Distributed Algorithm Over Time-Varying Unbalanced Topologies for Optimization Problem Subject to Multiple Local Constraints

Hongzhe Liu, Wenwu Yu, Guanghui Wen, and Wei Xing Zheng, Fellow, IEEE

Abstract

This paper studies the distributed optimization problem with possibly nonidentical local constraints, where its global objective function is composed of $N$ convex functions. The aim is to solve the considered optimization problem in a distributed manner over time-varying unbalanced directed topologies by using only local information and performing only local computations. Towards this end, a new distributed discrete-time algorithm is developed by synthesizing the row stochastic matrices sequence and column stochastic matrices sequence analysis technique. Furthermore, for the developed distributed discrete-time algorithm, its convergence property to the optimal solution as well as its convergence rate are established under some mild assumptions. Numerical simulations are finally presented to verify the theoretical results.

Index Terms

Convex optimization, distributed algorithm, time-varying unbalanced directed topologies, nonidentical local constraints.

I. INTRODUCTION

In recent decades, with the rapid development of various communication technologies, much more attention has been focused upon distributed control and optimization of multi-agent systems. As an important topic within this context, distributed optimization has recently received an increasing research interest due to its potential applications in numerous fields, such as formation control of robot networks, source location of sensor networks, machine learning, power grid dispatching, etc.

This work was supported in part by the National Natural Science Foundation of China under Grant No. 61673107, the General joint fund of the equipment advance research program of Ministry of Education under Grant No. 6141A020223, and the Jiangsu Provincial Key Laboratory of Networked Collective Intelligence under Grant No. BM2017002. (Corresponding author: Wenwu Yu.)

H. Liu, W. Yu, and G. Wen are with the Department of Systems Science, School of Mathematics, Southeast University, Nanjing 210096, China (e-mail: 230169408@seu.edu.cn; wwyu@seu.edu.cn; wenguanghui@gmail.com). G. Wen is also with the School of Engineering, RMIT University, Melbourne VIC 3001, Australia.

W. X. Zheng is with the School of Computer, Data and Mathematical Sciences, Western Sydney University, Sydney, NSW 2751, Australia (e-mail: w.zheng@westernsydney.edu.au).
In the past decade, many interesting results have been reported in the literature, where distributed optimization algorithms including continuous-time algorithms and discrete-time algorithms have been developed respectively for solving such optimization problems under different circumstances. The results on various distributed continuous-time algorithms can be found in [1]–[15] and references therein. This paper mainly focuses on designing a distributed discrete-time algorithm.

In the context of distributed discrete-time optimization, the distributed discrete-time algorithms were given in [16]–[21] to solve the optimization problems over balanced topologies under different scenarios. Later, the result was extended to the case with time-varying unbalanced directed topologies in [22] and [23], where the distributed algorithms were designed for the unconstrained optimization problem with the column stochastic matrices used. Subsequently, built on the work in [22], a distributed algorithm was developed in [24] over time-varying unbalanced directed topologies by integrating the dual averaging method into the push-sum mechanism, which can solve the optimization problem with a global closed convex set constraint. Moreover, the optimization problem with coupled linear equality constraint was addressed in a similar way in [25]. Additionally, based on [23], the novel push-pull algorithms with both row stochastic matrices and column stochastic matrices were proposed in [26]–[29] for the unconstrained optimization problem over unbalanced directed graphs which can be static or time-varying. However, those results cannot be applied in the constrained case. Recently, in [30] and [31], some attempts were made at adopting row stochastic iterative matrices for dealing with the optimization problems over static unbalanced directed topologies as the row stochastic property is very standard and easy to be satisfied in the distributed setting. In [30], an optimization problem with identical compact set constraint was considered and an algorithm based on the mechanism of estimating the left eigenvector associated with the eigenvalue 1 was proposed. The result in [30] was further extended in [31] to a more general case where nonidentical general closed convex set constraints were involved. To get a clear overview of the state of the art, Table I summarizes a detailed comparison of the

| References | Constraints | Linear convergence rate | Unbalanced topologies | Time-varying topologies |
|------------|-------------|-------------------------|-----------------------|-------------------------|
| [16]       | ×           | ×                       | ×                     | ✓                       |
| [17], [18] | ✓           | ×                       | ×                     | ✓                       |
| [19]       | ×           | ✓                       | ×                     | ✓                       |
| [20], [21] | ✓           | ×                       | ×                     | ✓                       |
| [22]       | ×           | ✓                       | ✓                     | ✓                       |
| [23]       | ×           | ✓                       | ✓                     | ✓                       |
| [24], [25] | ✓           | ×                       | ✓                     | ✓                       |
| [26]–[28]  | ×           | ✓                       | ✓                     | ✓                       |
| [29]       | ×           | ✓                       | ✓                     | ✓                       |
| [30], [31] | ✓           | ×                       | ✓                     | ✓                       |

✓ means that this feature is involved
× means that this feature is uninvolved
aforementioned results in [16], [17], [20]–[31].

Although the good result was achieved in [24], it is worth mentioning that the operation $\arg\min_{x \in X}$ was involved in the algorithm designed in [24], with $X$ being the constraint set. Thus, each iteration of the designed algorithm therein is required to solve a subproblem $\arg\min_{x \in X}$ first, which could be a numerically demanding task. Additionally, it should be noted that only the optimization problem with a global closed convex set constraint was considered in [24], but real-world optimization problems necessarily involve more general constraints quite often. Motivated by this, we aim to solve the optimization problem with more general constraints over time-varying unbalanced directed topologies in this paper, for which a desirable distributed discrete-time algorithm will be designed without involving any in-built operation $\arg\min$.

Specifically, the optimization problem with $N$ nonidentical convex inequality constraints and compact set constraints is considered in this paper. A new distributed discrete-time algorithm is proposed over time-varying unbalanced directed topologies and a rigorous analysis of its convergence to the optimal solution of the considered problem is made under some mild assumptions. Moreover, a detailed analysis of the convergence rate is also shown. Thus, the major contribution of this paper is that a challenging problem is successfully solved by a novel distributed discrete-time algorithm. In detail, first, although both row stochastic matrices and column stochastic matrices are involved, the designed distributed discrete-time algorithm herein is essentially different from the algorithms proposed in [26]–[29] since different analysis methods are adopted. As a result, the optimization problem with multiple nonidentical constraints can be solved over time-varying unbalanced directed topologies in this paper, which, however, could not in [26]–[29]. Second, the strong convexity of local objective functions is not necessarily required in this paper, while this is the key to the convergence analysis made in [26]–[29]. Third, unlike the results given in [30] and [31], auxiliary variables are introduced in the present paper to estimate the gradients of the global objective function rather than the left eigenvector for canceling out the gradients error of the global objective function caused by the asymmetry of time-varying unbalanced directed topologies. Thus, the dimensions of auxiliary variables as well as the total dimensions of all generated vectors associated with the proposed algorithm are considerably lower than those in [30] and [31] when large-scale networks are involved, which is of great importance in real applications. Fourth, compared to the algorithm designed in [24], the optimization problem with much more general constraints can be resolved by the algorithm proposed in this paper, where the in-built operation $\arg\min$ is not involved at all. Finally, this paper potentially provides an efficient mechanism for designing distributed algorithms to solve the optimization problem with very general constraints over time-varying unbalanced directed topologies.

The remaining parts are organized as follows. Section II recalls some preliminaries, including notations, graph theory, and useful lemmas. Section III shows the problem formulation, presents the algorithm development, and gives the main theorems on the convergence property and the convergence rate. In Section IV the detailed proofs of the main theorems are made. Section V provides some numerical simulations and Section VI gives a summary of the work finally. Besides, the proofs of intermediate results are furnished in the Appendix.
II. Preliminaries

A. Notations

Let $\mathbb{R}^n$ denote the $n$-dimensional real-valued vectors set and 1 denote the vector with proper dimension and all its entries being 1. Let $\mathbb{R}^{N \times N}$ represent the set of $N$-dimensional real-valued square matrices and $I_N$ represent the $N$-dimensional identity matrix. Furthermore, for a given matrix $M \in \mathbb{R}^{N \times N}$ and $i, j = 1, 2, \cdots, N$, $M_{ij}$ or $[M]_{ij}$ denotes the $ij$-th entry of matrix $M$. Particularly, for a given vector $z \in \mathbb{R}^n$, $z^T$ is the transpose of $z$, $||z||$ stands for the Euclidean norm of $z$, and the scalar $z^i$ represents the $i$-th entry of $z$. Moreover, for a given convex function $f: \mathbb{R}^n \to \mathbb{R}$, $\nabla f(x)$ (or $\partial f(x)$) stands for the gradient (sub-gradient) of function $f$ in $x$. Besides, for a scalar $a$, $a^+ = \max\{a, 0\}$. Additionally, for a given set $\Omega \subseteq \mathbb{R}^n$, $\text{relint} \Omega$ denotes the relative interior of the set $\Omega$.

B. Graph Theory

Let $\{G(t)\}$ stand for a directed graph sequence with a common node set $\mathcal{N} = \{1, 2, \cdots, N\}$ and a directed edge set sequence $\{E(t) \subseteq \mathcal{N} \times \mathcal{N}\}$, where $t = 0, 1, 2, \cdots$. Further, for a given natural number $t$, $(i, j) \in E(t)$ means that there exists a directed edge from node $j$ to node $i$, that is, node $i$ can receive information from node $j$, node $j$ is an in-neighbor of node $i$ and meanwhile node $i$ is an out-neighbor of node $j$. Moreover, define $\mathcal{N}^i_{\text{in}}(t)$ as the set of all node $i$’s in-neighbors and $d^+_i(t)$ as its cardinality at $t$. Similarly, define $\mathcal{N}^i_{\text{out}}(t)$ as the collection of all node $i$’s out-neighbors and $d^-_i(t)$ as its cardinality at $t$. Particularly, we always let $i \in \mathcal{N}^i_{\text{in}}(t) \cap \mathcal{N}^i_{\text{out}}(t)$, that is, $(i, i) \in E(t)$, for all $t \geq 0$. Further, for a given $t$, a directed path from node $i_m$ and node $i_1$ is defined as a sequence of $m$ distinct nodes $i_1, \cdots, i_m$ such that $(i_q, i_{q+1}) \in E(t)$ with $q = 1, 2, \cdots, m - 1$. A static directed graph $G$ is called a strongly connected graph if there exists a directed path between any two distinct nodes in graph $G$. For a given natural number $t$ and a positive integer $H$, the joint graph $\bigcup_{i=t}^{t+H-1} G(i)$ denotes the graph with the node set $\mathcal{N} = \{1, 2, \cdots, N\}$ and a directed edge set consisting of the union of the directed edge sets of graphs $G(t), G(t + 1), \cdots, G(t + H - 1)$. Furthermore, the directed graph sequence $\{G(t)\}$ is said to be uniformly jointly strongly connected if for all $t \geq 0$, there is a positive integer $H$ such that the joint graph $\bigcup_{i=t}^{t+H-1} G(i)$ is strongly connected.

