

$$\begin{aligned} \sum_{t=0}^{\infty} A_2\alpha(t)\sum_{s=0}^{t-1}\lambda^{t-1-s}\alpha(s) &\leq \tfrac{A_2}{1-\lambda}\sum_{t=0}^{\infty}\alpha^2(t) \\ &< \infty. \end{aligned} \tag{25}$$

Thus, $\{q(t)\}$ is summable. By (24), when T tends towards infinity,

$$\begin{cases} \sum_{t=0}^{\infty}\alpha(t)\left(\psi(\varpi(t)) - \psi(\varpi^*)\right) < \infty, \\ \sum_{t=0}^{\infty}\varphi(t) < \infty. \end{cases} \tag{26}$$

Combining with $\sum_{t=0}^{\infty}\alpha(t) = \infty$ and (26),

$$\begin{cases} \liminf_{t\to\infty}\left(\psi(\varpi(t)) - \psi(\varpi^*)\right) = 0, \\ \lim_{t\to\infty}\varphi(t) = 0. \end{cases} \tag{27}$$

Based on (4), (27), and $\lim_{t\to\infty}\alpha(t) = 0$, we have

$$\begin{cases} \lim_{t\to\infty}\beta(t) = 0, \\ \lim_{t\to\infty}\|\zeta_i(t) - \bar{\zeta}(t)\| = 0, \\ \lim_{t\to\infty}\delta(t) = 0. \end{cases}$$

By (27),

$$\lim_{k\to\infty}\psi(\varpi(t_k)) - \psi^* = 0. \tag{28}$$

Since $\{\varpi(t_k)\}$ is bounded, there exists a convergent subsequence $\{\varpi(t_l)\}$. Assuming that $\lim_{l\to\infty}\varpi(t_l) = \varpi_0$, we have $\psi^* = \lim_{k\to\infty}\psi(\varpi(t_k)) = \lim_{l\to\infty}\psi(\varpi(t_l)) = \psi(\varpi_0)$ and $\lim_{t\to\infty}\|\varpi(t) - \varpi_0\| = 0$. Selecting $\varpi^* = \varpi_0$ and considering that $\|\zeta_i(t) - \varpi_0\|^2 \leq 3(\|\zeta_i(t) - \bar{\zeta}(t)\|^2 + \|\bar{\zeta}(t) - \varpi(t)\|^2 + \|\varpi(t) - \varpi_0\|^2)$, we have

$$\begin{aligned} &\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi_0\|^2 \\ &\leq 3\left(\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \bar{\zeta}(t)\|^2\right. \\ &\qquad \left. + \sum_{i\in\mathcal{V}}\sigma_i\|\bar{\zeta}(t) - \varpi(t)\|^2 + \sum_{i\in\mathcal{V}}\sigma_i\|\varpi(t) - \varpi_0\|^2\right). \end{aligned} \tag{29}$$

Lastly, taking the limit of (29) as t approaches infinity yields $\lim_{t\to\infty}\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi_0\|^2 = c \leq 0$. Based the fact $c \geq 0$, there exists an optimal solution ϖ_0 that makes $c = 0$, which leads the conclusion. □

Remark 7: In contrast to the utilizing of supermartingale convergence lemma in [32], [36], we provide Lemma 6 to simplify (23) as (24), which is more practical for convergence analysis.

Next, we will discuss the convergence rate of the DPSG algorithm after selecting a certain step size. To begin with, we first define two variables as follows: $\hat{\varpi}(T) = \frac{2\sum_{t=1}^{T} t\varpi(t)}{T(T+1)}$ and $\hat{\zeta}_i(T) = \frac{2\sum_{t=1}^{T} t\zeta_i(t)}{T(T+1)}$.

Theorem 3: Suppose Assumptions 1–3 hold. With regard to the DPSG algorithm with $\alpha(t) = \frac{Q}{\mu t}$ for $t > 0$, there holds

$$\tfrac{Q}{\mu}\left(\psi(\hat{\varpi}(T)) - \psi(\varpi^*)\right) + H\|\hat{\zeta}_i(T) - \hat{\varpi}(T)\| = O\left(\tfrac{1}{T}\right),$$

where $Q = \max\{G, 16\}$ and $H = \frac{1}{4C(RN+1)}$. Moreover, there holds

$$|\psi(\hat{\zeta}_i(T)) - \psi(\varpi^*)| = O(\tfrac{1}{T}).$$

Proof: The process of proving is divided into five sequential steps. Firstly, according to global function ψ is μ-strongly convex and Theorem 1, we give an upper bound of the term $\sum_{t=1}^{T}\frac{Q}{\mu}t[\psi(\varpi(t)) - \psi(\varpi^*)] + \frac{\sigma_{\mathrm{m}}}{2}\sum_{t=1}^{T} t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2$. Then, through disposing of some variables in this bound and a new bound will be presented for $\sum_{t=t_1}^{T}\frac{Q}{\mu}t(\psi(\varpi(t)) - \psi(\varpi^*)) + \frac{\sigma_{\mathrm{m}}}{4}\sum_{t=t_1}^{T} t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2$ in step 2. Besides, another part of the relevant variables involved will be described to the upper bound for the term $\sum_{t=t_3}^{T}\frac{Q}{\mu}t[\psi(\varpi(t)) - \psi(\varpi^*)] + \frac{\sigma_{\mathrm{m}}}{16}\sum_{t=t_3}^{T} t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2$ in step 3. Moreover, the term $H\sum_{t=t_3}^{T} t\|\zeta_i(t) - \varpi(t)\|$ will be bounded in the subsequent step 4. At last, we will obtain a bound of $O(T)$ for $\sum_{t=t_3}^{T}\frac{Q}{\mu}t(\psi(\varpi(t)) - \psi(\varpi^*)) + H\sum_{t=t_3}^{T} t\|\zeta_i(t) - \varpi(t)\|$ and the proof will be completed via the usage of convexity and Lipschitz continuity of ψ.

Step 1: Give a bound of $\sum_{t=1}^{T}\frac{Q}{\mu}t[\psi(\varpi(t)) - \psi(\varpi^*)] + \frac{\sigma_{\mathrm{m}}}{2}\sum_{t=1}^{T} t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2$.

