SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

June 2026, Vol. 69, Iss. 6, 162206:1–162206:14
https://doi.org/10.1007/s11432-025-4887-6

Distributed constrained optimization algorithm for achieving linear convergence with less information storage

Hongzhe LIU[1], Mingqin CHENG[1], Wenwu YU[1,3*] & Wei Xing ZHENG[2*]

[1]*School of Mathematics, Southeast University, Nanjing 210000, China*
[2]*School of Computer, Data and Mathematical Sciences, Western Sydney University, Sydney 2751, Australia*
[3]*Purple Mountain Laboratories, Nanjing 211102, China*

Received 31 May 2025/Revised 13 October 2025/Accepted 24 November 2025/Published online 26 May 2026

Abstract This paper studies the optimization problem where the decision variable is contained in a closed convex set. For the researched constrained optimization problem, the aim of this work is to design a distributed algorithm that exhibits a linear convergence rate and requires less information storage for the iteration variables. Towards this end, based on the implicit gradient-tracking (IGT) technique, an auxiliary variable is introduced for each agent in this work, for which the iteration at the current step does not require the information of the state variable at the previous step. But such information is necessarily involved in the previous studies where the explicit gradient-tracking technique was implemented. Thus, less information storage is needed in the proposed distributed optimization algorithm with IGT employed. Moreover, in order to handle the closed convex set constraint with IGT employed, the indirect projection method is used in this work. Consequently, the distributed constrained optimization algorithm with IGT is successfully designed over an undirected graph. Additionally, the linear convergence rate is strictly proven under several common assumptions and the exact regions of the feasible constant step-sizes are also provided. Finally, a numerical simulation on the logistic regression problem is conducted to verify the effectiveness of the established theoretical results.

Keywords constrained optimization, distributed algorithm, linear convergence, less information storage

Citation Liu H Z, Cheng M Q, Yu W W, et al. Distributed constrained optimization algorithm for achieving linear convergence with less information storage. Sci China Inf Sci, 2026, 69(6): 162206, https://doi.org/10.1007/s11432-025-4887-6

1 Introduction

During the past decade, distributed optimization has been extensively investigated as a significant research topic in the field of systems science due to its wide applications in efficiently addressing various optimization problems involved in intelligent communication systems, smart grid systems, intelligent transportation systems, and other intelligent systems [1]. Accordingly, numerous important results regarding distributed optimization have been established (see, e.g., [2, 3] and references therein). Furthermore, in the existing results, the available distributed optimization algorithms mainly include two classes: continuous-time algorithms under the differential equation framework [4–17] and discrete-time algorithms under the difference equation framework [18–39]. In this work, we focus on designing a distributed algorithm under the discrete-time framework to address the constrained optimization problem. Thus, the subsequent literature overview is made only regarding distributed algorithms developed under the discrete-time framework.

The early work on distributed optimization was reported in [18], where the distributed algorithm with a non-negative decaying step-size was designed for the unconstrained optimization problem, and it was proven that the designed distributed algorithm achieves sub-linear convergence. Then, this result was extended to the cases with a global closed convex set constraint in [19] and with multiple global constraints including inequality constraint, linear equality constraint and closed convex set constraint in [20]. It is worth mentioning that only balanced graphs were involved in the above-mentioned studies.

Later on, distributed algorithms have been developed for optimization problems over unbalanced graphs. Over fixed unbalanced graphs, the distributed algorithm was designed for the optimization problem subject to nonidentical set constraints in [21] and further for the optimization problem subject to nonidentical inequality constraints, equality constraints, and closed convex set constraints in [22]. Moreover, over time-varying unbalanced graph sequences, a push-sum approach addressing unconstrained optimization was proposed in [23], and then this approach

---

* Corresponding author (email: wwyu@seu.edu.cn, w.zheng@westernsydney.edu.au)

was employed to solve the optimization problems with set constraints in [24, 25]. Specifically, an improved push-pull algorithm handling nonidentical multiple constraints including inequality constraints and set constraints was developed in [26]. Noteworthily, only sub-linear convergence could be achieved by those distributed optimization algorithms in the aforementioned studies [21–26].

Recently, various attempts have been made to design accelerated distributed optimization algorithms for different kinds of problems. In [27, 28] where the unconstrained optimization problems were studied, the distributed algorithms respectively with the Nesterov method and with the method of alternating direction method of multipliers (ADMM) were designed and they could achieve better sub-linear convergence under several assumptions on objective functions. Furthermore, in [29], the explicit gradient-tracking (EGT) technique was used for proposing the linearly convergent distributed algorithm over balanced graphs for the unconstrained optimization problem. Besides, the distributed optimization algorithms for the same problem setting as that in [29] were also developed over fixed unbalanced graphs in [30, 31] and over time-varying unbalanced graph sequences in [32, 33]. Additionally, for further accelerating the distributed algorithms with the EGT technique, the momentum method was introduced in [34–36]. Then, when the decision variables of the considered optimization problems are contained in a closed convex set, the distributed algorithms were designed respectively over fixed unbalanced graphs in [37] and over a time-varying unbalanced graph sequence in [38].

It is noteworthy that the EGT technique was employed in [29–38], and thus the iteration at step $l+1$ for the auxiliary variable required the information of the state variable at the previous step l, but the state variable at step l had already been updated. Therefore, more storage space is necessary for the operation of those distributed algorithms with the EGT technique. Recently, the distributed optimization algorithm with the implicit gradient-tracking (IGT) technique was successfully designed in [39], where only the updated information of the state variable was needed for the iteration at step $l+1$ for the auxiliary variable. However, only the unconstrained optimization problem was considered in [39]. Therefore, it is still challenging to design a distributed optimization algorithm with the IGT technique when constraints are involved.

The main contribution of this work is the successful design of the distributed constrained optimization algorithm with IGT (DCOAIGT) which can address the optimization problem with the set constraint. Moreover, for the designed DCOAIGT, the linear convergence rate is strictly proven, with the detailed feasible step-size regions being given. Specifically, the detailed contributions are given in the following.

(1) Given the difficulty in directly integrating the classical projection method into the distributed algorithm with the IGT technique, the indirect projection method is used to handle the constraint considered in this work. Moreover, the details on the above-mentioned difficulty are discussed after the development of the proposed algorithm.

(2) Unlike the work in [39] where the feasible step-size regions were given only by an implicit optimal solution set of an optimization problem, the feasible step-size regions are provided with a detailed form in this work.

The remainder of this paper is structured as follows. Section 2 presents the necessary background, including notations, relevant graph theory, and essential matrix theory. Section 3 then formulates the problem, proposes a distributed algorithm, states the main theoretical result, and provides a rigorous convergence analysis. Section 4 offers simulation results to validate the theoretical findings. Finally, Section 5 summarizes the contributions of this work, and Appendix A contains the proofs of several supporting lemmas and propositions.

2 Preliminaries

Some preliminaries including notations, graph theory and matrix theory are shown in this section.

2.1 Notations

The set of n-dimensional real vectors is denoted by $\mathbb{R}^n$ and the vector with all entries being 1 and proper dimension is denoted by $\mathbf{1}$. For a vector $\mathbf{s} \in \mathbb{R}^n$, $\|\mathbf{s}\|$ and $\mathbf{s}^{\mathrm{T}}$ respectively denote the 2-norm and the transpose of $\mathbf{s}$. Moreover, for a function $g(\mathbf{s})$ defined on $\mathbb{R}^n$, the gradient of g at $\mathbf{s}$ is represented by $\nabla g(\mathbf{s})$. The $(n \times n)$-dimensional real matrices set is denoted by $\mathbb{R}^{n\times n}$ and the identity matrix with proper dimension is denoted by $\mathbf{I}$. For a matrix $\mathbf{W} \in \mathbb{R}^{n\times n}$, the matrix norm induced from the 2-norm and the transpose of $\mathbf{W}$ are denoted by $\|\mathbf{W}\|$ and $\mathbf{W}^{\mathrm{T}}$, respectively. Furthermore, if all entries of $\mathbf{W}$ are nonnegative, then $\mathbf{W}$ is called a nonnegative matrix. Additionally, let $S_0 \subseteq \mathbb{R}^n$ be a closed convex set and $\mathbf{x} \in \mathbb{R}^n$ be a vector, then $P_{S_0}(\mathbf{x})$ denotes the projection of $\mathbf{x}$ on S_0. As discussed in [19], the projection operator has the non-expansive property, that is, $\|P_{S_0}(\mathbf{x}) - P_{S_0}(\mathbf{y})\| \leqslant \|\mathbf{x} - \mathbf{y}\|$, $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

2.2 Graph theory

An undirected graph $\mathcal{G}$ is depicted by a node set $\mathcal{M} = \{1, 2, \ldots, M\}$ and an edge set $\mathcal{E} \subseteq \mathcal{M} \times \mathcal{M}$ with the associated weighted matrix $\mathbf{W} = [w_{ji}] \in \mathbb{R}^{M \times M}$. If an edge exists between nodes j and i, meaning that node j can receive information from node i directly, then $(j, i) \in \mathcal{E}$ and $w_{ji} > 0$. On the other hand, if node j cannot receive information from node i directly, then $(j, i) \notin \mathcal{E}$ and $w_{ji} = 0$. Specifically, assume that $(j, j) \in \mathcal{E}$ for all $j = 1, 2, \ldots, M$. Moreover, a path is said to exist between two distinct nodes j and j' if there exist the edges $(j, j_1), (j_1, j_2), \ldots, (j_{k-1}, j_k), (j_k, j')$ with $j, j_1, \ldots, j_k, j'$ being distinct nodes. Besides, graph $\mathcal{G}$ is said to be connected if a path exists between any two distinct nodes j and j'.