C. Useful Lemmas

In this subsection, some auxiliary results are shown for facilitating analysis of the main results.

**Lemma 1** ([17], [34]): Suppose that $Y \subseteq \mathbb{R}^n$ is a closed convex set and $P_Y$ is the projection operator on $Y$, i.e., $P_Y(u) = \arg\min_{v \in Y} ||u - v||$. Then, for all $x_1, x_2 \in \mathbb{R}^n$, and $x_3 \in Y$, there hold

(a) $\|P_Y(x_1) - P_Y(x_2)\| \leq \|x_1 - x_2\|$

(b) $\|P_Y(x_1) - x_3\|^2 \leq \|x_1 - x_3\|^2 - \|P_Y(x_1) - x_1\|^2$.

**Lemma 2** ([22]):

(a) For a scalar sequence $\{\theta(k)\}$, suppose that $\lim_{k \to \infty} \theta(k) = \theta$ and $0 < \rho < 1$ hold. Then $\lim_{k \to \infty} \sum_{l=0}^{k} \rho^{k-l} \theta(l) = \frac{\theta}{1-\rho}$.

(b) For a strictly positive scalar sequence $\{\theta(k)\}$, assume that $\sum_{k=0}^{\infty} \theta(k) < \infty$ and $0 < \rho < 1$. Then $\sum_{k=0}^{\infty} \left( \sum_{l=0}^{k} \rho^{k-l} \theta(l) \right) < \infty$. 

DRAFT
Lemma 3 (22): For nonnegative scalar sequences \( \{a(k)\} \), \( \{b(k)\} \), \( \{c(k)\} \) and \( \{d(k)\} \) with \( \sum_{k=0}^{\infty} b(k) < \infty \) and \( \sum_{k=0}^{\infty} c(k) < \infty \), let the following condition hold:

\[
a(k + 1) \leq (1 + b(k))a(k) - d(k) + c(k), \quad \forall k \geq 0.
\]

Then, the scalar sequence \( \{a(k)\} \) converges to some \( a \geq 0 \) and \( \sum_{k=0}^{\infty} d(k) < \infty \).

In the next Lemmas 4 and 5 we reveal some important properties of the row stochastic matrices sequence and the column stochastic matrices sequence associated with a sequence of uniformly jointly strongly connected directed graphs, which are the key points for the subsequent analysis of consensus and our improved push-pull mechanism.

Lemma 4 (22): For a uniformly jointly strongly connected directed graphs sequence \( \{G(k)\} \) and \( k \geq s \geq 0 \), let \( A(k : s) = A(k) \cdots A(s) \) when \( k > s \), and \( A(k : s) = A(k) \) when \( k = s \), with \( \{A(k)\} \) being a sequence of nonnegative row stochastic matrices associated with \( \{G(k)\} \). Also, assume that there exists a positive constant \( a_0 \) such that \( A_{ii}(k) \geq a_0 \) hold for all \( i = 1, 2, \ldots, N \) and \( k \geq 0 \). Then, \( \forall k \geq 0 \), there exists a positive vector sequence \( \{\pi(k)\} \) satisfying \( \pi^T(k)1 = 1 \) and the following conditions:

(a) For all \( i, j \in \mathcal{N} \) and \( k \geq s \), there are two constants \( C_1 > 0 \) and \( 0 < \lambda_1 < 1 \) such that \( |a_{ij}(k : s) - \pi^j(s)| \leq C_1 \lambda_1^{k-s} \);

(b) For all \( i \in \mathcal{N} \) and \( k \geq 0 \), there is a strictly positive constant \( \theta_1 \) such that \( \pi^i(k) \geq \theta_1 \);

(c) For all \( k \geq 0 \), \( \pi^T(k) = \pi^T(k + 1)A(k) \).

Lemma 5 (22): For a uniformly jointly strongly connected directed graphs sequence \( \{G(k)\} \) and \( k \geq s \geq 0 \), let \( B(k : s) = B(k) \cdots B(s) \) when \( k > s \), and \( B(k : s) = B(k) \) when \( k = s \), with \( \{B(k)\} \) being a sequence of nonnegative column stochastic matrices associated with \( \{G(k)\} \). Furthermore, assume that there exists a positive constant \( b_0 \) such that \( B_{ii}(k) \geq b_0 \) hold for all \( i = 1, 2, \ldots, N \) and \( k \geq 0 \). Then, \( \forall k \geq 0 \), there exists a positive vector sequence \( \{\mu(k)\} \) satisfying \( \mu^T(k)1 = 1 \) and the following conditions:

(a) For all \( i, j \in \mathcal{N} \) and \( k \geq s \), there are two constants \( C_2 > 0 \) and \( 0 < \lambda_2 < 1 \) such that \( |b_{ij}(k : s) - \mu^j(k)| \leq C_2 \lambda_2^{k-s} \);

(b) For all \( i \in \mathcal{N} \) and \( k \geq 0 \), there is a strictly positive constant \( \theta_2 \) such that \( \mu^i(k) \geq \theta_2 \).

Moreover, in the following Lemma 6 we provide a key result for finding the point \( x \) satisfying \( g(x) \leq 0 \), with \( g(x) \) being a convex function.

Lemma 6 (23): Let \( Y \subseteq \mathbb{R}^n \) be an arbitrary closed convex set, \( g : \mathbb{R}^n \to \mathbb{R} \) be a convex function, and \( x \in \mathbb{R}^n \) be given by

\[
x = P_Y \left( v - \beta_0 \frac{g^+(v)}{\|d\|^2} d \right),
\]

where \( v \in \mathbb{R}^n \), \( 0 < \beta_0 < 2 \) is a constant, and \( d \in \partial g^+(v) \) when \( g^+(v) > 0 \) and \( d = d_0 \not= 0 \) otherwise, with \( d_0 \) being an arbitrary nonzero constant vector. Then, for all \( z \in Y \) satisfying \( g^+(z) = 0 \), there holds

\[
\|x - z\|^2 \leq \|v - z\|^2 - \beta_0 (2 - \beta_0) \frac{(g^+(v))^2}{\|d\|^2}.
\]
Based on (1) given in Lemma 6, in order to find the point \( x \) satisfying \( g(x) \leq 0 \) with \( g(x) \) being a convex function, an iterative rule can be correspondingly designed as
\[
x(t + 1) = P_Y \left( x(t) - \beta_t \frac{g^+(x(t))}{\|d^t\|^2} d^t \right),
\]
with an arbitrary initial value \( x(0) \), and \( d^t \in \partial g^+(x(t)) \) when \( g^+(x(t)) > 0 \) and \( d^t = d_0 \neq 0 \) otherwise. Clearly, the above iterative rule (3a) can be equivalently written as
\[
x(t + 1) = P_Y (x(t) - \beta_t \partial g^+(x(t))),
\]
which can be approximately seen as the classical projected subgradient decent iteration, with the nonnegative stepsize \( \beta_t = \beta_0 \frac{g^+(x(t))}{\|d^t\|^2} \) if \( g^+(x(t)) > 0 \) and \( \beta(x(t)) = 0 \) otherwise. Since \( \beta(x(t)) \) is positive provided that \( g^+(x(t)) > 0 \), the value \( g^+(x(t+1)) \) will become smaller than \( g^+(x(t)) \) when \( g^+(x(t)) > 0 \), which means that the distance from \( x(t+1) \) to the set \( \{ y | g(y) \leq 0 \} \) will be smaller than the distance from \( x(t) \) to the set \( \{ y | g(y) \leq 0 \} \).

Furthermore, for an arbitrary \( z \in \{ y | g(y) \leq 0 \} \), the relation between \( x(t + 1) \) and \( z \) can be characterized as
\[
\|x(t + 1) - z\|^2 \leq \|x(t) - z\|^2 - \beta_t (2 - \beta_0) \frac{(g^+(x(t)))^2}{\|d^t\|^2}.
\]
In addition, it can also be concluded that the above iteration rule (3b) will stop at the point \( x \) satisfying \( g^+(x(t)) = 0 \) or \( \partial g^+(x(t)) = 0 \), both of which can imply \( g(x) \leq 0 \) with the consideration that \( g(x) \) is convex.

III. MAIN RESULTS

A. Problem Formulation

In this paper, the optimization problem with \( N \) nonidentical inequality constraints and \( N \) nonidentical closed convex set constraints is researched, which is formulated as
\[
\min f(x) = \sum_{i=1}^{N} f_i(x)
\]
\[
s.t. \quad g_i(x) \leq 0,
\]
\[
x \in X_i, \quad i = 1, 2, \ldots, N,
\]
where \( x \in \mathbb{R}^n \), \( f_i : \mathbb{R}^n \to \mathbb{R} \) and \( g_i : \mathbb{R}^n \to \mathbb{R} \) are convex functions, and \( X_i \subseteq \mathbb{R}^n \) are closed convex sets. Furthermore, a directed graphs sequence \( \{G(k)\} \) is introduced to describe local interactions between nodes in this paper. Slater’s condition is first assumed for the problem (4).

Assumption 1: There exists a vector \( \hat{x} \in \text{relint}X = \bigcap_{i=1}^{N} X_i \) such that \( g_i(\hat{x}) < 0 \) for all \( i = 1, 2, \ldots, N \).

The optimality conditions for the problem setup are as follows. Let \( X_0 \) and \( X^* \) denote the feasible solutions set and the optimal solutions set of the considered problem, respectively. Specially, under Assumption 1 we can get that \( X_0 \) and \( X^* \) are nonempty. Meanwhile, let \( x^* \) be an optimal solution to the problem (4) and assume that the optimal function value \( f^* \) is finite. Particularly, the following properties of \( x^* \) play an important role in the subsequent analysis:
\[
f(x^*) \leq f(x) \quad \text{for all } x \in X_0, \quad g_0(x^*) \leq 0, \quad \text{and } x^* \in X,
\]
with \( g_0(x) = \max\{g_1(x), g_2(x), \ldots, g_N(x)\} \).
B. Algorithm Development

First, to solve the above optimization problem (4), a distributed discrete-time algorithm is developed as below:

\[ v_i(t) = \sum_{j=1}^{N} A_{ij}(t)x_j(t) - y_i(t), \quad \text{(5a)} \]

\[ x_i(t + 1) = P_{X_i}(v_i(t) - \beta k_i(t)), \quad \text{(5b)} \]

\[ y_i(t + 1) = \sum_{j=1}^{N} B_{ij}(t)y_j(t) + \alpha(t + 1)\nabla f_i(x_i(t + 1)) - \alpha(t)\nabla f_i(x_i(t)), \quad \text{(5c)} \]

where

\[ k_i(t) = \frac{g_i^+(v_i(t))}{\|d_i(t)\|^2}d_i(t), \]

with

\[ d_i(t) = \begin{cases} \partial g_i^+(v_i(t)), & g_i^+(v_i(t)) \neq 0, \\ d_0 \text{ (a nonzero constant vector)}, & g_i^+(v_i(t)) = 0. \end{cases} \]

Moreover, \( \{A(t)\} \subseteq \mathbb{R}^{N \times N} \) is a nonnegative row stochastic weight matrices sequence associated with the given directed graphs sequence \( \{G(t)\} \), whose entries can be defined as \( A_{ij}(t) = \frac{1}{d^i_j(t)} \) if \( (i,j) \subseteq \mathcal{E}(t) \) and \( A_{ij}(t) = 0 \) otherwise; and \( \{B(t)\} \subseteq \mathbb{R}^{N \times N} \) is a nonnegative column stochastic weight matrices sequence associated with the given directed graphs sequence \( \{G(t)\} \), whose entries can be defined as \( B_{ij}(t) = \frac{1}{d^j_i(t)} \) if \( (i,j) \subseteq \mathcal{E}(t) \) and \( B_{ij}(t) = 0 \) otherwise. In addition, \( \{\alpha(t)\} \) is a positive decaying step-size sequence satisfying \( \sum_{t=0}^{\infty} \alpha(t) = \infty \) and \( \sum_{t=0}^{\infty} \alpha^2(t) < \infty \), and \( \beta \) is a constant step-size satisfying \( 0 < \beta < 2 \). Specially, we select \( x_i(0) \) as an arbitrary value and \( y_i(0) = \alpha(0)\nabla f_i(x_i(0)) \).