Owing to ψ is μ-strongly convex,

$$\psi(\varpi(t)) - \psi(\varpi^*) \geq \frac{\mu}{2}\|\varpi(t) - \varpi^*\|^2. \tag{30}$$

Noting that $\varpi(t) - \varpi^* = \varpi(t) - \bar{\zeta}(t) + \bar{\zeta}(t) - \varpi^*$,

$$\begin{aligned}&\|\varpi(t) - \varpi^*\|^2\\ &\geq 2(\varpi(t) - \bar{\zeta}(t))^{\mathrm{T}}(\bar{\zeta}(t) - \varpi^*) + \|\bar{\zeta}(t) - \varpi^*\|^2\\ &\geq -2\|\varpi(t) - \bar{\zeta}(t)\|^2 + \tfrac{1}{2}\|\bar{\zeta}(t) - \varpi^*\|^2. \end{aligned}\tag{31}$$

Thus, considering (7), (30), and (31), we get

$$\begin{aligned}&\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1) - \varpi^*\|^2 + \tfrac{\sigma_{\mathrm{m}}}{2}\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2\\ &\leq (1 - \mu\alpha(t) + NCL\eta\lambda^{2t})\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi^*\|^2\\ &\quad - \alpha(t)\left(\psi(\varpi(t)) - \psi(\varpi^*)\right) + A_3\alpha(t)\lambda^t\\ &\quad + A_4\alpha^2(t) + A_2\alpha(t)\sum_{s=0}^{t-1}\lambda^{t-1-s}\alpha(s)\\ &\quad + ab(\delta(t) - \delta(t+1)) + \mu\alpha(t)\|\varpi(t) - \bar{\zeta}(t)\|^2\\ &\quad - \tfrac{\alpha(t)}{4}\mu\|\bar{\zeta}(t) - \varpi^*\|^2. \end{aligned}\tag{32}$$

Substituting $\alpha(t) = \frac{Q}{\mu t}$ into (32) yields

$$\begin{aligned}&\tfrac{Q}{\mu t}\left(\psi(\varpi(t)) - \psi(\varpi^*)\right)\\ &\leq \sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi^*\|^2 - \sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1) - \varpi^*\|^2\\ &\quad + (NCL\eta\lambda^{2t} - \tfrac{Q}{t})\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi^*\|^2\\ &\quad - \tfrac{\sigma_{\mathrm{m}}}{2}\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2 + ab(\delta(t) - \delta(t+1))\\ &\quad + \tfrac{QA_3}{\mu t}\lambda^t + \tfrac{Q^2A_2}{\mu^2 t}\sum_{s=0}^{t-1}\lambda^{t-1-s}\tfrac{1}{s} + \tfrac{Q^2A_4}{\mu^2t^2}\\ &\quad - \tfrac{Q}{4t}\|\bar{\zeta}(t) - \varpi^*\|^2 + \tfrac{Q}{t}\|\varpi(t) - \bar{\zeta}(t)\|^2. \end{aligned}\tag{33}$$

According to Lemma 6 and $Q = \max\{G, 16\}$, we have

$$\begin{aligned}&\tfrac{Q}{\mu t}\left(\psi(\varpi(t)) - \psi(\varpi^*)\right)\\ &\leq \sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi^*\|^2 - \sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1) - \varpi^*\|^2\\ &\quad + \tfrac{QA_3}{\mu t}\lambda^t - \tfrac{\sigma_{\mathrm{m}}}{2}\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2 + \tfrac{Q^2A_4}{\mu^2t^2}\\ &\quad - \tfrac{4}{t}\|\bar{\zeta}(t) - \varpi^*\|^2 + ab(\delta(t) - \delta(t+1))\\ &\quad + \tfrac{Q}{t}\|\varpi(t) - \bar{\zeta}(t)\|^2 + \tfrac{Q^2A_2}{\mu^2t}\sum_{s=0}^{t-1}\lambda^{t-1-s}\tfrac{1}{s}. \end{aligned}\tag{34}$$

Multiplying both sides of (34) by t^2,

$$\begin{aligned}&\tfrac{Q}{\mu}t\left(\psi(\varpi(t)) - \psi(\varpi^*)\right)\\ &\leq t^2\left(\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi^*\|^2 - \sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t+1) - \varpi^*\|^2\right)\\ &\quad + \tfrac{Q^2A_4}{\mu^2} + \tfrac{Q^2A_2}{\mu^2}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\tfrac{1}{s} - \tfrac{\sigma_{\mathrm{m}}}{2}t^2\\ &\quad \times\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2 - 4t\|\bar{\zeta}(t) - \varpi^*\|^2 + \tfrac{QA_3}{\mu}t\lambda^t\\ &\quad + abt^2(\delta(t) - \delta(t+1)) + Qt\|\varpi(t) - \bar{\zeta}(t)\|^2. \end{aligned}\tag{35}$$

Taking the summation of (35) from $t = 1$ to T and according to the fact $\sum_{t=1}^{T} t^2(\Xi(t) - \Xi(t+1)) \leq \sum_{t=1}^{T} 2t\Xi(t)$ with $\Xi(t) = \sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi^*\|^2$, yields

$$\begin{aligned}&\sum_{t=1}^{T}\tfrac{Q}{\mu}t\left[\psi(\varpi(t)) - \psi(\varpi^*)\right] + \tfrac{\sigma_{\mathrm{m}}}{2}\sum_{t=1}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2\\ &\leq \sum_{t=1}^{T}2t\left(\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi^*\|^2 - 2\|\bar{\zeta}(t) - \varpi^*\|^2\right)\\ &\quad + \sum_{t=1}^{T}\tfrac{QA_3}{\mu}t\lambda^t + \sum_{t=1}^{T}\tfrac{Q^2A_2}{\mu^2}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\tfrac{1}{s}\\ &\quad + \sum_{t=1}^{T}\tfrac{Q^2A_4}{\mu^2} + \sum_{t=1}^{T}abt^2(\delta(t) - \delta(t+1))\\ &\quad + Q\sum_{t=1}^{T}t\|\varpi(t) - \bar{\zeta}(t)\|^2. \end{aligned}\tag{36}$$

For the term $\sum_{t=1}^{T} 2t(\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi^*\|^2 - 2\|\bar{\zeta}(t) - \varpi^*\|^2)$ in (36), we have

$$\begin{aligned}&\sum_{t=1}^{T}2t\left(\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \varpi^*\|^2 - 2\|\bar{\zeta}(t) - \varpi^*\|^2\right)\\ &= \sum_{t=1}^{T}2t\left(\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \bar{\zeta}(t) + \bar{\zeta}(t) - \varpi^*\|^2\right.\\ &\quad \left. - 2\|\bar{\zeta}(t) - \varpi^*\|^2\right)\\ &\leq \sum_{t=1}^{T}2t\left[\sum_{i\in\mathcal{V}}\sigma_i(2\|\zeta_i(t) - \bar{\zeta}(t)\|^2 + 2\|\bar{\zeta}(t) - \varpi^*\|^2)\right.\\ &\quad \left. - 2\|\bar{\zeta}(t) - \varpi^*\|^2\right]\\ &= \sum_{t=1}^{T}4t\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \bar{\zeta}(t)\|^2. \end{aligned}\tag{37}$$