2.3 Matrix theory

In the subsequent analysis, the nonnegative double-stochastic matrix is involved and its property is significant. Thus, a necessary description of nonnegative double-stochasticity is introduced in the following.

Definition 1. Suppose that $\mathbf{W} \in \mathbb{R}^{M \times M}$ is a nonnegative matrix. $\mathbf{W}$ is said to be a nonnegative doubly-stochastic matrix if $\mathbf{W}\mathbf{1} = \mathbf{1}^{\mathrm{T}}\mathbf{W} = 1$.

Lemma 1 ([37])**.** Suppose that $\mathcal{G}$ is a connected graph and $\mathbf{W} \in \mathbb{R}^{M \times M}$ is the associated nonnegative doubly-stochastic weighted matrix. Then there holds

$$\left\| \mathbf{W} - \frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}} \right\| \leqslant \rho_0 < 1.$$

3 Main results

In this section, the studied problem formulation, the proposed distributed algorithm, the established main theorem, and detailed convergence analysis are given.

3.1 Problem formulation

In this work, we study the optimization problem where the decision variable is contained in a global closed convex set. Specifically, the problem is formulated as

$$\begin{aligned} &\min_{\mathbf{s}} g(\mathbf{s}) = \frac{1}{M}\sum_{j=1}^{M} g_j(\mathbf{s}) \\ &\text{s.t.} \quad \mathbf{s} \in S_0, \end{aligned} \tag{1}$$

where $\mathbf{s} \in \mathbb{R}^n$, $S_0 \subseteq \mathbb{R}^n$ is a closed convex set, and $g_j(\mathbf{s})$ is a convex function on $\mathbb{R}^n$. Specifically, in this paper, $n = 1$ is assumed as a matter of illustration convenience and the case when $n > 1$ is similar to that discussed in [37]. First, several key assumptions are introduced for problem (1) in the following.

Assumption 1. For problem (1), all $g_j(\mathbf{s})$, $j = 1, 2, \ldots, M$, are μ-strongly convex and L-smooth on $\mathbb{R}^n$ with μ and L being two positive constants, i.e., all $g_j(\mathbf{s})$ are differentiable and $\forall \mathbf{s}, \mathbf{s}' \in \mathbb{R}^n$,

$$\begin{aligned} &(\mathbf{s} - \mathbf{s}')^{\mathrm{T}}(\nabla g_j(\mathbf{s}) - \nabla g_j(\mathbf{s}')) \geqslant \mu\|\mathbf{s} - \mathbf{s}'\|^2, \\ &\|\nabla g_j(\mathbf{s}) - \nabla g_j(\mathbf{s}')\| \leqslant L\|\mathbf{s} - \mathbf{s}'\|. \end{aligned}$$

Clearly, if Assumption 1 holds, then the optimal solution to problem (1) exists and is unique, which is denoted as $\mathbf{s}^*$. Furthermore, as pointed out in [37], if $\mathbf{s}^*$ is the optimal solution to problem (1), then for an arbitrary positive constant α_0, there holds

$$\mathbf{s}^* = P_{S_0}(\mathbf{s}^* - \alpha_0 \nabla g(\mathbf{s}^*)). \tag{2}$$

Moreover, it can be obtained from (2) that for an arbitrary positive constant α_0 and a vector $\mathbf{u} \in \mathbb{R}^n$, if

$$\mathbf{u} = P_{S_0}(\mathbf{u}) - \alpha_0 \nabla g(P_{S_0}(\mathbf{u})), \tag{3}$$

then $P_{S_0}(\mathbf{u})$ is the optimal solution to problem (1) since it satisfies (2). Clearly, for arbitrary $\mathbf{u}, \mathbf{u}' \in \mathbb{R}^n$ satisfying (3), $P_{S_0}(\mathbf{u})$ and $P_{S_0}(\mathbf{u}')$ are the optimal solutions to problem (1). Then, under Assumption 1, it can be obtained that the optimal solution to problem (1) is unique and thus

$$P_{S_0}(\mathbf{u}) = P_{S_0}(\mathbf{u}').$$

Therefore, it holds

$$\mathbf{u} = P_{S_0}(\mathbf{u}) - \alpha_0 \nabla g(P_{S_0}(\mathbf{u})) = P_{S_0}(\mathbf{u}') - \alpha_0 \nabla g(P_{S_0}(\mathbf{u}')) = \mathbf{u}',$$

which means that the vector satisfying (3) is also unique under Assumption 1 and is denoted as $\mathbf{u}^*$.

3.2 Distributed algorithm and main theorem

It is worth mentioning that if $\mathbf{u}^*$ can be exactly obtained, then the optimal solution to problem (1) $\mathbf{s}^* = P_{S_0}(\mathbf{u}^*)$ can be easily acquired. Therefore, in this subsection, considering the difficulties on designing and analyzing a distributed optimization algorithm for converging to $\mathbf{s}^*$ directly, the following distributed optimization algorithm for converging to $\mathbf{u}^*$ rather than $\mathbf{s}^*$ is designed for problem (1) (see Algorithm 1).

Algorithm 1 DCOAIGT.

I: **Input**: The positive step-sizes α and β, the positive integer M, the associated weighted matrix $W = [w_{ji}] \in \mathbb{R}^{M\times M}$, and the number of iterations T;

II: **Initialize**: $\forall j = 1, 2, \ldots, M$, arbitrary $u_j(0) \in \mathbb{R}$, and $v_j(0) = 0$.

III: **Iterative rule**:

$$u_j(l+1) = s_j(l) - \alpha \nabla g_j(s_j(l)) - v_j(l), \tag{4a}$$

$$v_j(l+1) = v_j(l) + \beta\Big(u_j(l+1) - \sum_{i=1}^{M} w_{ji} u_i(l+1)\Big). \tag{4b}$$

IV: **Output:** $\forall j = 1, 2, \ldots, M$, $u_j(T)$, $v_j(T)$ and $s_j(T)$.

First, we provide some insights into the development of Algorithm 1. Clearly, in Algorithm 1, iterations (4a) and (4b) are designed for ensuring that $\mathbf{u}_j(l)$ converges to $\mathbf{u}^*$ and the consensus of all $\mathbf{u}_j(l)$, respectively. In [29–38], the EGT-based distributed algorithms for various optimization problems were proposed and the linear convergence rates were achieved. Specifically, the iterative rule of the EGT-based distributed optimization algorithm for the unconstrained optimization problem was proposed as

$$\mathbf{u}_j(l+1) = \sum_{i=1}^{M} w_{ji} \mathbf{u}_j(l) - \alpha \mathbf{v}_j(l), \tag{5a}$$

$$\mathbf{v}_j(l+1) = \sum_{i=1}^{M} w_{ji} \mathbf{v}_j(l) + \nabla g_j(\mathbf{u}_j(l+1)) - \nabla g_j(\mathbf{u}_j(l)). \tag{5b}$$

It is important to notice that the information of $\mathbf{u}_j(l)$ is still necessary for the iteration of $\mathbf{v}_j(l)$ after $\mathbf{u}_j(l)$ has already been updated into $\mathbf{u}_j(l+1)$, which implies that extra storage space is needed for the information of $\mathbf{u}_j(l)$ during the operation of iteration (5). Thus, in this paper, in order to save the storage space, only the information of the updated state $\mathbf{u}_j(l+1)$ is involved in the iteration of $\mathbf{v}_j(l)$. Furthermore, noting from (14) to be given in Lemma 5 that the gradient-tracking term $\nabla \mathbf{G}(l) - \nabla \mathbf{G}(l-1)$ also implicitly exists in Algorithm 1, we thus call Algorithm 1 as the distributed constrained optimization algorithm with implicit gradient-tracking method (DCOAIGT).