It is worth mentioning that the algorithm (5) is designed on the basis of the push-pull mechanism with some significant improvements. Thus, we first give some necessary description of the push-pull based algorithms before providing some insights into the algorithm (5). In (26), the push-pull/AB algorithm was developed as

\[ x_i(t + 1) = \sum_{j=1}^{N} A_{ij}x_j(t) - \alpha y_i(t), \]

\[ y_i(t + 1) = \sum_{j=1}^{N} B_{ij}y_j(t) + \nabla f_i(x_i(t + 1)) - \nabla f_i(x_i(t)), \]

where \( \{A_{ij}\} \) and \( \{B_{ij}\} \) are respectively a nonnegative row stochastic matrices sequence and a nonnegative column stochastic matrices sequence associated with the given static unbalanced graph, and \( \alpha \) is a desired constant step-size. Moreover, the initial values \( x_i(0) \) can be arbitrarily selected and \( y_i(0) = \nabla f_i(x_i(0)) \) for all \( i \). Different kinds of push-pull/AB algorithms were designed for different cases in (27)–(29) and the more detailed information on the push-pull/AB algorithms can be found in (37). Clearly, it can be seen from (26)–(29) that the constant step-sizes were usually employed in the iterations of the state variables \( x_i(t) \) in the push-pull/AB algorithms. Furthermore, it can also be noticed from the above existing results that the convergence analysis of the push-pull/AB algorithms depends heavily on constructing some linear matrix inequalities with the strong convexity properties being imposed on all local objective functions and only the unconstrained optimization problems can be dealt with. Nowadays,
it is still rather difficult to solve the constrained optimization problem by employing the push-pull/AB algorithms, which remains as an open issue.

In this work, in order to effectively integrate the methods of handling the nonidentical local inequality constraints and closed convex set constraints into the classical push-pull mechanism, the decaying step-size \( \alpha(t) \) is newly employed in our improved push-pull mechanism. Specifically, the decaying step-size \( \alpha(t) \) is involved in the iterations of the auxiliary variables \( y_i(t) \) rather than the state variables \( x_i(t) \) (see the iterative rule (5c)). Thus, the auxiliary variables \( y_i(t) \) can track the gradient terms \( \alpha(t)\mu^i(t-1)\sum_{i=1}^{N} \nabla f_i(x_i(t)) \) with \( \mu^i(t) \) as introduced in Lemma 5 and the tracking errors can be clearly characterized by some nonnegative variables which can be well analyzed under our improved push-pull mechanism now. Accordingly, the iterative rule (5b) can be approximately written as

\[
x_i(t+1) = P_{X_i}(v_i(t) - \beta k_i'(t)),
\]

where \( v_i'(t) \) is defined by replacing all terms \( y_i(t) \) in \( v_i(t) \) with \( \alpha(t)\mu^i(t-1)\sum_{i=1}^{N} \nabla f_i(x_i(t)) \) and \( k_i'(t) \) is defined by replacing all terms \( v_i(t) \) in \( k_i(t) \) with \( v_i'(t) \). Based on Lemma 5 it can be obtained that \( v_i'(t) \) will move to the set \( \{ y \mid g_i(y) \leq 0 \} \). Moreover, noting that \( \lim_{t \to \infty} \alpha(t) = 0 \) and \( \mu^i(t-1)\sum_{i=1}^{N} \nabla f_i(x_i(t)) \) is uniformly bounded with respect to \( t \), we can get that \( z_i(t) = \sum_{j=1}^{N} A_{ij}(t)x_j(t) \) will also move to the set \( \{ y \mid g_i(y) \leq 0 \} \). Additionally, since \( \{ A(t) \} \) is the nonnegative row stochastic matrices sequence, the term \( \sum_{j=1}^{N} A_{ij}(t)x_j(t) \) will enable \( x_i(t) \) to move towards the varying consensus state \( \sum_{i=1}^{N} \pi^i(t)x_i(t) \) with \( \pi^i(t) \) as introduced in Lemma 4. Thus, it follows that

\[
\lim_{t \to \infty} \left\| x_i(t) - \sum_{i=1}^{N} \pi^i(t)x_i(t) \right\| = \lim_{t \to \infty} \| x_i(t) - z_i(t) \| = 0,
\]

i.e., \( x_i(t) \) will also move to the set \( \{ y \mid g_i(y) \leq 0 \} \) and then to the set \( \{ y \mid g_i(y) \leq 0, i = 1, 2, \cdots, N \} \) since all \( x_i(t) \) will reach consensus. Furthermore, each varying state \( x_i(t) \) is also controlled by the projection operation on the set \( X_i \) and pushed by the gradient term \( \alpha(t)\mu^i(t-1)\sum_{i=1}^{N} \nabla f_i(x_i(t)) \) to move towards the set \( X_i \) and then to the optimal solution set of the optimization problem

\[
\min_{x} \sum_{i=1}^{N} f_i(x) \text{ s.t. } g_i(x) \leq 0 \quad x \in X_i, \quad i = 1, 2, \cdots, N.
\]

Consequently, upon reaching consensus, all states will converge to a common optimal solution of the considered problem 4.

Here, we also provide some insights into the algorithm 5 from the perspective of centralized iteration. The corresponding centralized iteration of the algorithm 5 can be formulated as

\[
x(t+1) = P_{X}(x(t) - \alpha(t)\nabla f(x(t)) - \beta k(t)),
\]

where

\[
k(t) = \begin{cases} \frac{g_0^+(v(t))}{\| \partial g_0^+(v(t)) \|^2} \partial g_0^+(v(t)), & g_0^+(v(t)) \neq 0, \\ 0, & \text{otherwise,} \end{cases}
\]

with \( v(t) = x(t) - \alpha(t)\nabla f(x(t)) \). Moreover, the convergence property of the above centralized iteration 8 can be deduced from Lemma 3 and Lemma 6 which can be found in 34. Indeed, in the remainder of the paper, we will
resort to the techniques used in the convergence analysis of the centralized iteration together with characterizing the error terms \(\|y_i(t) - \alpha(t)\mu^i(t-1) \sum_{i=1}^{N} \nabla f_i(x_i(t))\|\) and \(\alpha(t)\|x_i(t) - \sum_{i=1}^{N} \pi_i^i(t) x_i(t)\|\) by some well analyzed nonnegative variables to complete the convergence analysis of the developed distributed algorithm (5), which will be clearly shown in Section IV.

**Remark 1:** It can be concluded from the above discussion that some significant improvements on the classical push-pull/AB algorithms are made in our paper. First, the decaying step-size \(\alpha(t)\) is newly employed in our improved push-pull mechanism and the decaying step-size is involved in the iterations of the auxiliary variables \(y_i(t)\) rather than the state variables \(x_i(t)\). Second, the methods of handling the nonidentical local inequality constraints and closed convex set constraints are effectively integrated into our improved push-pull mechanism. As a consequence, the convergence analysis of the algorithm (5), as to be shown in Section IV, is conducted based on Lemma 3, which is absolutely distinct from the convergence analyses as given in [26]–[29] for the existing push-pull/AB algorithms. Third, it is not necessary to assume the strong convexity of the local objective functions or to construct any linear matrix inequalities for proceeding with the convergence analysis. More importantly, the optimization problem with multiple nonidentical local constraints is successfully addressed over a sequence of time-varying unbalanced graphs under our improved push-pull mechanism in this work, which is known as a very challenging issue in the field of distributed optimization so far. Furthermore, it will also be convenient to extend the results obtained in this paper to the case with other kinds of distributed constrained optimization problems.

### C. Main Theorems

Before going on, the definition of \(L\)-smooth function and several common assumptions are introduced for the establishment of the main theorems.

**Definition 1:** A differentiable function \(h : \mathbb{R}^n \rightarrow \mathbb{R}\) is said to be \(L\)-smooth if the following condition holds:
\[
\forall x_1, x_2 \in \mathbb{R}^n, \quad \| \nabla h(x_1) - \nabla h(x_2) \| \leq L \| x_1 - x_2 \|.
\]

**Assumption 2:** For all \(i = 1, 2, \cdots, N\), functions \(f_i(x)\) are \(L\)-smooth and functions \(g_i(x)\) have continuous gradients.

**Assumption 3:** For all \(i = 1, 2, \cdots, N\), \(X_i \subseteq \mathbb{R}^n\) are compact sets.

**Proposition 1:** There exists a constant \(R\) such that
\[
\text{dist}(x, X_0) \leq R \max \left\{ \max_{1 \leq i \leq N} \text{dist}(x, X_i), \max_{1 \leq i \leq N} g_i^+(x) \right\},
\]
for all \(x \in U = \text{conv} \left( \bigcup_{i=1}^{N} X_i \right)\).

Clearly, under Assumptions [3] Proposition [1] will naturally hold since the involved variables are uniformly bounded when \(x \in U\). Moreover, Proposition [1] (also known as the constraint regularity property) is very important and standard in dealing with the optimization problems involving nonidentical constraints, which can also be found in [31] and [34].

**Assumption 4:** The time-varying directed graph sequence \(\{\mathcal{G}(t)\}\) is uniformly jointly strongly connected.
Remark 2: Under Assumptions 2 and 3, we can conclude that $\nabla f_i(x_i(t))$ can be uniformly bounded by a positive constant $M$, i.e., $\|\nabla f_i(x_i(t))\| \leq M$, since $\nabla f_i(x_i(t))$ are continuous and $x_i(t)$ generated by the iteration rule (5b) are contained in the compact set $U$. Besides, Assumption 4 is also very common and standard in dealing with the optimization problems over time-varying unbalanced directed graphs.

Now, we are prepared to develop our first main theorem, which demonstrates the convergence property of the developed distributed discrete-time algorithm (5).