Substituting (37) into (36),

$$\begin{aligned}&\sum_{t=1}^{T}\tfrac{Q}{\mu}t\left[\psi(\varpi(t)) - \psi(\varpi^*)\right] + \tfrac{\sigma_{\mathrm{m}}}{2}\sum_{t=1}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2\\ &\leq \sum_{t=1}^{T}4t\sum_{i\in\mathcal{V}}\sigma_i\|\zeta_i(t) - \bar{\zeta}(t)\|^2 + \tfrac{Q^2A_2}{\mu^2}\sum_{t=1}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\tfrac{1}{s}\\ &\quad + \tfrac{QA_3}{\mu}\sum_{t=1}^{T}t\lambda^t + \tfrac{Q^2A_4T}{\mu^2} + ab\sum_{t=1}^{T}t^2(\delta(t) - \delta(t+1))\\ &\quad + Q\sum_{t=1}^{T}t\|\varpi(t) - \bar{\zeta}(t)\|^2. \end{aligned}\tag{38}$$

Step 2: Provide a bound for $\sum_{t=t_1}^{T}\frac{Q}{\mu}t(\psi(\varpi(t)) - \psi(\varpi^*)) + \frac{\sigma_{\mathrm{m}}}{4}\sum_{t=t_1}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2$.

Regarding the term $Q\sum_{t=1}^{T}t\|\varpi(t)-\bar{\zeta}(t)\|^2$ in (38), taking (19) into account, we have

$$\|\varpi(t)-\bar{\zeta}(t)\|^2 \leq NR^2\sum_{i\in\mathcal{V}}\|\zeta_i(t)-\bar{\zeta}(t)\|^2. \tag{39}$$

According to (4), we have

$$\begin{aligned}&\|\zeta_i(t)-\bar{\zeta}(t)\|^2\\ &\leq 2A_1^2\lambda^{2t}+2\left(\sum_{s=0}^{t-1}\lambda^{t-1-s}\left(NCL\eta\alpha(s)+C\beta(s)\right)\right)^2.\end{aligned} \tag{40}$$

Using Lemma 5 directly gives that

$$\begin{aligned}&\left(\sum_{s=0}^{t-1}\lambda^{t-1-s}\left(NCL\eta\alpha(s)+C\beta(s)\right)\right)^2\\ &\leq \frac{1}{1-\lambda}\sum_{s=0}^{t-1}\lambda^{t-1-s}\left(NCL\eta\alpha(s)+C\beta(s)\right)^2\\ &\leq \frac{2N^2C^2L^2\eta^2}{1-\lambda}\sum_{s=0}^{t-1}\lambda^{t-1-s}\alpha^2(s)\\ &\quad+\frac{2C^2}{1-\lambda}\sum_{s=0}^{t-1}\lambda^{t-1-s}\beta^2(s).\end{aligned} \tag{41}$$

Substituting (41) into (40), yields

$$\begin{aligned}&\|\zeta_i(t)-\bar{\zeta}(t)\|^2\\ &\leq 2A_1^2\lambda^{2t}+\frac{4N^2C^2L^2\eta^2}{1-\lambda}\sum_{s=0}^{t-1}\lambda^{t-1-s}\alpha^2(s)\\ &\quad+\frac{4C^2}{1-\lambda}\sum_{s=0}^{t-1}\lambda^{t-1-s}\beta^2(s).\end{aligned} \tag{42}$$

By (38), (39), (42), and $\beta(t)=\sum_{i\in\mathcal{V}}\|\phi_i(t)\|$,

$$\begin{aligned}&\sum_{t=1}^{T}\frac{Q}{\mu}t\left[\psi(\varpi(t))-\psi(\varpi^*)\right]+\frac{\sigma_{\mathrm{m}}}{2}\sum_{t=1}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2\\ &\leq ab\sum_{t=1}^{T}t^2(\delta(t)-\delta(t+1))+\frac{QA_3}{\mu}\sum_{t=1}^{T}t\lambda^t\\ &\quad+\frac{Q^2A_4T}{\mu^2}+\frac{Q^2A_2}{\mu^2}\sum_{t=1}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\frac{1}{s}\\ &\quad+A_5\sum_{t=1}^{T}t\lambda^{2t}+A_6\sum_{t=1}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\alpha^2(s)\\ &\quad+A_7\sum_{t=1}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\sum_{i\in\mathcal{V}}\|\phi_i(s)\|^2,\end{aligned} \tag{43}$$

where

$$A_5=2A_1^2(4+QN^2R^2),$$
$$A_6=\frac{4N^2C^2L^2\eta^2(4+QN^2R^2)}{1-\lambda},$$
$$A_7=\frac{4C^2N(4+QN^2R^2)}{1-\lambda}.$$

For the term $ab\sum_{t=1}^{T}t^2(\delta(t)-\delta(t+1))$ in (43), noting that $t^2(\delta(t)-\delta(t+1))=(t-1)^2\delta(t)-t^2\delta(t+1)+(2t-1)\delta(t)$, we have

$$\sum_{t=2}^{T}t^2(\delta(t)-\delta(t+1))\leq\delta(2)+2\sum_{t=2}^{T}t\delta(t). \tag{44}$$

Combining (43) and (44), it yields

$$\begin{aligned}&\sum_{t=2}^{T}\frac{Q}{\mu}t\left[\psi(\varpi(t))-\psi(\varpi^*)\right]+\frac{\sigma_{\mathrm{m}}}{2}\sum_{t=2}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2\\ &\leq\xi_1(T)+\frac{Q^2A_6}{\mu^2}\sum_{t=2}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\frac{1}{s^2}\\ &\quad+\frac{Q^2A_2}{\mu^2}\sum_{t=2}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\frac{1}{s}+2ab\sum_{t=2}^{T}t\delta(t)\\ &\quad+A_7\sum_{t=2}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\sum_{i\in\mathcal{V}}\|\phi_i(s)\|^2,\end{aligned} \tag{45}$$

where

$$\xi_1(T)=A_5\sum_{t=2}^{T}t\lambda^{2t}+\frac{QA_3}{\mu}\sum_{t=2}^{T}t\lambda^t+\frac{Q^2A_4T}{\mu^2}+ab\delta(2).$$