Moreover, as discussed in [29,37], $\mathbf{v}_j(l)$ in (5) is employed to approximately track the information of the gradient of the global objective function, i.e., $\frac{1}{M}\sum_{j=1}^{M} g_j(\mathbf{u}_j(l))$. Clearly, it is necessary to use the information of the gradient of the global objective function in the derived iteration of $\bar{\mathbf{u}}(l) = \frac{1}{M}\sum_{j=1}^{M} \mathbf{u}_j(l)$ since the aim is to solve problem (1) which has a global objective function with the form of $\frac{1}{M}\sum_{j=1}^{M} g_j(\mathbf{s})$. Additionally, considering that the classical projection method is usually employed to address the closed convex set constraint, DCOAIGT can be rewritten as

$$\mathbf{u}_j(l+1) = P_{S_0}(\mathbf{u}_j(l) - \alpha \nabla g_j(\mathbf{u}_j(l)) - \mathbf{v}_j(l)), \tag{6a}$$

$$\mathbf{v}_j(l+1) = \mathbf{v}_j(l) + \beta\Big(\mathbf{u}_j(l+1) - \sum_{i=1}^{M} w_{ji} \mathbf{u}_i(l+1)\Big). \tag{6b}$$

Furthermore, since the projection operators on most closed convex sets are nonlinear, the derived iteration

$$\bar{\mathbf{u}}(l+1) = \frac{1}{M}\sum_{i=1}^{M} P_{S_0}(\mathbf{u}_j(l) - \alpha\nabla g_j(\mathbf{u}_j(l)) - \mathbf{v}_j(l)),$$

cannot involve the term $\frac{1}{M}\sum_{j=1}^{M} g_j(\mathbf{u}_i(l))$. Therefore, it is infeasible for DCOAIGT to use the classical projection method to address the involved set constraint.

On the other hand, with the indirect projection method as employed in this paper, the derived iteration

$$\bar{\mathbf{u}}(l+1) = \frac{1}{M}\sum_{j=1}^{M} \mathbf{s}_j(l) - \alpha\frac{1}{M}\sum_{j=1}^{M} \nabla g_j(\mathbf{u}_j(l)) - \frac{1}{M}\sum_{j=1}^{M} \mathbf{v}_j(l),$$

could involve the information of the gradient of the global objective function and thus the indirect projection method is feasible for DCOAIGT developed in this paper.

Introducing the following variables:

$$\begin{aligned}
&\mathbf{u}(l) = (\mathbf{u}_1(l), \mathbf{u}_2(l), \ldots, \mathbf{u}_M(l))^{\mathrm{T}},\\
&\mathbf{v}(l) = (\mathbf{v}_1(l), \mathbf{v}_2(l), \ldots, \mathbf{v}_M(l))^{\mathrm{T}},\\
&\mathbf{s}(l) = (\mathbf{s}_1(l), \mathbf{s}_2(l), \ldots, \mathbf{s}_M(l))^{\mathrm{T}},\\
&\nabla\mathbf{G}(l) = (\nabla g_1(\mathbf{s}_1(l)), \ldots, \nabla g_M(\mathbf{s}_M(l)))^{\mathrm{T}},\\
&\nabla\bar{\mathbf{G}}(l) = (\nabla g_1(P_{S_0}(\bar{\mathbf{u}}(l))), \ldots, \nabla g_M(P_{S_0}(\bar{\mathbf{u}}(l))))^{\mathrm{T}},
\end{aligned}$$

and $S = S_0^M$, we can rewrite (4) into the compact form as

$$\mathbf{u}(l+1) = \mathbf{s}(l) - \alpha\nabla\mathbf{G}(l) - \mathbf{v}(l), \tag{7a}$$

$$\mathbf{v}(l+1) = \mathbf{v}(l) + \beta(\mathbf{I} - \mathbf{W})\mathbf{u}(l+1). \tag{7b}$$

Remark 1. In real-world applications, the information shared over the channels/links in graph $\mathcal{G}$ might be subject to quantization and thus it is significant to consider the quantization of the exchanged information when designing distributed optimization algorithms. Furthermore, based on [40, 41] where the distributed optimization algorithms with information quantization were designed, the result in this work could also be possibly extended to the cases with information quantization considered although the introduction of information quantization into the distributed algorithm may cause potential difficulty in the convergence analysis.

Next, one significant assumption on graph $\mathcal{G}$ is made for the subsequent analysis, based on which the main theorem describing the convergence property of $\mathbf{s}_j(l)$ can be established.

Assumption 2. Graph $\mathcal{G}$ is connected and its weighted matrix $\mathbf{W}$ is a nonnegative doubly-stochastic matrix.

It is worth mentioning that it is not necessary for the design of the weighted matrix $\mathbf{W}$ to obey the specific rule, and actually arbitrary nonnegative doubly-stochastic matrices associated with the given graph $\mathcal{G}$ could be selected for the proposed Algorithm 1.

Assumption 3. For the positive constants μ and L as introduced in Assumption 1, there holds

$$2L < (3 - \rho_0)\mu, \tag{8}$$

with $\rho_0 = \|\mathbf{W} - \frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\|$.

Theorem 1. Suppose that Assumptions 1–3 hold. If the step-sizes α and β satisfy

$$\frac{2}{(3-\rho_0)\mu} < \alpha < \frac{1}{L},\ 0 < \beta \leqslant 1, \tag{9}$$

then all $\mathbf{s}_j(l)$, $j = 1, 2, \cdots, M$, generated by DCOAIGT, converge to $\mathbf{s}^*$ with a linear rate.

Proof. The detailed proof is provided in the following Subsection 3.3.

Remark 2. It can be clearly seen from Theorem 1 that the inequality (8) in Assumption 3 is the key point to ensure that the feasible step-size region of α is nonempty. Moreover, it should be noted that the inequality (8) holds for some common strongly convex functions, such as the quadratic function on $\mathbb{R}$, since $2 < 3 - \rho_0$ with $\rho_0 < 1$ and $L = \mu$ for the quadratic function on $\mathbb{R}$.

3.3 Convergence analysis

In this subsection, Theorem 1 is proven by resorting to the linear matrix inequality theory. To this end, a linear matrix inequality is first established in the next lemma.

Lemma 2. Suppose that Assumptions 1–3 hold. If the step-sizes α and β satisfy (9), then for all $l \geqslant 1$, the variables $\mathbf{u}(l)$ and $\mathbf{v}(l)$ under the iterative rule (7) satisfy

$$\begin{pmatrix} \|\mathbf{u}(l+1)-\mathbf{1}\bar{\mathbf{u}}(l+1)\| \\ \sqrt{M}\|\bar{\mathbf{u}}(l+1)-\mathbf{u}^*\| \\ \|\mathbf{u}(l+1)-\mathbf{u}(l)\| \end{pmatrix} \leqslant \mathbf{A} \begin{pmatrix} \|\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\| \\ \sqrt{M}\|\bar{\mathbf{u}}(l)-\mathbf{u}^*\| \\ \|\mathbf{u}(l)-\mathbf{u}(l-1)\| \end{pmatrix}, \tag{10}$$

where

$$\mathbf{A} = \begin{pmatrix} \rho_1 & 0 & \gamma \\ \gamma & \gamma & 0 \\ 2\beta & 0 & \gamma \end{pmatrix},$$

with $\rho_1 = \|(1-\beta)\mathbf{I} + \beta\mathbf{W} - \frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\|$ and $\gamma = 1-\alpha\mu$. Furthermore, $\rho_1 \leqslant 1-\beta(1-\rho_0) < 1$.

Proof. The detailed proof is provided in Appendix A.

Based on Lemma 2, we are prepared to give the detailed proof of Theorem 1.

Proof of Theorem 1. Clearly, in order to complete the proof, it is sufficient to verify that the spectral radius $\rho(\mathbf{A})$ of $\mathbf{A}$ meets $\rho(\mathbf{A}) < 1$ if the step-sizes α and β satisfy (9). Furthermore, noting that $\mathbf{A}$ is nonnegative, we can obtain from [42, Theorem 8.3.1] that $\rho(\mathbf{A})$ is a real eigenvalue of $\mathbf{A}$. In the following, we will confirm that all real eigenvalues of $\mathbf{A}$ are smaller than 1 so as to show $\rho(\mathbf{A}) < 1$, thus completing the proof of Theorem 1.