**Theorem 1:** Let Assumptions 1, 2, 3, and 4 hold. Then for all $i = 1, 2, \cdots, N$, the states $x_i(t)$ generated by the algorithm (5) converge to a common optimal solution $s^*$ of the problem (4), i.e., $\lim_{t \to \infty} x_i(t) = s^*$.

**Proof.** The detailed proof will be given in the subsequent Subsection IV-C.$\square$

Next, defining the following two variables:

$$\tilde{x}_i(t) = \frac{\sum_{k=0}^{t} \alpha(k)x_i(k)}{\sum_{k=0}^{t} \alpha(k)}, \quad \tilde{s}(t) = \frac{\sum_{k=0}^{t} \alpha(k)s(k)}{\sum_{k=0}^{t} \alpha(k)},$$

with $s(t) = P_{X_0}(\bar{x}(t))$ and $\bar{x}(t) = \sum_{i=1}^{N} \pi(t)x_i(t)$, we state our second main theorem showing the rate at which the function values $f(\tilde{x}_i(t))$ converge to the optimal value $f^*$.

**Theorem 2:** Let Assumptions 1, 2, 3, and 4 hold. Then for the algorithm (5), there holds

$$H_1 \|\tilde{x}_i(t) - \tilde{s}(t)\| + f(\tilde{s}(t)) - f^* \leq L(t), \quad (9)$$

with

$$L(t) = \left( H_2 + H_3 \sum_{k=0}^{t} \alpha^2(k) \right) / \sum_{k=0}^{t} \alpha(k), \quad (10)$$

and $H_j, j = 1, 2, 3$, being positive constants which will be defined in the subsequent proof of Theorem 2. Furthermore, for each $i = 1, 2, \cdots, N$, there holds

$$|f(\tilde{x}_i(t)) - f^*| \leq \left( \frac{NM}{H_1} + 1 \right) L(t), \quad (11)$$

with $M$ being the positive constant as given in Remark 2.

**Proof.** The detailed proof will be given in the subsequent Subsection IV-D.$\square$

**Remark 3:** Clearly, we can select $O\left(\frac{1}{t+1}\right)^\sigma$ with $0.5 < \sigma \leq 1$ as the decaying step-size $\alpha(t)$. Now, let $\alpha(t) = O\left(\frac{1}{\sqrt{t+1}}\right)$. By direct computation, it is easy to get

$$L(t) \sim O\left(\frac{\ln t}{\sqrt{t}}\right),$$

which means that the terms $\|\tilde{x}_i(t) - \tilde{s}(t)\|, f(\tilde{s}(t)) - f^*$ and $|f(\tilde{x}_i(t)) - f^*|$ will all decay to zero at a rate of $O\left(\frac{\ln t}{\sqrt{t}}\right)$. This result for the convergence rate is very common and can also be found in the recent works [22] and [31].
IV. PROOFS OF MAIN THEOREMS

A. Road Map for Proofs of Main Theorems

To begin with, we briefly sketch the main ideas of the proofs of Theorems 1 and 2.

1) First, we will start the analysis in Lemmas 7 and 8 by discussing that under Assumptions 2 and 3, the auxiliary variables \( y_i(t) \) can track the global gradient terms \( \alpha(t)\mu^i(t-1) \sum_{i=1}^{N} \nabla f_i(x_i(t)) \).

2) Second, to provide some preparations for the subsequent analysis, we will argue in Remark 4 that \( k_i(t) \) in the proposed distributed discrete-time algorithm (5) can be uniformly bounded with respect to \( t \) under Assumption 1.

3) Third, we will preliminarily describe the consensus errors in Lemma 9, based on which we will further characterize the gradient error terms \( \|y_i(t) - \alpha(t)\mu^i(t-1) \sum_{i=1}^{N} \nabla f_i(x_i(t))\| \) in Lemma 10.

4) Fourth, we will analyze the evolution of all states \( x_i(t) \) in Lemmas 11 and 12 and show that, as a consequence of all states reaching consensus, the iteration (5) can be approximately seen as the ordinary iteration for solving the problem (7) and all states \( x_i(t) \) will converge to a common optimal solution of the considered problem (4).

5) Finally, with all these preparatory works, we will be ready to give the detailed proofs of our main Theorems 1 and 2.

In addition, all the proofs of several aforementioned intermediate results (including Lemmas 7–12 and Corollary 1) will be collected in the Appendix so as to maintain a smooth presentation flow.

B. Preparatory Works

First, we define the following two intermediate variables:

\[
\tilde{y}(t) = \sum_{i=1}^{N} y_i(t), \quad \tilde{f}(t) = \sum_{i=1}^{N} f_i(x_i(t)).
\]

Motivated by the results in [22], [35] and [36], we will preliminarily characterize the error between the auxiliary variable \( y_i(t) \) and the global gradient term \( \alpha(t)\mu^i(t-1) \sum_{i=1}^{N} \nabla f_i(x_i(t)) \) in the following Lemmas 7 and 8.

**Lemma 7:** For the algorithm (5) with given initial values and all \( t \geq 0 \), there holds

\[
\tilde{y}(t) = \alpha(t)\nabla \tilde{f}(t).
\]

Note that with Lemma 7 we have

\[
\alpha(t)\mu^i(t-1) \sum_{i=1}^{N} \nabla f_i(x_i(t)) = \mu^i(t-1)\tilde{y}(t).
\]

**Lemma 8:** Let Assumptions 1, 2, 3, and 4 hold. Then, for the algorithm (5), the following inequality holds:

\[
\|y_i(t) - \mu^i(t-1)\tilde{y}(t)\| \leq C_3\lambda_2^{t-1} + C_4 \sum_{s=0}^{t-1} \lambda_2^{t-1-s}\alpha(s),
\]

where \( C_3 = C_2 \sum_{j=1}^{N} \|y_j(0)\| \) and \( C_4 = 2NMC'_4 \) with \( C'_4 = \max\{\frac{C_2}{\lambda_2}, 1\} \).

Then, based on Lemmas 7 and 8, we give the following corollary where a bound is introduced for the variable \( y_i(t) \).
**Corollary 1:** Let Assumptions 1, 2, 3 and 4 hold. Then, for the algorithm (5), the following inequality holds:

\[ \|y_i(t)\| \leq C_3 t^{-1} + C_4 \sum_{s=0}^{t-1} t^{-1-s} \alpha(s) + N M \alpha(t). \]  

Furthermore, there holds \( \lim_{t \to \infty} \|y_i(t)\| = 0. \)

**Remark 4:** From the result in Corollary 1 and the iteration rule (5a), it is clearly seen that \( v_i(t) \) is uniformly bounded with respect to \( t \). Thus, under Assumption 2, \( g_i(v_i(t)) \) and \( \nabla g_i(v_i(t)) \) are also uniformly bounded. Let

\[ G_i = \{ x \mid \nabla g_i(x) = 0 \}, \quad G_{i0} = \{ x \mid g_i(x) < 0 \}. \]

Since \( g_i(x) \) is convex, \( G_i \) is a closed and convex set and \( G_{i0} \) is an open set. Moreover, under Assumption 1 there holds \( \min \{ \inf x_i \} < 0 \) for all \( i = 1, 2, \cdots, N \), which implies that \( G_{i0} \) contains \( G_i \) strictly. Then, noting that \( \nabla g_i(x) \) is continuous, we can conclude that the set \( \{ \| \nabla g_i(x) \| \mid x \notin G_{i0} \} \) has a strictly positive infimum, which means that there exists a positive constant \( M \leq \|d_0\| \) such that \( \| \nabla g_i(x) \| \geq M \) holds for all \( x \notin G_{i0} \), i.e., for all \( t \geq 0 \), there holds \( \|d_i(t)\| \geq M \). Taking into account that \( g_i(x_i(t)) \) and \( \nabla g_i(x_i(t)) \) are also uniformly bounded, we have that \( d_i(t) \) and \( k_i(t) \) are also uniformly bounded. Hence, for convenience of the subsequent analysis, we can assume that for all \( i = 1, 2, \cdots, N \),

\[ \max \{ \|g_i(v_i(t))\|, \|\nabla g_i(v_i(t))\|, \|d_i(t)\|, \|k_i(t)\| \} \leq M. \]

Next, to further characterize the error term \( \|y_i(t) - \mu^i(t-1)\tilde{y}(t)\| \), we introduce another two new variables:

\[ w_1(t) = \sum_{i=1}^{N} \|\phi_i(t)\|, \quad w_2(t) = \sum_{i=1}^{N} g_i^+(z_i(t)), \]

with \( \phi_i(t) = x_i(t+1) - (v_i(t) - \beta k_i(t)) \). Besides, define

\[ \gamma_i(t) = \alpha(t) \sum_{s=0}^{t-1} t^{-1-s} w_i(s), \quad i = 1, 2, \]

with \( \gamma_1(0) = \gamma_2(0) = 0 \). In the following lemma, we will deduce a common bound for all the terms \( \alpha(t)\|x_i(t) - \bar{x}(t)\| \).

**Lemma 9:** Let Assumptions 1, 2, 3 and 4 hold. Then, for the algorithm (5) and all \( i = 1, 2, \cdots, N \), there holds

\[ \alpha(t)\|x_i(t) - \bar{x}(t)\| \leq C_6 \gamma_1(t) + C_7 \gamma_2(t) + \xi_1(t), \]  

where \( C_6 = 2 N^3 C_1 \), \( C_7 = 2 N^2 C_1 \frac{\beta}{M} \), and

\[ \xi_1(t) = C_8 \alpha(t) \sum_{s=1}^{t-1} t^{-1-s} t^s - 1 + C_{10} \alpha(t) \sum_{s=1}^{t-1} t^{-1-s} \alpha(s) \]

\[ + C_5 \alpha(t) t^s - 1 + C_9 \alpha(t) \sum_{s=1}^{t-1} t^{-1-s} \sum_{r=0}^{s-1} t^r \alpha(r), \]

with

\[ C_5 = N C_1 \sum_{j=1}^{N} \|x_j(0) - \bar{x}(0)\| + 2 N^2 C_1 (w_1(0) + \sum_{j=1}^{N} y_j(0) + \beta \sum_{j=1}^{N} \|k_j(0)\|), \]

\[ C_8 = 2 N^2 C_1 C_3 \left( 1 + \frac{\beta M}{M} \right), \quad C_9 = 2 N^2 C_1 C_4 \left( 1 + \frac{\beta M}{M} \right), \quad C_{10} = 2 N^3 M C_1 \left( 1 + \frac{\beta M}{M} \right). \]
Based on Lemma 9 we will provide a more concrete bound of the term \( \| y_i(t) - \mu^i(t-1)\hat{y}(t) \| \) in the next lemma. To this end, we need to introduce several intermediate variables:

\[
\eta_1(t) = \sum_{s=0}^{t-1} \lambda_2^{-1-s} \gamma_1(s), \quad i = 1, 2, \quad \eta_2(t) = \sum_{s=0}^{t-1} \lambda_2^{-1-s} \alpha(s)w_2(s),
\]

with \( \eta_1(0) = \eta_2(0) = \eta_3(0) = 0. \)

**Lemma 10:** Let Assumptions 1, 2, 3, and 4 hold. Then, for the algorithm (5) and all \( i = 1, 2, \cdots, N \), the following inequality holds:

\[
\| y_i(t) - \mu^i(t-1)\hat{y}(t) \| 
\leq C_3 \lambda_2^{-1} + C_{17} \sum_{s=0}^{t-1} \lambda_2^{-1-s} \xi_2(s) + C_{18} \eta_1(t) + C_{19} \eta_2(t) + C_{20} \eta_3(t),
\]

where

\[
\xi_2(t) = M(\alpha(t) - \alpha(t+1)) + L(1+N)\xi_1(t) + C_{13} \alpha(t) \lambda_2^{-1} + C_{14} \alpha(t) \sum_{s=0}^{t-1} \lambda_2^{-1-s} \alpha(s) + C_{15} \alpha^2(t),
\]

with \( C_{11} = L(1+N)C_6, C_{12} = L(1+N)C_7, \)

\[
C_{13} = L\left(1 + \frac{\beta M}{M}\right) C_3, \quad C_{14} = L\left(1 + \frac{\beta M}{M}\right) C_4, \quad C_{15} = LNM\left(1 + \frac{\beta M}{M}\right),
\]

\[
C_{16} = \frac{\beta L}{M}, \quad C_{17} = NC_{14}', \quad C_{18} = NC_{11}C_4', \quad C_{19} = NC_{12}C_4', \quad C_{20} = NC_{16}C_4'.
\]

Now, we are prepared to examine the evolution of all states \( x_i(t) \) in the below lemma.