For the term $A_7\sum_{t=2}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\sum_{i\in\mathcal{V}}\|\phi_i(s)\|^2$ in (45), there exists a positive constant $t_1\geq 2$ such that for any $t\geq t_1$,

$$\begin{aligned}&A_7\sum_{t=2}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\sum_{i\in\mathcal{V}}\|\phi_i(s)\|^2\\ &<\frac{A_7T}{1-\lambda}\sum_{t=0}^{T-1}\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2\\ &\leq\frac{\sigma_{\mathrm{m}}}{4}\sum_{t=2}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2.\end{aligned} \tag{46}$$

Thus, we gain

$$\begin{aligned}&\sum_{t=t_1}^{T}\frac{Q}{\mu}t\left[\psi(\varpi(t))-\psi(\varpi^*)\right]+\frac{\sigma_{\mathrm{m}}}{4}\sum_{t=t_1}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2\\ &\leq\xi_2(T)+\frac{Q^2A_6}{\mu^2}\sum_{t=t_1}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\frac{1}{s^2}\\ &\quad+\frac{Q^2A_2}{\mu^2}\sum_{t=t_1}^{T}t\sum_{s=0}^{t-1}\lambda^{t-1-s}\frac{1}{s}+2ab\sum_{t=t_1}^{T}t\delta(t),\end{aligned} \tag{47}$$

where $\xi_2(T)$ is defined by replacing $t=t_1$ in all terms of $\xi_1(T)$.

Step 3: Offer a bound of $\sum_{t=t_3}^{T}\frac{Q}{\mu}t[\psi(\varpi(t))-\psi(\varpi^*)]+\frac{\sigma_{\mathrm{m}}}{16}\sum_{t=t_3}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2$.

Letting $\theta(t)=t\eta(t)$, we have

$$\begin{aligned}\theta(t+1)&=(t+1)\left(\sum_{s=0}^{t-1}\lambda^{t-s}\varphi(s)+\varphi(t)\right)\\ &=\frac{t+1}{t}t\sum_{s=0}^{t-1}\lambda^{t-s}\varphi(s)+(t+1)\varphi(t)\\ &=\frac{t+1}{t}\lambda\theta(t)+(t+1)\varphi(t).\end{aligned} \tag{48}$$

Taking into account that $0<\lambda<1$ and $\lim_{t\to\infty}\frac{t+1}{t}=1$, there exist λ_1 and t_2 satisfying $\lambda<\lambda_1<1$ and $t_2\geq t_1$. Then for any $t\geq t_2$,

$$0<\tfrac{t+1}{t}\lambda\leq\lambda_1<1. \tag{49}$$

By (48) and (49),

$$\theta(t+1)\leq\lambda_1\theta(t)+(t+1)\varphi(t),\quad t\geq t_2, \tag{50}$$

which indicates that for all $T\geq t_2+1$,

$$\sum_{t=t_2}^{T-1}\theta(t+1)\leq\lambda_1\sum_{t=t_2}^{T-1}\theta(t)+\sum_{t=t_2}^{T-1}(t+1)\varphi(t), \tag{51}$$

that is,

$$\sum_{t=t_2+1}^{T}\theta(t)\leq\lambda_1\sum_{t=t_2+1}^{T}\theta(t)+\sum_{t=t_2+1}^{T}(t+1)\varphi(t)+D_1, \tag{52}$$

where $D_1=\lambda_1\theta(t_2)+(t_2+1)\varphi(t_2)$. Then,

$$\sum_{t=t_2+1}^{T}\theta(t)\leq\tfrac{1}{1-\lambda_1}\sum_{t=t_2+1}^{T}(t+1)\varphi(t)+\tfrac{D_1}{1-\lambda_1}. \tag{53}$$

As a result, if $\varphi(t)=\frac{1}{t}$, we deduce that

$$\sum_{t=t_2+1}^{T}t\sum_{s=0}^{t-1}\lambda^{t-s-1}\tfrac{1}{s}\leq\tfrac{1}{1-\lambda_1}\sum_{t=t_2+1}^{T}\tfrac{t+1}{t}+\tfrac{D_2}{1-\lambda_1}, \tag{54}$$

where $D_2=\lambda_1 t_2\sum_{s=0}^{t_2-1}\lambda^{t_2-s-1}\frac{1}{s}+(t_2+1)\varphi(t_2)$. Similarly, if $\varphi(t)=\frac{1}{t^2}$, we have

$$\sum_{t=t_2+1}^{T}t\sum_{s=0}^{t-1}\lambda^{t-s-1}\tfrac{1}{s^2}\leq\tfrac{1}{1-\lambda_1}\sum_{t=t_2+1}^{T}\tfrac{t+1}{t^2}+\tfrac{D_3}{1-\lambda_1}, \tag{55}$$

where $D_3=\lambda_1 t_2\sum_{s=0}^{t_2-1}\lambda^{t_2-s-1}\frac{1}{s^2}+(t_2+1)\varphi(t_2)$. Moreover, if $\varphi(t)=\sum_{i\in\mathcal{V}}\|\phi_i(t)\|$, we also have

$$\begin{aligned}&\sum_{t=t_2+1}^{T}t\sum_{s=0}^{t-1}\lambda^{t-s-1}\sum_{i\in\mathcal{V}}\|\phi_i(s)\|\\&\leq\tfrac{1}{1-\lambda_1}\sum_{t=t_2+1}^{T}(t+1)\sum_{i\in\mathcal{V}}\|\phi_i(t)\|+\tfrac{D_4}{1-\lambda_1},\end{aligned} \tag{56}$$

where $D_4=\lambda_1 t_2\sum_{s=0}^{t_2-1}\lambda^{t_2-s-1}\sum_{i\in\mathcal{V}}\|\phi_i(s)\|+(t_2+1)\varphi(t_2)$. By (47), (54)–(56),

$$\begin{aligned}&\sum_{t=t_2+1}^{T}\tfrac{Q}{\mu}t\left[\psi(\varpi(t))-\psi(\varpi^*)\right]+\tfrac{\sigma_{\mathrm{m}}}{4}\sum_{t=t_2+1}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2\\&\leq\xi_3(T),\end{aligned} \tag{57}$$

where

$$\begin{aligned}\xi_3(T)&=\xi_{31}(T)+\xi_{32}(T),\\ \xi_{31}(T)&=\frac{2Qab}{\mu(1-\lambda_1)}\sum_{t=t_2+1}^{T}\tfrac{t+1}{t}\sum_{i\in\mathcal{V}}\|\phi_i(t)\|,\\ \xi_{32}(T)&=\frac{Q^2(A_2D_2+A_6D_3)}{\mu^2(1-\lambda_1)}+ab\delta(2)+\frac{Q^2A_4T}{\mu^2}\\&\quad+\frac{QA_3}{\mu}\sum_{t=t_2+1}^{T}t\lambda^t+A_5\sum_{t=t_2+1}^{T}t\lambda^{2t}\\&\quad+\frac{2QabD_4}{\mu(1-\lambda_1)t}+\frac{Q^2A_6}{\mu^2(1-\lambda_1)}\sum_{t=t_2+1}^{T}\frac{t+1}{t^2}\\&\quad+\frac{Q^2A_2}{\mu^2(1-\lambda_1)}\sum_{t=t_2+1}^{T}\frac{t+1}{t}.\end{aligned}$$