By direct computation, we have

$$|\lambda\mathbf{I} - \mathbf{A}| = (\lambda-\gamma)[\lambda^2 - (\rho_1+\gamma)\lambda + (\rho_1\gamma - 2\beta\gamma)]. \tag{11}$$

Introduce the quadratic polynomial

$$p_0(\lambda) = \lambda^2 - (\rho_1+\gamma)\lambda + (\rho_1\gamma - 2\beta\gamma).$$

Since $\gamma = 1-\alpha\mu < 1$ is a real eigenvalue of $\mathbf{A}$ based on (11), it suffices to confirm that all real solutions to $p_0(\lambda) = 0$ are less than 1 or the nonexistence of those real solutions for completing the proof.

For the quadratic polynomial $p_0(\lambda)$ with respect to λ, since the coefficient of the quadratic term is $1 > 0$ and the axis of symmetry of the curve of $p_0(\lambda)$ denoted as $\lambda = \frac{\rho_1+\gamma}{2}$ is less than 1, $p_0(1) > 0$ with the selected step-sizes α and β satisfying (9) can deduce the conclusion that all real solutions to $p_0(\lambda) = 0$ are less than 1 or the nonexistence of those real solutions. Directly, with the selected step-sizes α and β satisfying (9), we have

$$\begin{aligned} p_0(1) &= 1 - (\rho_1+\gamma) + (\rho_1\gamma - 2\beta\gamma) \\ &= 1 - (1-\gamma)\rho_1 - \gamma - 2\beta\gamma \\ &= 1 - \alpha\mu\rho_1 - (1-\alpha\mu) - 2\beta(1-\alpha\mu) \\ &= -\alpha\mu\rho_1 + \alpha\mu - 2\beta + 2\alpha\beta\mu. \end{aligned}$$

Then, considering that $\rho_1 \leqslant 1-\beta(1-\rho_0)$ as given in Lemma 2, we have

$$\begin{aligned} p_0(1) &= -\alpha\mu\rho_1 + \alpha\mu - 2\beta + 2\alpha\beta\mu \\ &\geqslant -\alpha\mu[1-\beta(1-\rho_0)] + \alpha\mu - 2\beta + 2\alpha\beta\mu \\ &= \alpha\beta\mu(1-\rho_0) - 2\beta + 2\alpha\beta\mu \\ &= \beta[\alpha\mu(3-\rho_0) - 2]. \end{aligned}$$

Clearly, since $\alpha > \frac{2}{(3-\rho_0)\mu}$ as assumed in (9), it follows that

$$p_0(1) > \beta\Big[\frac{2}{(3-\rho_0)\mu}\mu(3-\rho_0) - 2\Big] = 0,$$

which completes the proof.

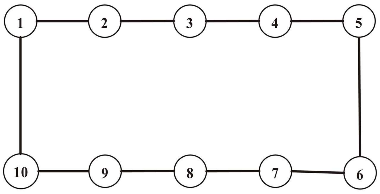

Figure 1 Graph $\mathcal{G}$ with ten interacting agents.

Remark 3. From the convergence analysis of the proposed Algorithm 1, it can be concluded that the computational complexity of Algorithm 1 is at most $\mathcal{O}(M^2n^2)$, thus being of polynomial order. In fact, a similar computational complexity with polynomial order can also be found in most studies on distributed optimization algorithms. Therefore, the distributed optimization algorithms could be efficiently applied in real-time large-scale systems, such as intelligent communication systems, smart grid systems, and intelligent transportation systems.

Remark 4. In order to focus on the main contribution that a distributed optimization algorithm with linear rate and less information storage can be designed to successfully address the constrained optimization problem, only the balanced and time-invariant network is considered in this work. On the other hand, it is worth mentioning that based on the method of estimating the left eigenvector with respect to eigenvalue 1 of a nonnegative stochastic matrix (see [21,22,31,37]) and Push-Pull or Push-DIGing framework (see [32,33,38]), the result in this work could be potentially extended to the cases with fixed unbalanced networks and uniformly-connected time-varying unbalanced networks, respectively.

Remark 5. It can be obtained from Lemma 2 that the value $\rho(\mathbf{A})$ affected by ρ_0 determines the convergence rate of the proposed distributed algorithm. Furthermore, it can also be noted that ρ_0 is the norm of matrix $\mathbf{W}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}$ and the weighted matrix $\mathbf{W}$ is associated with graph $\mathcal{G}$. Therefore, the structure of the graph $\mathcal{G}$ would have some effect on the convergence rate, although an arbitrary graph $\mathcal{G}$ satisfying Assumption 2 can ensure the linear convergence property of the proposed Algorithm 1 with feasible step-sizes. Specifically, researching how the structure of graph $\mathcal{G}$ affects the convergence rate based on [43] would be an interesting future topic.

4 Simulations

In this section, simulation results about applying the proposed distributed algorithm DCOAIGT to the logistic regression problem are presented to verify its good performance. Specifically, as discussed in [37], the logistic regression problem on $\mathbb{R}^2$ can be formulated as

$$\min_{\mathbf{s}\in S_0} g(\mathbf{s})=\sum_{j=1}^{M}\ln\left[1+\mathrm{e}^{-a_j\mathbf{b}_j^{\mathrm{T}}\mathbf{s}}\right]+\frac{c}{2}\|\mathbf{s}\|^2, \tag{12}$$

where $a_j\in\mathbb{R}$ and $\mathbf{b}_j\in\mathbb{R}^2$ respectively represent the corresponding labels and the feature vectors, and c is a positive constant. Specifically, in the simulation, a_j is selected as $(-1)^j$ and $\mathbf{b}_j$ is set as $(0.1j,0.2j)^{\mathrm{T}}$ for each $j=1,2,\ldots,M$. Moreover, $M=10$, $c=10$, and $S_0=[-2,2]$ are chosen.

Additionally, for the proposed distributed algorithm (4), the step-sizes α and β are set as 0.08 and 1, respectively. The undirected graph $\mathcal{G}$ with 10 nodes as depicted in Figure 1 is introduced for DCOAIGT, and correspondingly, its associated weighted matrix $\mathbf{W}$ can be set as

$$\begin{pmatrix}
\frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3}\\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\
\frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3}
\end{pmatrix}.$$

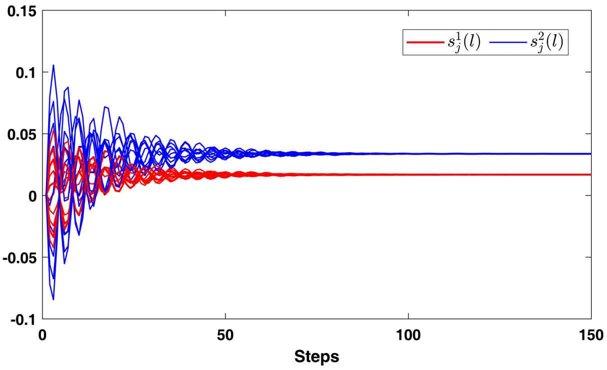

Figure 2 (Color online) Behaviors of the states $\mathbf{s}_j(l)$ under DCOAIGT.

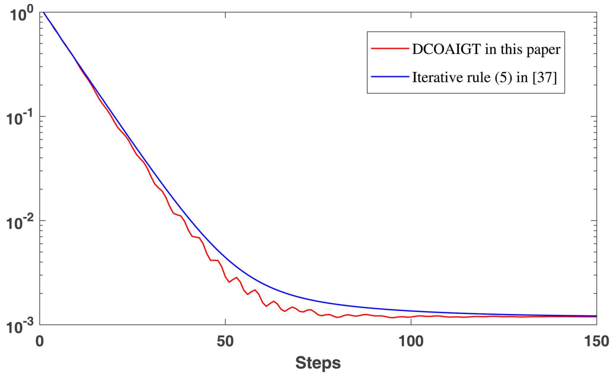

Figure 3 (Color online) Behaviors of the convergence index under DCOAIGT in this paper and under the iterative rule (5) in [37].