**Lemma 11:** Under Assumptions 1, 2, 3, and 4 for all \( v \in X^* \) and \( t \geq 0 \), there holds

\[
\sum_{i=1}^{N} \pi^i(t+1) \| x_i(t+1) - v \|^2 
\leq \sum_{i=1}^{N} \pi^i(t) \| x_i(t) - v \|^2 - 2\theta_2 \alpha(t) (f(s(t)) - f(v)) + \xi_3(t) + D_4 \gamma_1(t) + D_5 \gamma_2(t)
\]

\[
- \sum_{i=1}^{N} \frac{\theta_1 \beta (2 - \beta)}{2M^2} (g_i^\ast(z_i(t)))^2 + D_6 \eta_1(t) + D_7 \eta_2(t) + D_8 \eta_3(t) - \sum_{i=1}^{N} \theta_1 \| \phi_i(t) \|^2,
\]

where \( D_4 = 2C_6D_1N, D_5 = 2C_7D_1N, D_6 = C_{18}D_3N, D_7 = C_{19}D_3N, D_8 = C_{20}D_3N, \) and

\[
\xi_3(t) = 2D_1 N \xi_1(t) + D_2 \alpha^2(t) + C_3 D_3 N \lambda_2^{-1} + C_1 \gamma_3(t) \sum_{s=0}^{t-1} \lambda_2^{-1-s} \xi_2(s),
\]

with \( D_1 = MN + LM + (MN+1)MNR, D_3 = 2M + 8M \beta (2 - \beta), \) and

\[
D_2 = \frac{4M^4 N^3 R^2}{\theta_1 \beta (2 - \beta)} + 2M^2 N^2 + 8M^2 N^2 \beta (2 - \beta).
\]

It should be noted that in Lemma 11 the summability of \( \gamma_i(t), i = 1, 2, \) and \( \eta_j(t), j = 1, 2, 3, \) is still unknown for us. Hence, these terms need to be further analyzed. Some properties of \( \gamma_i(t), i = 1, 2, \) and \( \eta_j(t), j = 1, 2, 3, \) will be revealed in the following proposition whose proof is omitted for brevity.
**Proposition 2:** For all \( t \geq 0 \), the below inequalities hold:

\[
\begin{align*}
\gamma_i(t + 1) &\leq \lambda_1 \gamma_i(t) + \alpha(t) w_i(t), \quad i = 1, 2, \\
\eta_i(t + 1) &\leq \lambda_2 \eta_i(t) + \gamma_i(t), \quad i = 1, 2, \\
\eta_3(t + 1) &\leq \lambda_2 \eta_3(t) + \alpha(t) w_2(t).
\end{align*}
\]  

(17a) \hspace{1cm} (17b) \hspace{1cm} (17c)

To continue, based on the results in Proposition 2 and 3, we give another key lemma to further characterize the evolution of all states \( x_i(t) \).

**Lemma 12:** Under Assumptions 1, 2, 3, and 4, for all \( v \in X^* \) and \( t \geq 0 \), there holds

\[
\begin{align*}
\sum_{i=1}^{N} \pi_i(t + 1) ||x_i(t + 1) - v||^2 &+ e(t + 1) \\
\leq & \sum_{i=1}^{N} \pi_i(t) ||x_i(t) - v||^2 + e(t) + 2\theta_2 \alpha(t) (f(v) - f(s(t))) \\
&+ \xi_4(t) - \frac{\theta_1}{2} \sum_{i=1}^{N} ||\phi_i(t)||^2 - \frac{\theta_1 \beta(2 - \beta)}{4M^2} \sum_{i=1}^{N} (g_i^+(z_i(t)))^2,
\end{align*}
\]

(18)

where \( e(t) = \sum_{i=1}^{2} (a_i b_i \gamma_i(t) + c_i \eta_i(t)) + a_3 \beta \eta_3(t) \) and \( \xi_4(t) = \xi_3(t) + \frac{1}{2} (a_1^2 + a_2^2 + a_3^2) \alpha^2(t) \) with

\[
\begin{align*}
a_i &= \frac{c_i + D_{3+i}}{(1 - \lambda_2) b_i}, \quad i = 1, 2, \quad a_3 = \frac{D_8}{(1 - \lambda_2) b_3}, \quad b_1 = \sqrt{\frac{\theta_1}{N}}, \\
b_2 = b_3 &= \sqrt{\frac{\theta_1 \beta(2 - \beta)}{4M^2 N}}, \quad c_i = \frac{D_{3+i}}{1 - \lambda_2}, \quad i = 1, 2.
\end{align*}
\]

C. Proof of Theorem 7

In this subsection, based on the above intermediate results and the proof of Theorem 4, we are now ready to present the detailed proof of Theorem 1 regarding the convergence property of the algorithm (5).

**Proof of Theorem 7** First, the nonnegative sequences \( \{a(t)\} \), \( \{b(t)\} \), \( \{c(t)\} \), and \( \{d(t)\} \) in Lemma 3 can be defined as follows:

\[
\begin{align*}
a(t) &= \sum_{i=1}^{N} \pi_i(t) ||x_i(t) - v||^2 + e(t), \quad b(t) = 0, \quad c(t) = \xi_4(t), \\
d(t) &= 2\theta_2 \alpha(t) (f(s(t)) - f(v)) + \frac{\theta_1}{2} \sum_{i=1}^{N} ||\phi_i(t)||^2 + \frac{\theta_1 \beta(2 - \beta)}{4M^2} \sum_{i=1}^{N} (g_i^+(z_i(t)))^2.
\end{align*}
\]

Then, noting that \( \alpha^2(t) \) is summable, we can easily obtain from part (b) of Lemma 2 that \( \xi_1(t) \) is summable. In a similar manner, we can get that \( \xi_2(t) \) and \( \xi_3(t) \) are also summable, which directly implies the summability of \( \xi_4(t) \) (i.e., \( c(t) \)). So by Lemma 3 there exists a nonnegative constant \( \delta_0 \) such that

\[
\lim_{t \to \infty} \left( \sum_{i=1}^{N} \pi_i(t) ||x_i(t) - v||^2 + e(t) \right) = \delta_0,
\]

(19)

\[
\sum_{t=0}^{\infty} \alpha(t) (f(s(t)) - f(v)) < \infty,
\]

(20)
As a result, we attain
\[ \sum_{t=0}^{\infty} \left[ \frac{\theta_1}{2} \sum_{i=1}^{N} \| \phi_i(t) \|^2 + \frac{\theta_1 \beta(2 - \beta)}{4M^2} \sum_{i=1}^{N} (g_i^+(z_i(t)))^2 \right] < \infty. \] (21)

Clearly, it follows from (21) that
\[ \lim_{t \to \infty} \sum_{i=1}^{N} \| \phi_i(t) \|^2 = 0, \quad \lim_{t \to \infty} \sum_{i=1}^{N} (g_i^+(z_i(t)))^2 = 0, \] (22)

which yields that \( \lim_{t \to \infty} \gamma_i(t) = 0, \quad i = 1, 2, \) and \( \lim_{t \to \infty} \eta_i(t) = 0, \quad i = 1, 2, 3, \) according to part (a) of Lemma 2. Thus, we have \( \lim_{t \to \infty} c(t) = 0, \) and from (55) to be given in the proof of Lemma 9 in the Appendix, we can conclude that
\[ \lim_{t \to \infty} \| x_i(t) - \bar{x}(t) \| = 0. \] (23)

Furthermore, considering (19), we have
\[ \lim_{t \to \infty} \sum_{i=1}^{N} \xi_i(t) = \delta_0. \]

Given that \( \sum_{i=1}^{\infty} \alpha(t) = \infty, \) we can get from (20) that \( \liminf_{t \to \infty} f(s(t)) = f(v). \) As \( \{s(t)\} \) is contained in the compact set \( X, \) it has a convergent subsequence \( \{s(t)\}, \) that is, there exists \( s^* \) such that \( \lim_{k \to \infty} s(t_k) = s^*. \) Moreover, since \( X_0 \) is a closed set, we have \( s^* \in X_0. \) Thus, it follows from the continuity property of \( f \) that \( \lim_{t \to \infty} f(s(t_k)) = f(s^*) = f(v), \) which implies that \( s^* \in X^*. \) Since \( v \) is arbitrarily chosen in \( X^* \), we can substitute \( v \) by \( s^*. \) Then, it can be verified that \( \delta_0 = 0 \) when \( v = s^*. \) So we have
\[ \| x_i(t) - s^* \|^2 \leq 3(\| x_i(t) - \bar{x}(t) \|^2 + \| \bar{x}(t) - s(t) \|^2 + \| s(t) - s^* \|^2). \]

Utilizing (66) to be given in the proof of Lemma 11 in the Appendix, we arrive at
\[ \| \bar{x}(t) - s(t) \|^2 \leq 2R^2(NM + 1)^2 N \sum_{i=1}^{N} \| x_i(t) - \bar{x}(t) \|^2 + 2R^2 N \sum_{j=1}^{N} (g_j^+(z_j(t)))^2. \]

As a result, we attain
\[ \frac{1}{3} \| x_i(t) - s^* \|^2 \leq \| x_i(t) - \bar{x}(t) \|^2 + 2R^2(NM + 1)^2 N \sum_{i=1}^{N} \| x_i(t) - \bar{x}(t) \|^2 + 2R^2 N \sum_{i=1}^{N} (g_i^+(z_i(t)))^2 + \| s(t) - s^* \|^2. \] (24)