Moreover, we have

$$\begin{aligned}&\sum_{i\in\mathcal{V}}\|\phi_i(t)\|\\&\leq\tfrac{N(t+1)}{2\sigma_{\mathrm{m}}(1-\lambda_1)t^2}+\tfrac{\sigma_{\mathrm{m}}(1-\lambda_1)t^2}{2N(t+1)}\left(\sum_{i\in\mathcal{V}}\|\phi_i(t)\|\right)^2\\&\leq\tfrac{N(t+1)}{2\sigma_{\mathrm{m}}(1-\lambda_1)t^2}+\tfrac{\sigma_{\mathrm{m}}(1-\lambda_1)t^2}{2(t+1)}\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2.\end{aligned} \tag{58}$$

For the term $\frac{2Qab}{\mu(1-\lambda_1)}\sum_{t=t_2+1}^{T}\frac{t+1}{t}\sum_{i\in\mathcal{V}}\|\phi_i(t)\|$ in (57), we obtain

$$\begin{aligned}&\frac{2Qab}{\mu(1-\lambda_1)}\sum_{t=t_2+1}^{T}\tfrac{t+1}{t}\sum_{i\in\mathcal{V}}\|\phi_i(t)\|\\&\leq\tfrac{QabN}{\mu\sigma_{\mathrm{m}}(1-\lambda_1)^2}\sum_{t=t_2+1}^{T}\tfrac{(t+1)^2}{t^3}\\&\quad+\tfrac{Qab\sigma_{\mathrm{m}}}{\mu}\sum_{t=t_2+1}^{T}t\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2.\end{aligned} \tag{59}$$

There exists $t_3\geq t_2+1$ such that, $\forall t\geq t_3$,

$$\tfrac{Qab\sigma_{\mathrm{m}}}{\mu}t\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2\leq\tfrac{\sigma_{\mathrm{m}}}{8}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2. \tag{60}$$

Consequently, we have

$$\begin{aligned}&\sum_{t=t_3}^{T}\tfrac{Q}{\mu}t\left[\psi(\varpi(t))-\psi(\varpi^*)\right]+\tfrac{\sigma_{\mathrm{m}}}{8}\sum_{t=t_3}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2\\&\leq\xi_4(T),\end{aligned} \tag{61}$$

where

$$\begin{aligned}\xi_4(T)&=\xi_{41}(T)+\xi_{42}(T),\\ \xi_{41}(T)&=\tfrac{Q^2A_2}{\mu^2(1-\lambda_1)}\sum_{t=t_3}^{T}\tfrac{t+1}{t}+\tfrac{Q^2A_6}{\mu^2(1-\lambda_1)}\sum_{t=t_3}^{T}\tfrac{t+1}{t^2}\\&\quad+\tfrac{2QabD_4}{\mu(1-\lambda_1)t}+\tfrac{Q^2A_4T}{\mu^2}+\tfrac{QabN}{\mu\sigma_{\mathrm{m}}(1-\lambda_1)^2}\\&\quad\times\sum_{t=t_3}^{T}\tfrac{(t+1)^2}{t^3},\\ \xi_{42}(T)&=\tfrac{QA_3}{\mu}\sum_{t=t_3}^{T}t\lambda^t+A_5\sum_{t=t_3}^{T}t\lambda^{2t}+ab\delta(2)\\&\quad+\tfrac{Q^2(A_2D_2+A_6D_3)}{\mu^2(1-\lambda_1)}.\end{aligned}$$

Step 4: Bound the term $H\sum_{t=t_3}^{T} t\|\zeta_i(t)-\varpi(t)\|$.
Note that

$$\begin{aligned}&\|\zeta_i(t)-\varpi(t)\|\\&\le\|\zeta_i(t)-\bar{\zeta}(t)\|+\|\bar{\zeta}(t)-\varpi(t)\|\\&\le\|\zeta_i(t)-\bar{\zeta}(t)\|+R\sum_{i\in\mathcal{V}}\|\zeta_i(t)-\bar{\zeta}(t)\|\\&\le(RN+1)\left[A_1\lambda^t+C\sum_{s=0}^{t-1}\lambda^{t-1-s}\left(NL\eta\alpha(s)\right.\right.\\&\quad\left.\left.+\beta(s)\right)\right].\end{aligned}\tag{62}$$

where the first inequality follows from adding and subtracting $\|\bar{\zeta}(t)\|$ and triangle inequality, the second one holds based on (19) and the last inequality follows from (4). Consequently,

$$\begin{aligned}&H\sum_{t=t_3}^{T}t\|\zeta_i(t)-\varpi(t)\|\\&\le\tfrac{A_1}{4C}\sum_{t=t_3}^{T}t\lambda^t+\tfrac{QNL\eta}{4\mu}\sum_{t=t_3}^{T}t\sum_{s=0}^{t-1}\lambda^{t-s-1}\tfrac{1}{s}\\&\quad+\tfrac{1}{4}\sum_{t=t_3}^{T}t\sum_{s=0}^{t-1}\lambda^{t-s-1}\sum_{i\in\mathcal{V}}\|\phi_i(s)\|,\end{aligned}\tag{63}$$

with $H=\frac{1}{4C(RN+1)}$. By (54), (56), (58), and (63), we have

$$\begin{aligned}&H\sum_{t=t_3}^{T}t\|\zeta_i(t)-\varpi(t)\|\\&\le\xi_5(T)+\tfrac{\sigma_{\rm m}}{8}\sum_{t=t_3}^{T}t^2\sum_{i\in\mathcal{V}}\|\phi_i(t)\|^2,\end{aligned}\tag{64}$$

where

$$\begin{aligned}\xi_5(T)=&\tfrac{A_1}{4C}\sum_{t=t_3}^{T}t\lambda^t+\tfrac{QNL\eta D_2}{4\mu(1-\lambda_1)}+\tfrac{D_4}{4(1-\lambda_1)}\\&+\tfrac{N}{8\sigma_{\rm m}(1-\lambda_1)^2}\sum_{t=t_3}^{T}\tfrac{(t+1)^2}{t^2}\\&+\tfrac{QNL\eta}{4\mu(1-\lambda_1)}\sum_{t=t_3}^{T}\tfrac{t+1}{t}.\end{aligned}$$