Then the transient behaviors of all $\mathbf{s}_j(l)$ under the proposed DCOAIGT are displayed in Figure 2, where all $s_j(l)$ converge to $\mathbf{s}^* = (0.0169, 0.0338)^{\mathrm{T}}$. Specifically, $s_j^1(l)$ and $s_j^2(l)$ involved in Figure 2 respectively represent the first entry and second entry of the vector $\mathbf{s}_j(l)$.

Noting that the same optimization problem as researched in this work was also solved linearly by designing the distributed algorithm with EGT in [37], we further make a simulation comparison between DCOAIGT and the algorithm (5) designed in [37]. To this end, the convergence index (CI)

$$\mathrm{CI} = \frac{\left\|\frac{1}{M}\sum_{j=1}^{M}\mathbf{s}_j(l) - \mathbf{s}^*\right\|}{\|\mathbf{s}^*\|}$$

is introduced to measure the convergence performance of these two algorithms. Then, the transient behaviors of CI under the proposed distributed algorithm DCOAIGT and the iterative rule (5) designed in [37] are exhibited in Figure 3, which indicates that the considered two algorithms have a similar linear convergence rate while DCOAIGT could save the storage space as discussed after the development of DCOAIGT.

On the other hand, it can be observed that some oscillations occur in the curve of the transient behavior of CI under DCOAIGT before the algorithm is absolutely convergent. Noting (14), the potential reason for the oscillations is that the term $\|(\mathbf{s}(l) - \mathbf{s}(l-1)) - \alpha(\nabla\mathbf{G}(l) - \nabla\mathbf{G}(l-1))\|$ may be large since the adjustable small coefficient is absent for the term $\mathbf{s}(l) - \mathbf{s}(l-1)$. This makes it difficult for DCOAIGT to achieve the desired consensus performance before it is absolutely convergent. In order to avoid such oscillations existing in the curve of the transient behaviors of CI under DCOAIGT, we improve the proposed distributed algorithm DCOAIGT by redefining $\mathbf{s}_j(l) = P_{S_0}\left(\sum_{i=1}^{M} w_{ji}\mathbf{u}_i(l)\right)$, which can benefit achieving consensus of all $\mathbf{u}_j(l)$ while increasing information exchange. Note that the convergence analysis of Algorithm 1 with the term $\mathbf{s}_j(l) = P_{S_0}(\sum_{i=1}^{M} w_{ji}\mathbf{u}_i(l))$ can be similarly completed. Furthermore, for comparison purposes, the transient behaviors of CI under DCOAIGT with $\mathbf{s}_j(l) = P_{S_0}(\mathbf{u}_i(l))$ and $\mathbf{s}_j(l) = P_{S_0}\left(\sum_{i=1}^{M} w_{ji}\mathbf{u}_i(l)\right)$, and the iterative rule (5) designed in [37] are shown in Figure 4, from which it can be seen that those oscillations disappear for DCOAIGT with $\mathbf{s}_j(l) = P_{S_0}\left(\sum_{i=1}^{M} w_{ji}\mathbf{u}_i(l)\right)$. Moreover, it can be seen from Figure 4 that the increased information exchange accompanied by the modified term could benefit from eliminating the oscillations present in the CI curve.

5 Conclusion

This paper has studied the optimization problem under a global closed convex set constraint. To address the studied problem, we have introduced a distributed algorithm DCOAIGT over an undirected graph, which curtails the information storage demands inherent in many EGT-based distributed optimization strategies. The linear convergence rate of DCOAIGT has been formally proven under standard assumptions, with feasible step-size regions being explicitly provided. Simulation studies on a logistic regression problem have empirically confirmed the effectiveness of our distributed approach, aligning with the theoretical analysis. Notably, compared with [39] which focused on an unconstrained problem, this study has developed a linearly convergent distributed algorithm for constrained optimization problem, thus making a significant contribution. Finally, limitations of the current algorithm include

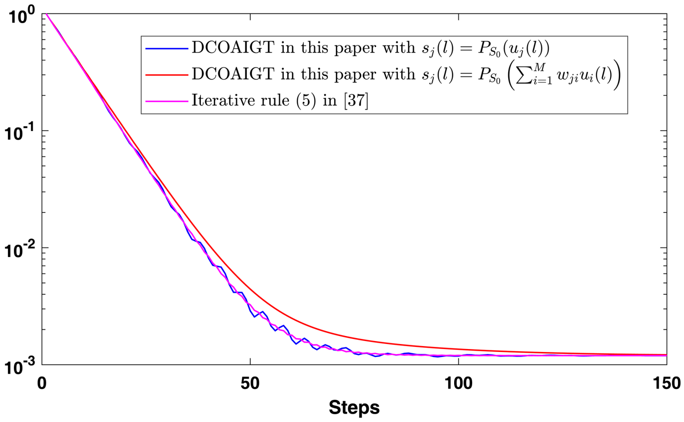

Figure 4 (Color online) Behaviors of the convergence index under DCOAIGT with $\mathbf{s}_j(l)=P_{S_0}(\mathbf{u}_j(l))$ and $\mathbf{s}_j(l)=P_{S_0}\left(\sum_{i=1}^{M}w_{ji}\mathbf{u}_i(l)\right)$, and the iterative rule (5) designed in [37].

its inability to address more complicated constraints or unbalanced graphs, and these challenges will be a focus of future research.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62573120, 62233004), Jiangsu Provincial Scientific Research Center of Applied Mathematics (Grant No. BK20233002), and Natural Science Foundation of Jiangsu Province (Grant No. BK20242027).

References

1 Yu W W, Chen D X, Liu H Z, et al. Systems science in the new era: intelligent systems and big data. Sci China Inf Sci, 2024, 67: 136201

2 Yi P, Hong Y. Distributed cooperative optimization and its applications (in Chinese). Sci Sin Math, 2016, 46: 1547–1564

3 Chai T Y, Yang T. Research status and prospects of distributed collaborative optimization (in Chinese). Sci Sin Tech, 2020, 50: 1414–1425

4 Wang J, Elia N. A control perspective for centralized and distributed convex optimization. In: Proceeding of 2011 50th IEEE Conference on Decision and Control and European Control Conference, 2011. 3800–3805

5 Gharesifard B, Cortes J. Distributed continuous-time convex optimization on weight-balanced digraphs. IEEE Trans Automat Contr, 2014, 59: 781–786

6 Hong H, Yu X, Yu W, et al. Distributed convex optimization on state-dependent undirected graphs: homogeneity technique. IEEE Trans Control Netw Syst, 2020, 7: 42–52

7 Zhu Y, Yu W, Wen G, et al. Continuous-time coordination algorithm for distributed convex optimization over weight-unbalanced directed networks. IEEE Trans Circuits Syst II, 2019, 66: 1202–1206

8 Kia S S, Cortés J, Martínez S. Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. Automatica, 2015, 55: 254–264

9 Liu Q, Wang J. A second-order multi-agent network for bound-constrained distributed optimization. IEEE Trans Automat Contr, 2015, 60: 3310–3315

10 Shi G, Johansson K H, Hong Y. Reaching an optimal consensus: dynamical systems that compute intersections of convex sets. IEEE Trans Automat Contr, 2013, 58: 610–622

11 Qiu Z, Liu S, Xie L. Distributed constrained optimal consensus of multi-agent systems. Automatica, 2016, 68: 209–215

12 Lin P, Ren W, Farrell J A. Distributed continuous-time optimization: nonuniform gradient gains, finite-time convergence, and convex constraint set. IEEE Trans Automat Contr, 2017, 62: 2239–2253

13 Yuan D, Ho D W C, Xu S. Regularized primal-dual subgradient method for distributed constrained optimization. IEEE Trans Cybern, 2016, 46: 2109–2118

14 Yi P, Hong Y, Liu F. Distributed gradient algorithm for constrained optimization with application to load sharing in power systems. Syst Control Lett, 2015, 83: 45–52

15 Liu Q, Yang S, Wang J. A collective neurodynamic approach to distributed constrained optimization. IEEE Trans Neural Netw Learn Syst, 2017, 28: 1747–1758

16 Yang S, Liu Q, Wang J. A multi-agent system with a proportional-integral protocol for distributed constrained optimization. IEEE Trans Automat Contr, 2017, 62: 3461–3467

17 Zhu Y, Yu W, Wen G, et al. Continuous-time distributed subgradient algorithm for convex optimization with general constraints. IEEE Trans Automat Contr, 2019, 64: 1694–1701

18 Nedic A, Ozdaglar A. Distributed subgradient methods for multi-agent optimization. IEEE Trans Automat Contr, 2009, 54: 48–61