Multiplying both sides of the above inequality (24) by \( \pi_i'(t) \) and summing over from \( i = 1 \) to \( i = N \) produces
\[ \sum_{i=1}^{N} \frac{\pi_i'(t)}{3} \| x_i(t) - s^* \|^2 \leq \sum_{i=1}^{N} \pi_i'(t) \| x_i(t) - \bar{x}(t) \|^2 + \sum_{i=1}^{N} \pi_i'(t) \| s(t) - s^* \|^2 + 2R^2(M + 1)^2 N \sum_{i=1}^{N} \pi_i'(t) \sum_{i=1}^{N} \| x_i(t) - \bar{x}(t) \|^2 + 2R^2 N \sum_{i=1}^{N} \pi_i'(t) \sum_{i=1}^{N} (g_i^+(z_i(t)))^2. \] (25)

Then, taking \( \liminf \) for both sides of (25) and using (22) as well as (23) yields \( \frac{\pi_1}{3} \leq \liminf_{t \to \infty} \| s(t) - s^* \|. \)

Finally, since \( \lim_{k \to \infty} s(t_k) = s^* \), we have
\[ \liminf_{t \to \infty} \| s(t) - s^* \| = 0, \]
which implies that $\delta_0 = 0$. Noting that $\pi^i(t) \geq \theta_1 > 0$ for all $i = 1, \ldots, N$, and $t \geq 0$, we can get that 
$$\lim_{t \to \infty} x_i(t) = s^* \quad \text{for all } i = 1, \ldots, N,$$ 
thus establishing the convergence of the algorithm (5). The proof is complete.

\[ \square \]

\textbf{D. Proof of Theorem 2}

In this subsection, with the inspirations from the analysis of the convergence rate shown in [22] and [31] and the above intermediate results in Subsection IV-B, we will give the detailed proof of Theorem 2 about the convergence rate of the algorithm (5).

\textbf{Proof of Theorem 2}: First, according to (18), by defining 
$$\varphi_1(t) = \frac{\theta_1}{4\theta_2} \sum_{i=1}^{N} \|\phi_i(t)\|^2, \quad \varphi_2(t) = \frac{\theta_1\beta(2-\beta)}{8\theta_2M^2} \sum_{i=1}^{N} (g_i^+(z_i(t)))^2,$$
we have 
$$\sum_{k=0}^{t} \left[ \alpha(k)(f(s(k)) - f^*) + \varphi_1(k) + \varphi_2(k) \right] \leq \frac{1}{2\theta_2} \left( \sum_{i=1}^{N} \pi^i(0)\|x_i(0) - v\|^2 + \epsilon(0) + \sum_{k=0}^{t} \xi_4(k) \right). \quad (26)$$

Obviously, for any nonnegative summable scalar sequence \{h(t)\} and any constant $0 < \rho < 1$, there holds 
$$\sum_{k=0}^{t} \sum_{s=0}^{k} \rho^{k-s} h(s) = \sum_{s=0}^{t} h(s) \sum_{k=s}^{t} \rho^{k-s} \leq \sum_{k=0}^{t} \frac{h(k)}{1-\rho}, \quad (27)$$
which implies that 
$$\sum_{k=0}^{t} \xi_1(k) \leq \frac{C_8\alpha(0)}{(1-\lambda_1)(1-\lambda_2)} + \frac{C_{10}}{1-\lambda_1} \sum_{k=0}^{t} \alpha^2(k)$$
$$+ \frac{C_5\alpha(0)}{1-\lambda_1} + \frac{C_9}{(1-\lambda_1)(1-\lambda_2)} \sum_{k=0}^{t} \alpha^2(k)$$
$$= M_1 + M_2 \sum_{k=0}^{t} \alpha^2(k), \quad (28)$$

$$\sum_{k=0}^{t} \xi_2(k) \leq M_1 + M_2 \sum_{k=0}^{t} \alpha^2(k) + L(1+N) \left( M_1 + M_2 \sum_{k=0}^{t} \alpha^2(k) \right)$$
$$+ \frac{C_{13}\alpha(0)}{1-\lambda_2} + \frac{C_{14}}{1-\lambda_2} \sum_{k=0}^{t} \alpha^2(k) + \sum_{k=0}^{t} \alpha^2(k)$$
$$= M_3 + M_4 \sum_{k=0}^{t} \alpha^2(k), \quad (29)$$

$$\sum_{k=0}^{t} \xi_3(k) \leq 2ND_1 \left( M_1 + M_2 \sum_{k=0}^{t} \alpha^2(k) \right) + D_2 \sum_{k=0}^{t} \alpha^2(k)$$
$$+ \frac{C_3D_3\alpha(0)}{1-\lambda_2} + \frac{C_{17}D_3\alpha(0)}{1-\lambda_2} \left( M_3 + M_4 \sum_{k=0}^{t} \alpha^2(k) \right)$$
$$= M_5 + M_6 \sum_{k=0}^{t} \alpha^2(k), \quad (30)$$
\[ \sum_{k=0}^{t} \xi_4(k) \leq M_5 + M_7 \sum_{k=0}^{t} \alpha^2(k), \]

where

\[
M_1 = \frac{D_8\alpha(0)}{1 - \lambda_1(1 - \lambda_2)} + \frac{C_5\alpha(0)}{1 - \lambda_2}, \quad M_2 = \frac{C_{10}}{1 - \lambda_1} + \frac{C_9}{(1 - \lambda_1)(1 - \lambda_2)},
\]

\[
M_3 = M\alpha(0) + L(1 + N)M_1 + \frac{C_{13}\alpha(0)}{1 - \lambda_2}, \quad M_4 = L(1 + N)M_2 + \frac{C_{14}}{1 - \lambda_2} + C_{15},
\]

\[
M_5 = 2D_1 M_1 N + \frac{C_5 D_3 N}{1 - \lambda_2} + \frac{C_{17} D_3 M_3 N}{1 - \lambda_2}, \quad M_6 = 2D_1 M_2 N + D_2 + \frac{C_{17} D_3 M_4 N}{1 - \lambda_2},
\]

\[
M_7 = M_6 + \frac{1}{2}(\alpha_1 + \alpha_2 + \alpha_3).
\]

Thus, it follows from (26) that

\[
\sum_{k=0}^{t} \alpha(k)(f(s(k)) - f^*) + \varphi_1(k) + \varphi_2(k) \leq M_8 + M_9 \sum_{k=0}^{t} \alpha^2(k), \tag{32}
\]

with

\[
M_8 = \frac{1}{2\theta_2} \left( \sum_{i=1}^{N} \pi^i(t) \| x_i(0) - v \|^2 + e(0) + M_3 \right), \quad M_9 = \frac{M_7}{2\theta_2}.
\]

Additionally, according to (66) to be given in the proof of Lemma 11 in the Appendix, we can get

\[
\| x_i(t) - s(t) \|
\leq \| x_i(t) - \bar{x}(t) \| + \| \bar{x}(t) - s(t) \|
\leq \| x_i(t) - \bar{x}(t) \| + R \sum_{i=1}^{N} g_i^+(z_i(t)) + R(MN + 1) \sum_{i=1}^{N} \| x_i(t) - \bar{x}(t) \|. \tag{33}
\]

Thus, by Lemma 9, we have

\[
\frac{\alpha(t)\| x_i(t) - s(t) \|}{\kappa(RM^2N^2 + RN + 1)} \leq \gamma_1(t) + \gamma_2(t) + \frac{1}{\kappa} \xi_1(t) + \frac{R\alpha(t)w_2(t)}{\kappa(RM^2N^2 + RN + 1)}, \tag{34}
\]

with \( \kappa = \max\{C_6, C_7\} \). Hence, combining (26) and (34) together yields

\[
\frac{1}{\kappa(RM^2N^2 + RN + 1)} \sum_{k=0}^{t} \alpha(k)\| x_i(k) - s(k) \|
\leq \sum_{k=0}^{t} \gamma_1(k) + \sum_{k=0}^{t} \gamma_2(k) + \frac{1}{\kappa} \left( M_1 + M_2 \sum_{k=0}^{t} \alpha^2(k) \right)
+ \frac{R}{\kappa(RM^2N^2 + RN + 1)} \sum_{k=0}^{t} \alpha(k)w_2(t). \tag{35}
\]

Clearly, we have

\[
\sum_{k=0}^{t} \gamma_i(k) \leq \lambda_1 \sum_{k=0}^{t-1} \gamma_i(k) + \sum_{k=0}^{t-1} \alpha(k)w_i(k) \leq \lambda_1 \sum_{k=0}^{t} \gamma_i(k) + \sum_{k=0}^{t} \alpha(k)w_i(k), \tag{36}
\]

which implies that

\[
\sum_{k=0}^{t} \gamma_i(k) \leq \frac{1}{1 - \lambda_1} \sum_{k=0}^{t} \alpha(k)w_i(k) \leq \frac{1}{2(1 - \lambda_1)} \sum_{k=0}^{t} \alpha^2(k) + \frac{1}{2(1 - \lambda_1)} \sum_{k=0}^{t} w_i^2(k). \tag{37}
\]

Thus, it follows from (35) that
Furthermore, it is straightforward to deduce

\[
\frac{2(1 - \lambda_1)}{\kappa(RMN^2 + RN + 1)} \sum_{k=0}^{t} \alpha(k)\|x_i(k) - s(k)\| \\
\leq \sum_{k=0}^{t} \alpha^2(k) + \sum_{k=0}^{t} w_1^2(k) + \sum_{k=0}^{t} \alpha^2(k) + \sum_{k=0}^{t} w_2^2(k) + \sum_{k=0}^{t} w_2^2(k) \\
+ \frac{2(1 - \lambda_1)}{\kappa} \left( M_1 + M_2 \sum_{k=0}^{t} \alpha^2(k) \right) + \frac{R^2(1 - \lambda_1)^2}{\kappa^2(RMN^2 + RN + 1)^2} \sum_{k=0}^{t} \alpha^2(k). \tag{38}
\]

Now, introducing \( \psi_1(t) = \sum_{i=1}^{N} \|\phi_i(t)\|^2 \) and \( \psi_2(t) = \sum_{i=1}^{N} (q_i^+(z_i(t)))^2 \), and noting that \( w_1^2(t) \leq N^2 \psi_i(t), \ i = 1, 2 \), we obtain

\[
H_1 \sum_{k=0}^{t} \alpha(k)\|x_i(k) - s(k)\| \leq M_{10} + M_{11} \sum_{k=0}^{t} \alpha^2(k) + \delta(\psi_1(t) + \psi_2(t)), \tag{39}
\]

where \( \delta = \min \left\{ \frac{\delta_0}{2M^2}, \frac{\delta_1 \beta(2 - \beta)}{8M^2} \right\} \).

\[
H_1 = \frac{\delta(1 - \lambda_1)}{\kappa(RMN^2 + RN + 1)^N}, \quad M_{10} = \frac{(1 - \lambda_1)M_1 \delta}{\kappa N}, \\
M_{11} = \frac{\delta}{N} + \frac{(1 - \lambda_1)M_2 \delta}{\kappa N} + \frac{R^2(1 - \lambda_1)^2 \delta}{2\kappa^2(RMN^2 + RN + 1)^2 N}.
\]

Then, it follows from (32) and (39) that

\[
\sum_{k=0}^{t} \left[H_1 \alpha(k)\|x_i(k) - s(k)\| + \alpha(k)(f(s(k)) - f^*)\right] \leq H_2 + H_3 \sum_{k=0}^{t} \alpha^2(k), \tag{40}
\]

where \( H_2 = M_8 + M_{10} \) and \( H_3 = M_9 + M_{11} \).