Step 5: Obtain ultimate convergence rate result.
Combining (61) and (64) gives

$$\begin{aligned}&\sum_{t=t_3}^{T}\tfrac{Q}{\mu}t\left(\psi(\varpi(t))-\psi(\varpi^*)\right)+H\sum_{t=t_3}^{T}t\|\zeta_i(t)-\varpi(t)\|\\&\le\xi_4(T)+\xi_5(T).\end{aligned}\tag{65}$$

Because of $\sum_{t=1}^{t_3}\frac{Q}{\mu}t(\psi(\varpi(t))-\psi(\varpi^*))+H\sum_{t=1}^{t_3}t\|\zeta_i(t)-\varpi(t)\|\le O(T)$, $\xi_4(T)=O(T)$, and $\xi_5(T)=O(T)$,

$$\frac{\sum_{t=1}^{T}[\frac{Q}{\mu}t(\psi(\varpi(t))-\psi(\varpi^*))+Ht\|\zeta_i(t)-\varpi(t)\|]}{\frac{T(T+1)}{2}}=O\left(\tfrac{1}{T}\right).\tag{66}$$

Considering the convexity of ψ, we have

$$\psi(\hat{\varpi}(T))-\psi(\varpi^*)\le\frac{\sum_{t=1}^{T}t(\psi(\varpi(t))-\psi(\varpi^*))}{\frac{T(T+1)}{2}},\tag{67}$$

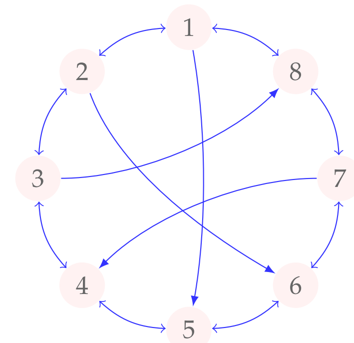

Fig. 1. Communication graph of the example.

and

$$\|\hat{\zeta}_i(T)-\hat{\varpi}(T)\|\le\frac{\sum_{t=1}^{T}t\|\zeta_i(t)-\varpi(t)\|}{\frac{T(T+1)}{2}}.\tag{68}$$

Therefore, we completed the proof. □

Remark 8: For the proposed algorithm, we provided an $O(1/T)$ convergence rate from the perspective of the weighted averaging $\hat{\varpi}(T)=\frac{2\sum_{t=1}^{T}t\varpi(t)}{T(T+1)}$, which is better than the unclear convergence rate in [16], the $O(\ln T/\sqrt{T})$ convergence rate in [25], [27], [30], [36] and the $O(1/\sqrt{T})$ counterpart in [17], [19], [20], [31], [33]. Moreover, compared with the gradient Lipschitz continuous in [26], this paper only needs the objective function being Lipschitz continuous and also achieves the same convergence rate.

V. Simulation

In this section, to verify the effectiveness of the previous theoretical analysis, we consider a machine learning problem as follows

$$\min_{\zeta\in\Omega}\quad\psi(\zeta):=\sum_{i=1}^{N}\psi_i(\zeta),\tag{69}$$

where $\zeta=[\zeta^1,\zeta^2]^\top$ and

$$\psi_i(\zeta)=\begin{cases}\ln\left[1+e^{-a_i(b_i\zeta^1+\zeta^2)}\right]+\frac{1}{2}\|\zeta\|^2, & i=1,\\ \ln\left[1+e^{-a_i(b_i\zeta^1+\zeta^2)}\right], & \text{otherwise},\end{cases}$$

with $a_i=(-1)^i$ and $b_i=0.01i$. Therefore, the global objective function $\psi(\zeta)$ is strongly convex.

Let $N=8$ and Fig. 1 provides the network communication topology. Additionally, we choose N constrained sets as

$$\zeta\in\Omega_i=\begin{bmatrix}0.1i-0.5\\0.1i+0.5\end{bmatrix}\times\begin{bmatrix}0.1i\\0.1i+1\end{bmatrix},\tag{70}$$

which implies the corresponding intersection $\Omega=[0.3,0.6]\times[0.8,1.1]$.

Firstly, we investigate the convergence performance of the proposed DPSG with different step sizes. By designing the positive nonincreasing step sizes as $\alpha(t)=\frac{2\times10^{-4}}{(t+1)^{0.6}}$ and $\alpha(t)=\frac{0.01}{t}$, $t\ge0$, we provide the trajectories of $\zeta_i(t)$ in Fig. 2. From Fig. 2,

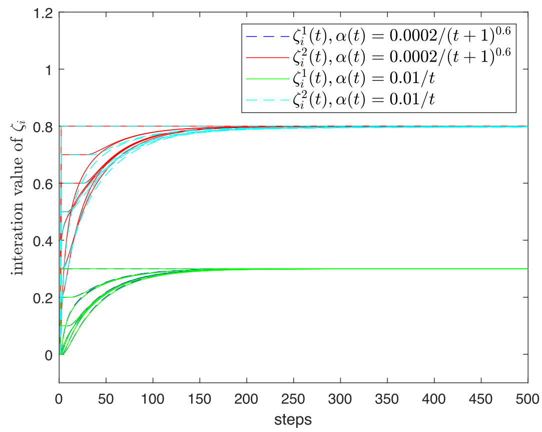

Fig. 2. Performances of variables $\zeta_i(t)$.

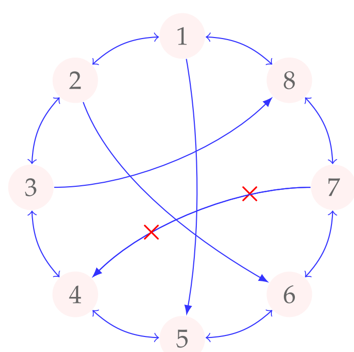

Fig. 3. Communication graph with a broken path.

the performances of the variables $\zeta_i(t)$ under different step sizes eventually converge to the optimal solution $\zeta^* = (0.3, 0.8)^{\mathrm{T}}$.

Moreover, one of the main contributions of this paper is increasing the robustness of the optimization algorithm for communication topology by modifying the column stochastic matrix as a row stochastic one. In the next, we take a comparison with the column stochastic matrix based method, which is called "PSCOA" in [32]. We consider the following two circumstances with $\alpha(t) = \frac{2\times10^{-4}}{(t+1)^{0.6}}$ and $\alpha(t) = \frac{0.01}{t}$.