19 Nedic A, Ozdaglar A, Parrilo P A. Constrained consensus and optimization in multi-agent networks. IEEE Trans Automat Contr, 2010, 55: 922–938

20 Zhu M, Martinez S. On distributed convex optimization under inequality and equality constraints. IEEE Trans Automat Contr, 2012, 57: 151–164

21 Mai V S, Abed E H. Distributed optimization over directed graphs with row stochasticity and constraint regularity. Automatica, 2019, 102: 94–104

22 Liu H, Zheng W X, Yu W. Distributed discrete-time algorithms for convex optimization with general local constraints on weight-unbalanced digraph. IEEE Trans Control Netw Syst, 2021, 8: 51–64

23 Nedic A, Olshevsky A. Distributed optimization over time-varying directed graphs. IEEE Trans Automat Contr, 2015, 60: 601–615

24 Liang S, Wang L Y, Yin G. Dual averaging push for distributed convex optimization over time-varying directed graph. IEEE Trans Automat Contr, 2020, 65: 1785–1791

25 Yu W, Liu H, Zheng W X, et al. Distributed discrete-time convex optimization with nonidentical local constraints over time-varying unbalanced directed graphs. Automatica, 2021, 134: 109899

26 Liu H, Yu W, Wen G, et al. Distributed algorithm over time-varying unbalanced graphs for optimization problem subject to multiple local constraints. IEEE Trans Control Netw Syst, 2025, 12: 387–402

27 Jakovetic D, Xavier J, Moura J M F. Fast distributed gradient methods. IEEE Trans Automat Contr, 2014, 59: 1131–1146

28 Aybat N S, Wang Z, Lin T, et al. Distributed linearized alternating direction method of multipliers for composite convex consensus optimization. IEEE Trans Automat Contr, 2018, 63: 5–20

29 Qu G, Li N. Harnessing smoothness to accelerate distributed optimization. IEEE Trans Control Netw Syst, 2018, 5: 1245–1260

30 Xi C, Xin R, Khan U A. ADD-OPT: accelerated distributed directed optimization. IEEE Trans Automat Contr, 2018, 63: 1329–1339

31 Xi C, Mai V S, Xin R, et al. Linear convergence in optimization over directed graphs with row-stochastic matrices. IEEE Trans Automat Contr, 2018, 63: 3558–3565

32 Nedić A, Olshevsky A, Shi W. Achieving geometric convergence for distributed optimization over time-varying graphs. SIAM J Optim, 2017, 27: 2597–2633

33 Saadatniaki F, Xin R, Khan U A. Decentralized optimization over time-varying directed graphs with row and column-stochastic matrices. IEEE Trans Automat Contr, 2020, 65: 4769–4780

34 Xin R, Khan U A. Distributed heavy-ball: a generalization and acceleration of first-order methods with gradient tracking. IEEE Trans Automat Contr, 2020, 65: 2627–2633

35 Li H, Cheng H, Wang Z, et al. Distributed nesterov gradient and heavy-ball double accelerated asynchronous optimization. IEEE Trans Neural Netw Learn Syst, 2021, 32: 5723–5737

36 Gao J, Liu X, Dai Y H, et al. A family of distributed momentum methods over directed graphs with linear convergence. IEEE Trans Automat Contr, 2023, 68: 1085–1092

37 Liu H, Yu W, Chen G. Discrete-time algorithms for distributed constrained convex optimization with linear convergence rates. IEEE Trans Cybern, 2022, 52: 4874–4885

38 Liu H, Yu W, Zheng W X, et al. Distributed constrained optimization algorithms with linear convergence rate over time-varying unbalanced graphs. Automatica, 2024, 159: 111346

39 Ghaderyan D, Aybat N S, Aguiar A P, et al. A fast row-stochastic decentralized method for distributed optimization over directed graphs. IEEE Trans Automat Contr, 2024, 69: 275–289

40 Doostmohammadian M, Qureshi M I, Hossein Khalesi M, et al. Notice of removal: log-scale quantization in distributed first-order methods: gradient-based learning from distributed data. IEEE Trans Automat Sci Eng, 2025, 22: 10948–10959

41 Doostmohammadian M, Rabiee H R. Momentum-based accelerated algorithm for distributed optimization under sector-bound nonlinearity. J Franklin Institute, 2025, 362: 107857

42 Horn R A, Johnson C R. Matrix Analysis. Cambridge: Cambridge University Press, 2012

43 Doostmohammadian M, Kharazmi S, Rabiee H R. How clustering affects the convergence of decentralized optimization over networks: a Monte-Carlo-based approach. Soc Netw Anal Min, 2024, 14: 135

Appendix A

In the appendix, we give the detailed proof of Lemma 2. For this purpose, several intermediate results are first presented. To begin with, an important property of strongly convex and smooth functions is given, which is the key point for the subsequent convergence analysis.

Lemma 3. Let function g be μ-strongly convex and L-smooth on $\mathbb{R}^n$. For any $\mathbf{s}, \mathbf{s}' \in \mathbb{R}^n$ and $0 < \alpha < \frac{2}{L}$, there holds that

$$\|(\mathbf{s} - \mathbf{s}') - \alpha(\nabla g(\mathbf{s}) - \nabla g(\mathbf{s}'))\| \leqslant \lambda \|\mathbf{s} - \mathbf{s}'\|,$$

where $\lambda = \max\{|1 - \alpha\mu|, |1 - \alpha L|\}$.

Proof. The proof of this lemma can be directly completed based on the proof of [29, Lemma 10] and is thus omitted here.

Then, for establishing the inequality relationship between $\|\mathbf{u}(l+1) - \mathbf{1}\bar{\mathbf{u}}(l+1)\|$ and $\|\mathbf{u}(l) - \mathbf{1}\bar{\mathbf{u}}(l)\|$, several intermediate results are established in the following Lemmas 4–6.

Lemma 4. For all $l \geqslant 0$, $\mathbf{v}(l)$ under the iterative rule (7) and Assumption 2 satisfy

$$\mathbf{1}^{\mathrm{T}}\mathbf{v}(l) = 0. \tag{13}$$

Proof. Due to the doubly-stochasticity of the associated weighted matrix $\mathbf{W}$, we have $\mathbf{1}^{\mathrm{T}}\mathbf{W} = \mathbf{1}^{\mathrm{T}}$. Then for all $l \geqslant 0$, it can be obtained from the iteration (7) that

$$\begin{aligned}\mathbf{1}^{\mathrm{T}}\mathbf{v}(l+1) &= \mathbf{1}^{\mathrm{T}}\mathbf{v}(l) + \beta\mathbf{1}^{\mathrm{T}}(\mathbf{I} - \mathbf{W})\mathbf{u}(l+1)\\ &= \mathbf{1}^{\mathrm{T}}\mathbf{v}(l) + \beta(\mathbf{1}^{\mathrm{T}} - \mathbf{1}^{\mathrm{T}}\mathbf{W})\mathbf{u}(l+1)\\ &= \mathbf{1}^{\mathrm{T}}\mathbf{v}(l) + \beta(\mathbf{1}^{\mathrm{T}} - \mathbf{1}^{\mathrm{T}})\mathbf{u}(l+1)\\ &= \mathbf{1}^{\mathrm{T}}\mathbf{v}(l).\end{aligned}$$

Thus, for all $l \geqslant 0$, $\mathbf{1}^{\mathrm{T}}\mathbf{v}(l) = \mathbf{1}^{\mathrm{T}}\mathbf{v}(l-1) = \cdots = \mathbf{1}^{\mathrm{T}}\mathbf{v}(0) = 0$, where all $v_i(0) = 0$ are taken into consideration. The proof is complete.

Lemma 5. For all $l \geqslant 1$, $\mathbf{u}(l)$ under the iterative rule (7) satisfy

$$\mathbf{u}(l+1) = \mathbf{R}\mathbf{u}(l) + (\mathbf{s}(l) - \mathbf{s}(l-1)) - \alpha(\nabla\mathbf{G}(l) - \nabla\mathbf{G}(l-1)), \tag{14}$$

where $\mathbf{R} = (1-\beta)\mathbf{I} + \beta\mathbf{W}$.