Since \( f(x) \) and the norm function are both convex, we have

\[
f(\tilde{s}(t)) - f^* \leq \sum_{k=0}^{t} \frac{\alpha(k)(f(s(k)) - f^*)}{\sum_{k=0}^{t} \alpha(k)}, \tag{41}
\]

\[
\|\tilde{x}_i(t) - \tilde{s}(t)\| \leq \sum_{k=0}^{t} \frac{\alpha(k)\|x_i(k) - s(k)\|}{\sum_{k=0}^{t} \alpha(k)}. \tag{42}
\]

Dividing both sides of (40) by \( \sum_{k=0}^{t} \alpha(k) \), we can get (9) from (41) and (42), with \( L(t) \) being as given in (10).

Furthermore, it is straightforward to deduce

\[
|f(\tilde{x}_i(t)) - f^*| \leq |f(\tilde{x}_i(t)) - f(\tilde{s}(t))| + |f(\tilde{s}(t)) - f^*|
\leq MN\|\tilde{x}_i(t) - \tilde{s}(t)\| + |f(\tilde{s}(t)) - f^*|
\leq MN\frac{H_1}{H_1}L(t) + L(t) = \left(\frac{MN}{H_1} + 1\right)L(t). \tag{43}
\]

We have thus completed the proof of the convergence rate.

\[\Box\]

V. COMPUTER SIMULATIONS

Similar to the numerical example shown in [31], we consider a constrained optimization problem involved in the machine learning problems with its global objective function defined on \( \mathbb{R}^3 \) as following:

\[
\min_{x \in \mathcal{X}} f(x) = \sum_{i=1}^{N} f_i(x), \tag{44a}
\]

DRAFT
where

\[ f_i(x) = \ln[1 + e^{-a_i(b_i^T u + x^3)}] + p_i(x), \]  

(44b)

with \( p_i(x) = \frac{(x_i^1)^2}{4} \) for \( i = 1, \ldots, r_1 - 1 \), \( p_i(x) = \frac{(x_i^2)^2}{4} \) for \( i = r_1, \ldots, N \), \( u = (x_1^1, x_2^1)^T \), \( b_i = (0.01i, 0.02i)^T \) are the feature vectors, and \( a_i = (-1)^i \) are the corresponding labels.

**Case A:** \( N = 8 \). In order to clearly show the effectiveness of the algorithm (5) in the case involving the time-varying unbalanced graphs with only the union of graphs being strongly connected, we first carry out numerical simulation when \( N = 8 \) and \( r_1 = 5 \). Additionally, for \( i = 1, 2, \cdots, N \), the inequality constraints \( g_i(x) \) are selected as

\[ g_i(x) = (x_1^1)^2 + ix^2 + x^3 - 10 \leq 0, \]  

(45)

and the closed convex sets \( X_i \) are chosen as

\[ X_i = \left[ \frac{i}{2} - 3, \frac{i}{2} + 1 \right] \times \left[ \frac{i}{2} - 3.5, \frac{i}{2} + 0.5 \right] \times \left[ \frac{i}{2} - 1, \frac{i}{2} + 2.5 \right]. \]

Thus, \( X = \bigcap_{i=1}^{N} X_i = [1, 2] \times [0.5, 1] \times \{3\} \). To describe the time-varying unbalanced connections between eight nodes, four graphs with each having eight interacting agents are depicted in Fig. 1 in which only the union of four graphs is strongly connected. Also, it is stipulated that the communication graph \( G(t) \) at step \( t \) is respectively selected as Fig. 1a, Fig. 1b, Fig. 1c, and Fig. 1d when \( \text{mod} (t, 4) = 1 \), \( \text{mod} (t, 4) = 2 \), \( \text{mod} (t, 4) = 3 \), and \( \text{mod} (k, 4) = 0 \). Moreover, the step-sizes are chosen as \( \alpha(t) = \frac{10^{-3}}{(t+1)^{\alpha_0}} \) and \( \beta = 1 \) for simulation. The transient behaviors of all states \( x_i(t) \) are displayed in Fig. 2, which clearly shows that all states \( x_i(t) \) under the algorithm (5) converge to the common optimal solution \((1, 0.5, 3)^T\).

**Case A’:** \( N = 8 \). We now compare the algorithm (5) with Algorithm 1 developed in [31]. Case A’ is the same as Case A except for two changes that a strongly connected unbalanced graph as depicted in Fig. 3 is involved instead and the inequality constraints in (45) are removed. This is because Algorithm 1 in [31] is applicable only to the optimization problem subject to nonidentical local closed convex set constraints and under the strongly connected unbalanced graph. Moreover, all \( x_i(t) \) under Algorithm 1 in [31] also converge to the common optimal solution \((1, 0.5, 3)^T\) within the setting as given in Case A’. Note that the algorithm (5) is still implemented under the same setting as described in Case A. The behaviors of the convergence criterion \( \frac{1}{N} \sum_{i=1}^{N} \frac{\|x_i(t) - x^*\|}{\|x^*\|} \) under the algorithm (5), in Case A and Algorithm 1 in [31] in Case A’ are plotted in Fig. 4. It can be observed from Fig. 4 that these...
two algorithms exhibit a similar convergence rate although the extra inequality constraints in (45) and the time-varying unbalanced graphs in Fig. 1 are involved in the numerical simulation of Case A for the algorithm (5). This observation agrees well with the discussion in Remark 1.

Case B: $N = 100$. In order to verify the effectiveness of the algorithm (5) in large-scale optimization problems, we reconsider the objective functions in (44) but with $N = 100$ and $r_1 = 51$. Furthermore, for $i = 1, 2, \cdots, N$, the inequality constraints are reselected as

$$g_i(x) = (x_1)^2 + 0.1ix^2 + x^3 - 10 \leq 0$$

and the closed convex sets are reselected as

$$X_i = [0.06i - 5, 0.06i + 1.94] \times [0.06i - 5.5, 0.06i + 0.94] \times [0.06i - 3, 0.06i + 2.94].$$

Thus, $X = \bigcap_{i=1}^{N} X_i = [1, 2] \times [0.5, 1] \times \{3\}$. Besides, the time-varying unbalanced directed graphs sequence with $N = 100$ nodes is considered here, where the connections between nodes are time-varying in the sense that the
Fig. 5. Behaviors of the state $x_i(t)$ under the algorithm (5) in Case B.

associated iterative matrices $A(t)$ and $B(t)$ are randomly selected at each step $t$. With the same step-sizes $\alpha(t)$ and $\beta$ as chosen in Case A, the transient behaviors of all states $x_i(t)$ under the algorithm (5) are illustrated in Fig. 5, from which it is clearly seen that all states $x_i(t)$ also converge to the common optimal solution $x^* = (1, 0.5, 3)^T$.

VI. CONCLUSION

In this paper, the optimization problem with nonidentical local convex inequality constraints and compact set constraints has been studied. Specifically, a distributed discrete-time algorithm has been developed over time-varying unbalanced directed topologies and its convergence property to the optimal solution has been rigorously confirmed. Moreover, the detailed analysis of the convergence rate has also been clearly shown for the proposed distributed algorithm. The advantage of the proposed algorithm is that the auxiliary variables have been introduced to estimate the gradients of the global objective function so as to offset the effect of the network-induced asymmetry. As a result, the optimization problem with multiple nonidentical local constraints has been successfully resolved over time-varying unbalanced topologies. More importantly, the work in this paper has shown an efficient mechanism for designing distributed algorithms for optimization problems with various types of constraints over time-varying unbalanced topologies. In the future, attention will be paid to addressing the distributed optimization problem formulated in this paper with a view to achieving a better convergence rate in the case strongly convex objective functions.

APPENDIX

In this Appendix, the proofs of all the intermediate results including Lemmas 7–12 and Corollary 1 are shown in detail.

Proof of Lemma 7 From the iteration rule (5c) and the definition of $\tilde{y}(t)$, we have

$$\tilde{y}(t+1) = \sum_{i=1}^{N} \sum_{j=1}^{N} B_{ij}(t)y_j(t) + \alpha(t+1) \sum_{i=1}^{N} \nabla f_i(x_i(t+1)) - \alpha(t) \sum_{i=1}^{N} \nabla f_i(x_i(t)).$$

Considering that $B(t)$ are column stochastic matrices for all $t \geq 0$, we can get

$$\tilde{y}(t+1) = \tilde{y}(t) + \alpha(t+1)\nabla \tilde{f}(t+1) - \alpha(t)\nabla \tilde{f}(t).$$

Then, direct computation can yield

$$\tilde{y}(t+1) - \alpha(t+1)\nabla \tilde{f}(t+1) = \tilde{y}(t) - \alpha(t)\nabla \tilde{f}(t),$$

DRAFT
As a result, we can obtain
\[\hat{y}(t) - \alpha(t)\nabla \hat{f}(t) = \hat{y}(0) - \alpha(0)\nabla \hat{f}(0).\]

Therefore, with the given initial values, we have
\[\hat{y}(t) - \alpha(t)\nabla \hat{f}(t) = \hat{y}(0) - \alpha(0)\nabla \hat{f}(0) = 0,\]
which completes the proof.

**Proof of Lemma 8** First, we can rewrite (5c) as
\[y_i(t + 1) = \sum_{j=1}^{N} B_{ij}(t)y_j(t) + \epsilon_i(t),\]
with \(\epsilon_i(t) = (t + 1)\nabla f_i(x_i(t + 1)) - \alpha(t)\nabla f_i(x_i(t))\) being bounded by \(2\alpha(t)M\). Then, we have
\[y_i(t) = \sum_{j=1}^{N} B(t - 1 : 0)_{ij}y_j(0) + \epsilon_i(t - 1) + \sum_{s=1}^{t-1} \sum_{j=1}^{N} B(t - 1 : s)_{ij}\epsilon_j(s - 1),\]
\[\hat{y}(t) = \hat{y}(0) + \sum_{s=0}^{t-1} \sum_{j=1}^{N} \epsilon_j(s).\]
As a result, we can obtain
\[y_i(t) - \mu^i(t - 1)\hat{y}(t) = \sum_{j=1}^{N} D(t - 1 : 0)_{ij}y_j(0) + \sum_{s=1}^{t-1} \sum_{j=1}^{N} D(t - 1 : s)_{ij}\epsilon_j(s - 1)\]
\[+ \epsilon_i(t - 1) - \mu^i(t - 1) \sum_{j=1}^{N} \epsilon_j(t - 1),\]
where \(D(t : s) = B(t : s) - \mu(t)1^T\). Next, from part (a) of Lemma 5 we can derive
\[\|y_i(t) - \mu^i(t - 1)\hat{y}(t)\| \leq \sum_{j=1}^{N} C_2\lambda_2^{j-1}\|y_j(0)\| + \sum_{j=1}^{N} \|\epsilon_j(t - 1)\|\]
\[+ \sum_{s=1}^{t-1} \sum_{j=1}^{N} C_2\lambda_2^{j-1-s}\|\epsilon_j(s - 1)\|\]
\[\leq C_3\lambda_2^{t-1} + C_4\sum_{s=0}^{t-1} \sum_{j=1}^{N} \lambda_2^{j-1-s}\|\epsilon_j(s)\|,\]
which implies (12). Thus, the proof is complete.