1) The communication graph is shown in Fig. 1 and each agent exactly knows its out-degree and in-degree information;
2) The communication graph is shown in Fig. 3, where the path from agent 7 to 4 is broken but agent 7 does not detect the broken link. Therefore, each agent exactly knows its in-degree information and agent 7 has incorrect out-degree information.

Figs. 4 and 5 show the relative errors $\frac{\|\zeta(t)-\zeta^*\|^2}{\|\zeta(1)-\zeta^*\|^2}$ of DPSG and PSCOA under the mentioned two circumstances, respectively. In Fig. 4, only if each agent under DPSG knows its in-degree and the counterpart under PSCOA knows its out-degree, both DPSG and PSCOA can achieve the optimal solution. Nevertheless, as depicted in Fig. 5, the disrupted path has no impact on the convergence of DPSG, but it leads to unstable convergence in the case of PSCOA.

The simulation results illustrate the effectiveness of the proposed method for solving optimization problems over

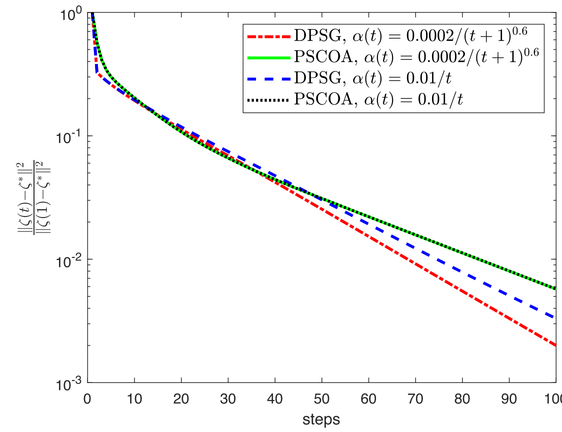

Fig. 4. Performances of DPSG and PSCOA under Fig. 1.

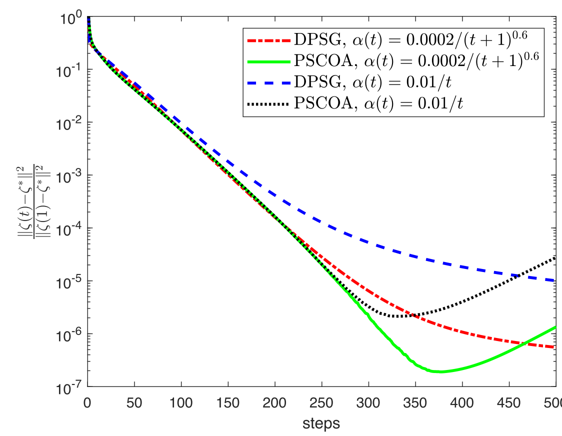

Fig. 5. Performances of DPSG and PSCOA with a broken link.

nonidentical constraints and unbalanced graphs, especially increasing the robustness of the algorithm for the communication topology.

VI. Conclusion

In this paper, we have solved the distributed strongly convex optimization problems with nonidentical, closed and convex sets constraint over unbalanced graphs by the distributed projected subgradient algorithm, which includes a row stochastic weight matrix. Besides, we have established a convergence rate of $O\left(\frac{1}{T}\right)$, which coincides with the convergence rate in the centralized base method. For future work, we will focus on designing the relevant distributed algorithm with a row stochastic weight matrix to solve optimization problems with communication delays and event-triggered communication.

References

[1] P. Yi, Y. Hong, and F. Liu, "Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems," *Automatica*, vol. 74, pp. 259–269, 2016.

[2] M. Luan, G. Wen, Y. Lv, J. Zhou, and C. P. Chen, "Distributed constrained optimization over unbalanced time-varying digraphs: A randomized constraint solving algorithm," *IEEE Trans. Autom. Control*, early access, doi: 10.1109/TAC.2023.3347328.