Proof. From (7b), for all $l \geqslant 1$, we have

$$\mathbf{v}(l) = \mathbf{v}(l-1) + \beta(\mathbf{I} - \mathbf{W})\mathbf{u}(l). \tag{15}$$

Substituting (15) into (7a) yields

$$\mathbf{u}(l+1) = \mathbf{s}(l) - \alpha\nabla\mathbf{G}(l) - \mathbf{v}(l-1) - \beta(\mathbf{I} - \mathbf{W})\mathbf{u}(l). \tag{16}$$

From (7a), we also get

$$\mathbf{v}(l) = \mathbf{s}(l) - \mathbf{u}(l+1) - \alpha\nabla\mathbf{G}(l), \tag{17}$$

which means that

$$\mathbf{v}(l-1) = \mathbf{s}(l-1) - \mathbf{u}(l) - \alpha\nabla\mathbf{G}(l-1). \tag{18}$$

Substituting (18) into (16) leads to that (14) holds, thus completing the proof.

Lemma 6. Let $0 < \beta \leqslant 1$. Then the matrix $\mathbf{R}$ introduced in Lemma 5 is nonnegative doubly-stochastic.

Proof. Utilizing the doubly-stochastic property of the associated weighted matrix $\mathbf{W}$, by direct computation, we obtain

$$\mathbf{R}\mathbf{1} = (1-\beta)\mathbf{I}\mathbf{1} + \beta\mathbf{W}\mathbf{1} = (1-\beta)\mathbf{1} + \beta\mathbf{1} = \mathbf{1}, \tag{19}$$

$$\mathbf{1}^{\mathrm{T}}\mathbf{R} = \mathbf{1}^{\mathrm{T}}[(1-\beta)\mathbf{I} + \beta\mathbf{W}] = \mathbf{1}^{\mathrm{T}}(1-\beta) + \beta\mathbf{1}^{\mathrm{T}}\mathbf{W} = \mathbf{1}^{\mathrm{T}}. \tag{20}$$

Thus, the matrix $\mathbf{R}$ also possesses the doubly-stochastic property. Moreover, since $0 < \beta \leqslant 1$ and $\mathbf{W}$ is a nonnegative matrix, we have that all entries of $\mathbf{R}$ are nonnegative with consideration of the definition of $\mathbf{R}$. Therefore, $\mathbf{R}$ is a nonnegative doubly-stochastic matrix.

In the following Lemma 7, the inequality relationship between $\|\mathbf{u}(l+1) - \mathbf{1}\bar{\mathbf{u}}(l+1)\|$ and $\|\mathbf{u}(l) - \mathbf{1}\bar{\mathbf{u}}(l)\|$ is established.

Lemma 7. Let $0 < \alpha < \frac{1}{L}$. For all $l \geqslant 1$, $\mathbf{u}(l)$ and $\bar{\mathbf{u}}(l)$ under the iterative rule (7) and Assumption 1 satisfy

$$\|\mathbf{u}(l+1) - \mathbf{1}\bar{\mathbf{u}}(l+1)\| \leqslant \rho_1\|\mathbf{u}(l) - \mathbf{1}\bar{\mathbf{u}}(l)\| + \gamma\|\mathbf{u}(l) - \mathbf{u}(l-1)\|. \tag{21}$$

Proof. Due to the doubly-stochastic property of the matrix $\mathbf{R}$, it follows from (14) that

$$\bar{\mathbf{u}}(l+1)=\bar{\mathbf{u}}(l)+\frac{1}{M}\mathbf{1}^{\mathrm{T}}(\mathbf{s}(l)-\mathbf{s}(l-1))-\alpha\frac{1}{M}\mathbf{1}^{\mathrm{T}}(\nabla\mathbf{G}(l)-\nabla\mathbf{G}(l-1)). \tag{22}$$

Thus, noting (14), we get

$$\begin{aligned}\|\mathbf{u}(l+1)-\mathbf{1}\bar{\mathbf{u}}(l+1)\|=&\Big\|\mathbf{R}\mathbf{u}(l)+(\mathbf{s}(l)-\mathbf{s}(l-1))-\alpha(\nabla\mathbf{G}(l)-\nabla\mathbf{G}(l-1))\\&-\mathbf{1}\bar{\mathbf{u}}(l)-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}(\mathbf{s}(l)-\mathbf{s}(l-1))\\&+\alpha\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}(\nabla\mathbf{G}(l)-\nabla\mathbf{G}(l-1))\Big\|\\\leqslant&\|\mathbf{R}\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\|+\Big\|\mathbf{I}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\Big\|\\&\times\|(\mathbf{s}(l)-\mathbf{s}(l-1))-\alpha(\nabla\mathbf{G}(l)-\nabla\mathbf{G}(l-1))\|.\end{aligned} \tag{23}$$

For the term $\|\mathbf{R}\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\|$ involved in (23), since $\mathbf{R}$ is nonnegative doubly-stochastic, we have

$$\begin{aligned}\|\mathbf{R}\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\|&=\Big\|\Big(\mathbf{R}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\Big)(\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l))\Big\|\\&\leqslant\Big\|\mathbf{R}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\Big\|\|\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\|\\&=\rho_1\|\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\|.\end{aligned} \tag{24}$$

Then, we consider the term $\|\mathbf{I}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\|\|(\mathbf{s}(l)-\mathbf{s}(l-1))-\alpha(\nabla\mathbf{G}(l)-\nabla\mathbf{G}(l-1))\|$ involved in (23). First, from the properties of $\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}$, it can be obtained that $\|\mathbf{I}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\|=1$. Furthermore, under Assumption 1 and with $0<\alpha<\frac{1}{L}$, it can be obtained from Lemma 3 that

$$\begin{aligned}&\|(\mathbf{s}(l)-\mathbf{s}(l-1))-\alpha(\nabla\mathbf{G}(l)-\nabla\mathbf{G}(l-1))\|\\&\quad\leqslant\gamma\|\mathbf{s}(l)-\mathbf{s}(l-1)\|\\&\quad\leqslant\gamma\|\mathbf{u}(l)-\mathbf{u}(l-1)\|.\end{aligned} \tag{25}$$

Finally, substituting (24) and (25) into (23) yields (21), which completes the proof.

In the following Lemma 8, the inequality relationship between $\|\bar{\mathbf{u}}(l+1)-\mathbf{u}^*\|$ and $\|\bar{\mathbf{u}}(l)-\mathbf{u}^*\|$ is established.

Lemma 8. Let $0<\alpha<\frac{1}{L}$. For all $l\geqslant 0$, $\bar{\mathbf{u}}(l)$ under the iterative rule (7) and Assumption 1 satisfy

$$\sqrt{M}\|\bar{\mathbf{u}}(l+1)-\mathbf{u}^*\|\leqslant\sqrt{M}\gamma\|\bar{u}(l)-\mathbf{u}^*\|+\gamma\|\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\|. \tag{26}$$

Proof. From (7a) and Lemma 4, it follows that

$$\bar{\mathbf{u}}(l+1)=\frac{1}{M}\mathbf{1}^{\mathrm{T}}\mathbf{s}(l)-\alpha\frac{1}{M}\mathbf{1}^{\mathrm{T}}\nabla\mathbf{G}(l). \tag{27}$$

Thus, noting that

$$\mathbf{u}^*=\mathbf{s}^*-\alpha\nabla g(\mathbf{s}^*), \tag{28}$$

we obtain

$$\begin{aligned}\|\bar{\mathbf{u}}(l+1)-\mathbf{u}^*\|=&\Big\|\frac{1}{M}\mathbf{1}^{\mathrm{T}}\mathbf{s}(l)-\alpha\frac{1}{M}\mathbf{1}^{\mathrm{T}}\nabla\mathbf{G}(l)-\mathbf{s}^*+\alpha\nabla g(\mathbf{s}^*)\Big\|\\\leqslant&\|P_{S_0}(\bar{\mathbf{u}}(l))-\alpha\nabla g(P_{S_0}(\bar{\mathbf{u}}(l)))-\mathbf{s}^*+\alpha\nabla g(\mathbf{s}^*)\|\\&+\Big\|\frac{1}{M}\mathbf{1}^{\mathrm{T}}\mathbf{s}(l)-\alpha\frac{1}{M}\mathbf{1}^{\mathrm{T}}\nabla\mathbf{G}(l)\\&-P_{S_0}(\bar{\mathbf{u}}(l))+\alpha\nabla g(P_{S_0}(\bar{\mathbf{u}}(l)))\Big\|\\=&\|P_{S_0}(\bar{\mathbf{u}}(l))-\alpha\nabla g(P_{S_0}(\bar{\mathbf{u}}(l)))-\mathbf{s}^*+\alpha\nabla g(\mathbf{s}^*)\|\end{aligned}$$