[3] Z. Wang and H. Li, "Edge-based stochastic gradient algorithm for distributed optimization," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1421–1430, Jul.–Sep. 2020.
[4] Y. Wang, Z. Tu, and H. Qin, "Distributed communication-sliding mirror-descent algorithm for nonsmooth resource allocation problem," *J. Syst. Sci. Complexity*, vol. 35, no. 4, pp. 1244–1261, 2022.
[5] L.-N. Liu and G.-H. Yang, "Distributed optimal economic environmental dispatch for microgrids over time-varying directed communication graph," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1913–1924, Apr.–Jun. 2021.
[6] Y. Zhang, Y. Lou, Y. Hong, and L. Xie, "Distributed projection-based algorithms for source localization in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3131–3142, Jun. 2015.
[7] S. Liang, X. Zeng, and Y. Hong, "Distributed nonsmooth optimization with coupled inequality constraints via modified lagrangian function," *IEEE Trans. Autom. Control*, vol. 63, no. 6, pp. 1753–1759, Jun. 2018.
[8] X. Zeng, P. Yi, Y. Hong, and L. Xie, "Distributed continuous-time algorithms for nonsmooth extended monotropic optimization problems," *SIAM J. Control Optim.*, vol. 56, no. 6, pp. 3973–3993, 2018.
[9] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM J. Optim.*, vol. 25, no. 2, pp. 944–966, 2015.
[10] C. Xi and U. A. Khan, "DEXTRA: A fast algorithm for optimization over directed graphs," *IEEE Trans. Autom. Control*, vol. 62, no. 10, pp. 4980–4993, Oct. 2017.
[11] J. Xu, S. Zhu, Y. C. Soh, and L. Xie, "Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes," in *Proc. 54th IEEE Conf. Decis. Control*, 2015, pp. 2055–2060.
[12] X. Ma, P. Yi, and J. Chen, "Distributed gradient tracking methods with finite data rates," *J. Syst. Sci. Complexity*, vol. 34, pp. 1927–1952, 2021.
[13] S. Pu and A. Nedić, "Distributed stochastic gradient tracking methods," *Math. Program.*, vol. 187, pp. 409–457, 2021.
[14] J. Wang, K. Fu, Y. Gu, and T. Li, "Convergence of distributed gradient-tracking-based optimization algorithms with random graphs," *J. Syst. Sci. Complexity*, vol. 34, pp. 1438–1453, 2021.
[15] H. Li, Q. Lü, and T. Huang, "Convergence analysis of a distributed optimization algorithm with a general unbalanced directed communication network," *IEEE Trans. Netw. Sci. Eng.*, vol. 6, no. 3, pp. 237–248, Jul.–Sep. 2019.
[16] J. Lei, H.-F. Chen, and H.-T. Fang, "Primal–dual algorithm for distributed constrained optimization," *Syst. Control Lett.*, vol. 96, pp. 110–117, 2016.
[17] S. Liu, Z. Qiu, and L. Xie, "Convergence rate analysis of distributed optimization with projected subgradient algorithm," *Automatica*, vol. 83, pp. 162–169, 2017.
[18] H. Li, Q. Lü, and T. Huang, "Distributed projection subgradient algorithm over time-varying general unbalanced directed graphs," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1309–1316, Mar. 2019.
[19] D. Yuan, S. Xu, H. Zhao, and L. Rong, "Distributed dual averaging method for multi-agent optimization with quantized communication," *Syst. Control Lett.*, vol. 61, no. 11, pp. 1053–1061, 2012.
[20] S. Liang, L. Wang, and G. Yin, "Dual averaging push for distributed convex optimization over time-varying directed graph," *IEEE Trans. Autom. Control*, vol. 65, no. 4, pp. 1785–1791, Apr. 2020.
[21] C. Wang, S. Xu, D. Yuan, B. Zhang, and Z. Zhang, "Push-sum distributed dual averaging for convex optimization in multiagent systems with communication delays," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 53, no. 3, pp. 1420–1430, Mar. 2023.
[22] J. Hou, X. Zeng, G. Wang, J. Sun, and J. Chen, "Distributed momentum-based Frank-Wolfe algorithm for stochastic optimization," *IEEE/CAA J. Automatica Sinica*, vol. 10, no. 3, pp. 685–699, Mar. 2023.
[23] G. Chen, P. Yi, Y. Hong, and J. Chen, "Distributed optimization with projection-free dynamics: A Frank-Wolfe perspective," *IEEE Trans. Cybern.*, vol. 54, no. 1, pp. 599–610, Jan. 2024.
[24] D. Yuan, Y. Hong, D. W. Ho, and G. Jiang, "Optimal distributed stochastic mirror descent for strongly convex optimization," *Automatica*, vol. 90, pp. 196–203, 2018.
[25] Y. Pang and G. Hu, "Randomized gradient-free distributed optimization methods for a multiagent system with unknown cost function," *IEEE Trans. Autom. Control*, vol. 65, no. 1, pp. 333–340, Jan. 2020.
[26] S. Cheng, S. Liang, Y. Fan, and Y. Hong, "Distributed gradient tracking for unbalanced optimization with different constraint sets," *IEEE Trans. Autom. Control*, vol. 68, no. 6, pp. 3633–3640, Jun. 2023.
[27] L. Romao, K. Margellos, G. Notarstefano, and A. Papachristodoulou, "Subgradient averaging for multi-agent optimisation with different constraint sets," *Automatica*, vol. 131, 2021, Art. no. 109738.
[28] Y. Wang, W. Zhao, Y. Hong, and M. Zamani, "Distributed subgradient-free stochastic optimization algorithm for nonsmooth convex functions over time-varying networks," *SIAM J. Control Optim.*, vol. 57, no. 4, pp. 2821–2842, 2019.
[29] X. Wu and J. Lu, "Fenchel dual gradient methods for distributed convex optimization over time-varying networks," *IEEE Trans. Autom. Control*, vol. 64, no. 11, pp. 4629–4636, Nov. 2019.
[30] C. Xi and U. A. Khan, "Distributed subgradient projection algorithm over directed graphs," *IEEE Trans. Autom. Control*, vol. 62, no. 8, pp. 3986–3992, Aug. 2017.
[31] S. Liang, L. Wang, and G. Yin, "Distributed quasi-monotone subgradient algorithm for nonsmooth convex optimization over directed graphs," *Automatica*, vol. 101, pp. 175–181, 2019.
[32] W. Yu, H. Liu, W. X. Zheng, and Y. Zhu, "Distributed discrete-time convex optimization with nonidentical local constraints over time-varying unbalanced directed graphs," *Automatica*, vol. 134, 2021, Art. no. 109899.
[33] D. Wang, J. Liu, J. Lian, Y. Liu, Z. Wang, and W. Wang, "Distributed delayed dual averaging for distributed optimization over time-varying digraphs," *Automatica*, vol. 150, 2023, Art. no. 110869.
[34] S. Pu, W. Shi, J. Xu, and A. Nedić, "Push- -pull gradient methods for distributed optimization in networks," *IEEE Trans. Autom. Control*, vol. 66, no. 1, pp. 1–16, Jan. 2021.
[35] Y. Liao, Z. Li, K. Huang, and S. Pu, "A compressed gradient tracking method for decentralized optimization with linear convergence," *IEEE Trans. Autom. Control*, vol. 67, no. 10, pp. 5622–5629, Oct. 2022.
[36] V. S. Mai and E. H. Abed, "Distributed optimization over directed graphs with row stochasticity and constraint regularity," *Automatica*, vol. 102, pp. 94–104, 2019.
[37] C. Xi, V. S. Mai, R. Xin, E. H. Abed, and U. A. Khan, "Linear convergence in optimization over directed graphs with row-stochastic matrices," *IEEE Trans. Autom. Control*, vol. 63, no. 10, pp. 3558–3565, Oct. 2018.
[38] A. Nedić, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.

Qing Huang received the B.Eng. degree in building electricity and intelligence from Anhui Science and Technology University, Bengbu, China, in 2022. She is currently working toward the M.Eng. degree with the School of Electrical Engineering and Automation, Anhui University, Hefei, China. Her research interests mainly include distributed control and optimization.

Yuan Fan (Senior Member, IEEE) received the B.Eng. degree in automation from the University of Science and Technology of China, Hefei, China, in 2006, and the Ph.D. degree in control science and engineering from the University of Science and Technology of China and the City University of Hong Kong, Kowloon, Hong Kong, in 2011. In 2013 and 2018, he was a Senior Research Associate with the Department of Mechanical and Biomedical Engineering, City University of Hong Kong, Hong Kong, and was a Research Fellow with the Department of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2015. Since 2011, he has been with Anhui University, Hefei, where he is currently a Professor. His research interests mainly include distributed control, microgrid control, and robot control.

Songsong Cheng (Member, IEEE) received the Ph.D. degree in engineering from the University of Science and Technology of China, Hefei, China, in 2018. From 2018 to 2020, he was a Postdoctoral Fellow with the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China. He was a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore in 2022. Since 2020, he has been with the School of Electrical Engineering and Automation, Anhui University, Hefei, China, where he is currently an Associated Professor. His research interests include distributed computation, optimization, and games.