$$\begin{aligned}
&+\Big\|\frac{1}{M}\mathbf{1}^{\mathrm{T}}\mathbf{s}(l)-\frac{1}{M}\mathbf{1}^{\mathrm{T}}P_S(\mathbf{1}\bar{\mathbf{u}}(l))\\
&-\alpha\Big(\frac{1}{M}\mathbf{1}^{\mathrm{T}}\nabla\mathbf{G}(l)-\frac{1}{M}\mathbf{1}^{\mathrm{T}}\nabla\bar{\mathbf{G}}(l)\Big)\Big\|\\
\leqslant{}&\big\|P_{S_0}(\bar{\mathbf{u}}(l))-\alpha\nabla g(P_{S_0}(\bar{\mathbf{u}}(l)))-\mathbf{s}^*+\alpha\nabla g(\mathbf{s}^*)\big\|\\
&+\Big\|\frac{1}{M}\mathbf{1}^{\mathrm{T}}\Big\|\big\|\mathbf{s}(l)-P_S(\mathbf{1}\bar{\mathbf{u}}(l))-\alpha(\nabla\mathbf{G}(l)-\nabla\bar{\mathbf{G}}(l))\big\|\\
\leqslant{}&\big\|P_{S_0}(\bar{\mathbf{u}}(l))-\alpha\nabla g(P_{S_0}(\bar{\mathbf{u}}(l)))-\mathbf{s}^*+\alpha\nabla g(\mathbf{s}^*)\big\|\\
&+\frac{1}{\sqrt{M}}\big\|\mathbf{s}(l)-P_S(\mathbf{1}\bar{\mathbf{u}}(l))-\alpha(\nabla\mathbf{G}(l)-\nabla\bar{\mathbf{G}}(l))\big\|. \qquad (29)
\end{aligned}$$

For the terms $\|P_{S_0}(\bar{\mathbf{u}}(l))-\alpha\nabla g(P_{S_0}(\bar{\mathbf{u}}(l)))-\mathbf{s}^*+\alpha\nabla g(\mathbf{s}^*)\|$ and $\frac{1}{\sqrt{M}}\|\mathbf{s}(l)-P_S(\mathbf{1}\bar{\mathbf{u}}(l))-\alpha(\nabla\mathbf{G}(l)-\nabla\bar{\mathbf{G}}(l))\|$ involved in (29), under Assumption 1 and with $0<\alpha<\frac{1}{L}$, it can be obtained from Lemma 3 that

$$\begin{aligned}
&\big\|P_{S_0}(\bar{\mathbf{u}}(l))-\alpha\nabla g(P_{S_0}(\bar{\mathbf{u}}(l)))-\mathbf{s}^*+\alpha\nabla g(\mathbf{s}^*)\big\|\\
\leqslant{}&\gamma\big\|P_{S_0}(\bar{\mathbf{u}}(l))-\mathbf{s}^*\big\|\\
\leqslant{}&\gamma\big\|\bar{\mathbf{u}}(l)-\mathbf{u}^*\big\| \qquad (30)
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{\sqrt{M}}\big\|\mathbf{s}(l)-P_S(\mathbf{1}\bar{\mathbf{u}}(l))-\alpha(\nabla\mathbf{G}(l)-\nabla\bar{\mathbf{G}}(l))\big\|\\
\leqslant{}&\frac{1}{\sqrt{M}}\gamma\big\|\mathbf{s}(l)-P_S(\mathbf{1}\bar{\mathbf{u}}(l))\big\|\\
\leqslant{}&\frac{1}{\sqrt{M}}\gamma\big\|\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\big\|. \qquad (31)
\end{aligned}$$

Thus, substituting (31) and (30) into (29) produces (26) and we have completed the proof.

In the following Lemma 9, the inequality relationship between $\|\mathbf{u}(l+1)-\mathbf{u}(l)\|$ and $\|\mathbf{u}(l)-\mathbf{u}(l-1)\|$ is established.

Lemma 9. Let $0<\alpha<\frac{1}{L}$. For all $l\geqslant 1$, $\mathbf{u}(l)$ and $\bar{u}(l)$ under the iterative rule (7) and Assumption 1 satisfy

$$\|\mathbf{u}(l+1)-\mathbf{u}(l)\|\leqslant\gamma\|\mathbf{u}(l)-\mathbf{u}(l-1)\|+2\beta\|\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\|. \qquad (32)$$

Proof. First, it can be obtained from (7a) that

$$\|\mathbf{u}(l+1)-\mathbf{u}(l)\|=\|\mathbf{s}(l)-\alpha\nabla\mathbf{G}(l)-\mathbf{s}(l-1)+\alpha\nabla\mathbf{G}(l-1)-\mathbf{v}(l)+\mathbf{v}(l-1)\|. \qquad (33)$$

Further, it follows from (33) and (7b) that

$$\|\mathbf{u}(l+1)-\mathbf{u}(l)\|\leqslant\|\mathbf{s}(l)-\alpha\nabla\mathbf{G}(l)-\mathbf{s}(l-1)+\alpha\nabla\mathbf{G}(l-1)\|+\beta\|(\mathbf{I}-\mathbf{W})\mathbf{u}(l)\|. \qquad (34)$$

For the term $\|\mathbf{s}(l)-\alpha\nabla\mathbf{G}(l)-\mathbf{s}(l-1)+\alpha\nabla\mathbf{G}(l-1)\|$ involved in (34), applying Assumption 1, the step-size condition $0<\alpha<\frac{1}{L}$ and Lemma 3, we get

$$\begin{aligned}
&\|\mathbf{s}(l)-\alpha\nabla\mathbf{G}(l)-\mathbf{s}(l-1)+\alpha\nabla\mathbf{G}(l-1)\|\\
&\leqslant\gamma\|\mathbf{s}(l)-\mathbf{s}(l-1)\|\\
&\leqslant\gamma\|\mathbf{u}(l)-\mathbf{u}(l-1)\|. \qquad (35)
\end{aligned}$$

For the term $\beta\|(\mathbf{I}-\mathbf{W})\mathbf{u}(l)\|$ involved in (34), from the doubly-stochastic property of the weighted matrix $\mathbf{W}$, it follows that $(\mathbf{I}-\mathbf{W})\mathbf{1}\bar{\mathbf{u}}(l)=0$. Then, we have

$$\begin{aligned}
\beta\|(\mathbf{I}-\mathbf{W})\mathbf{u}(l)\|&=\beta\|(\mathbf{I}-\mathbf{W})\mathbf{u}(l)-(\mathbf{I}-\mathbf{W})\mathbf{1}\bar{\mathbf{u}}(l)\|\\
&\leqslant\beta\|\mathbf{I}-\mathbf{W}\|\|\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\|\\
&\leqslant 2\beta\|\mathbf{u}(l)-\mathbf{1}\bar{\mathbf{u}}(l)\|. \qquad (36)
\end{aligned}$$

Finally, substituting (35) and (36) into (34) yields (32), thereby completing the proof.

With the results established in Lemmas 4–9 in hands, we are now ready to provide the detailed proof of Lemma 2 in the following.

Proof of Lemma 2. Since the step-sizes α and β satisfy (9), under Assumptions 1–3, it can be obtained directly from Lemmas 7–9 that (10) holds with $\mathbf{A}$, γ and ρ_1 as defined in Lemma 2. Thus, we only need to confirm that $\rho_1 \leqslant 1-\beta(1-\rho_0) < 1$ for completing the proof.

Clearly, for the matrix $\mathbf{R}$ as defined in Lemma 5, we obtain

$$\begin{aligned}\mathbf{R}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}} &= (1-\beta)\mathbf{I}+\beta\mathbf{W}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\\ &= (1-\beta)\Big(\mathbf{I}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\Big)+\beta\Big(\mathbf{W}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\Big),\end{aligned}\tag{37}$$

which implies that

$$\Big\|\mathbf{R}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\Big\| \leqslant (1-\beta)\Big\|\mathbf{I}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\Big\|+\beta\Big\|\mathbf{W}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\Big\|.\tag{38}$$

Furthermore, as discussed in the proof of Lemma 7, $\|\mathbf{I}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\|=1$. Then it follows from (38) that

$$\Big\|\mathbf{R}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\Big\| \leqslant (1-\beta)+\beta\rho_0 = 1-\beta(1-\rho_0).\tag{39}$$

Therefore, noting that $0<\beta\leqslant 1$ and $\rho_0<1$, we arrive at

$$\Big\|\mathbf{R}-\frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}\Big\| \leqslant 1-\beta(1-\rho_0) < 1.\tag{40}$$

So far, we have completed the proof.