**Proof of Corollary 1** According to Lemma 7 we can directly get
\[\|\hat{y}(t)\| \leq N M \alpha(t).\]
Observing that \(\mu^i(t) < 1\) and
\[\|y_i(t)\| \leq \|y_i(t) - \mu^i(t - 1)\hat{y}(t)\| + \mu^i(t - 1)\|\hat{y}(t)\|,\]
we can easily deduce (13) from (12) and (47).
Clearly, the first term $C_3 \lambda_{2}^{t-1}$ and the third term $N M \alpha(t)$ on the right-hand side of (13) converge to zero when $t$ tends to infinity. Therefore, noting from part (a) of Lemma 2 that
\[
\lim_{t \to \infty} C_4 \sum_{s=0}^{\ell-1} \lambda_{2}^{t-1-s} \alpha(s) = 0,
\]
we have that $\lim_{t \to \infty} \|y_i(t)\| = 0$, which completes the proof. 

**Proof of Lemma 9** First, with $u_i(t) = \phi_i(t) - y_i(t) - \beta k_i(t)$, we rewrite (56) as
\[
x_i(t+1) = \sum_{j=1}^{N} A_{ij}(t)x_j(t) + u_i(t),
\]
which corresponds to the form formulated in [32, Lemma 6]. Thus, from the proof of [32, Lemma 6], we can directly obtain
\[
\|x_i(t) - \bar{x}(t)\| \leq \sum_{j=1}^{N} NC_1 \lambda_{1}^{t-1} \|x_j(0) - \bar{x}(0)\| + \sum_{s=0}^{t-1} \sum_{j=1}^{N} NC_1 \lambda_{1}^{t-1-s} \|u_j(s) - \hat{u}(s)\|,
\]
where $\hat{u}(t) = \sum_{j=1}^{N} \pi^j(t+1) u_j(t)$. Then, for the term $\|u_j(s) - \hat{u}(s)\|$, we have
\[
\|u_j(s) - \hat{u}(s)\| \leq \|u_j(s)\| + \left\| \sum_{j=1}^{N} \pi^j(s+1) u_j(s) \right\|
\leq \sum_{j=1}^{N} \|u_j(s)\| + \sum_{j=1}^{N} \pi^j(s+1) \|u_j(s)\|
\leq 2 \sum_{j=1}^{N} \|u_j(s)\|,
\]
where the second inequality is deduced from the convex property of the norm function and the last inequality can be obtained from the fact that $0 < \pi^i(t) < 1$ for all $i = 1, 2, \cdots, N$ and $t \geq 0$. Moreover, considering that
\[
\|u_i(s)\| \leq \|\phi_i(s)\| + \|y_i(s)\| + \beta \|k_i(s)\|,
\]
we can get from (49) that
\[
\|x_i(t) - \bar{x}(t)\| \leq C_5 \lambda_{1}^{t-1} + 2N^2C_1 \sum_{s=1}^{t-1} \lambda_{1}^{t-1-s} \left( w_1(s) + \sum_{j=1}^{N} \|y_j(s)\| + \sum_{j=1}^{N} \beta \|k_j(s)\| \right).
\]
Furthermore, observing that
\[
g_i^+(v_i(t)) \leq g_i^+(z_i(t)) + g_i^+(v_i(t)) - g_i^+(z_i(t))
\leq g_i^+(z_i(t)) + (\partial g_i^+(v_i(t)))^T (v_i(t) - z_i(t))
\leq g_i^+(z_i(t)) + \|\partial g_i^+(v_i(t))\| \|y_i(t)\|
\leq g_i^+(z_i(t)) + M \|y_i(t)\|,
\]
we arrive at
\[
\|k_i(t)\| \leq \frac{1}{\|d_i(t)\|} g_i^+(v_i(t))
\]
According to part (a) of Lemma 1 and the iteration rules (5a)–(5b), we get

\[
\frac{1}{M}(g_i^+(z_i(t)) + g_i^+(v_i(t)) - g_i^+(z_i(t))) \\
\leq \frac{1}{M}g_i^+(z_i(t)) + \frac{M}{M}||y_i(t)||,
\]

which implies that

\[
||x_i(t) - \bar{x}(t)|| \\
\leq C_6\lambda_1^{-1} + 2N^2C_5 \sum_{s=1}^{l-1} \lambda_1^{l-1-s}w_1(s) + C_7 \sum_{s=0}^{l-1} \lambda_1^{l-1-s}w_2(s) \\
+ C_8 \sum_{s=1}^{l-1-s} \lambda_2^{-1} + C_9 \sum_{s=1}^{l-1-s} \sum_{r=0}^{l-1-s} \lambda_2^{l-1-r}\alpha(r) + C_10 \sum_{s=1}^{l-1-s} \alpha(s),
\]

Finally, combining (53) with (54) produces

\[
\frac{1}{M}(x_i(t) - \bar{x}(t)) \\
\leq C_6\lambda_1^{-1} + \frac{M}{M}||y_i(t)||,
\]

thereby directly completing the proof.

**Proof of Lemma 10.** Noticing that

\[
\alpha(t + 1)\nabla f_i(x_i(t + 1)) - \alpha(t)\nabla f_i(x_i(t)) \\
\leq \alpha(t + 1)\nabla f_i(x_i(t + 1)) - \alpha(t)\nabla f_i(x_i(t + 1)) \\
+ \alpha(t)\nabla f_i(x_i(t + 1)) - \alpha(t)\nabla f_i(x_i(t)),
\]

under Assumption 2, we have

\[
||e_i(t)|| \leq M(\alpha(t) - \alpha(t + 1)) + \alpha(t)||\nabla f_i(x_i(t + 1)) - \nabla f_i(x_i(t))|| \\
\leq M(\alpha(t) - \alpha(t + 1)) + \alpha(t)\beta||x_i(t + 1) - x_i(t)||.
\]

According to part (a) of Lemma 1 and the iteration rules (53)–(55), we get

\[
||x_i(t + 1) - x_i(t)|| \\
\leq ||v_i(t) - \beta k_i(t) - x_i(t)|| \\
\leq ||z_i(t) - \bar{x}(t)|| + ||x_i(t) - \bar{x}(t)|| + ||y_i(t)|| + \beta ||k_i(t)|| \\
\leq \sum_{j=1}^{N} A_{ij}(t)||x_j(t) - \bar{x}(t)|| + ||x_i(t) - \bar{x}(t)|| + ||y_i(t)|| + \beta ||k_i(t)||,
\]

where the third inequality is obtained from the convex property of the norm function. Noticing (53) and \(0 \leq A_{ij}(t) < 1\) for all \(i, j = 1, 2, \cdots, N\) and \(t \geq 0\), we can also obtain

\[
||e_i(t)|| \leq M(\alpha(t) - \alpha(t + 1)) + L\alpha(t)\sum_{j=1}^{N}||x_j(t) - \bar{x}(t)|| + L\alpha(t)||x_i(t) - \bar{x}(t)|| \\
+ L\left(1 + \frac{M}{\beta M}\right)\alpha(t)||y_i(t)|| + \frac{\beta L}{M}\alpha(t)g_i^+(z_i(t)).
\]
Then, based on Corollary 1 and Lemma 9 we finally reach
\[
\|z(t)\| \leq M(\alpha(t) - \alpha(t + 1)) + C_{11} \gamma_1(t) + C_{12} \gamma_2(t) + L(1 + N) \xi_1(t) + C_{13} \alpha(t) \lambda_2^{-1}
\]
\[
+C_{14} \alpha(t) \sum_{s=0}^{t-1} \lambda_2^{1-s} \alpha(s) + C_{15} \alpha^2(t) + C_{16} \alpha(t) w_2(t).
\]
(60)

Therefore, we can derive from (46) that
\[
\|y_i(t) - \mu^i(t-1)\tilde{y}(t)\|
\]
\[
\leq C_3 \lambda_2^{t-1} + C_4 \sum_{s=0}^{t-1} \sum_{j=1}^{N} \lambda_2^{1-s} \|\epsilon_j(s)\|
\]
\[
\leq C_3 \lambda_2^{t-1} + C_{17} \sum_{s=0}^{t-1} \lambda_2^{1-s} \xi_2(s) + C_{18} \eta_1(t) + C_{19} \eta_2(t) + C_{20} \eta_3(t).
\]
(61)

So far, we have completed the proof.

\[\Box\]

Remark 5: In (58), for the purpose of dealing with the term \(\|x_i(t+1) - x_i(t)\|\), the property \(x_i(t) \in X_i\) is taken into account. It can be observed that the property \(x_i(t) \in X_i\) holds for all \(t \geq 1\). Thus, although it does not really matter, we can select the initial values \(x_i(0) \in X_i\) to assure the preciseness of the whole analysis.

Proof of Lemma 11: From the iteration rules (5a)–(5b), and Lemmas 1 and 6, we can get
\[
\|x_i(t+1) - v\|^2 \leq \|z_i(t) - y_i(t) - \beta k_i(t) - v\|^2 - \|\phi_i(t)\|^2
\]
\[
\leq \|z_i(t) - y_i(t) - v\|^2 - \beta(2 - \beta) \frac{(g_i^+(t))^2}{\|d_i(t)\|^2} - \|\phi_i(t)\|^2
\]
\[
\leq \|z_i(t) - y_i(t) - v\|^2 - \frac{\beta(2 - \beta)}{M^2} (g_i^+(t))^2 - \|\phi_i(t)\|^2.
\]
(62)

The first term on the right-hand side of (62) equals
\[
\|z_i(t) - v\|^2 + 2y_i^T(t)(v - z_i(t)) + \|y_i(t)\|^2.
\]
(63)

Letting \(g(t) = \alpha(t) \sum_{i=1}^{N} \nabla f_i(\bar{x}(t))\), for the second term in (63), we can attain
\[
2y_i^T(t)(v - z_i(t))
\]
\[
= 2\mu^i(t-1)g^T(t)(v - \bar{x}(t)) + 2\mu^i(t-1)g^T(t)(\bar{x}(t) - z_i(t))
\]
\[
+ 2\mu^i(t-1)(\tilde{y}(t) - g(t))^T(v - z_i(t)) + 2(y_i(t) - \mu^i(t-1)\tilde{y}(t))^T(v - z_i(t)).
\]
(64)

Under Assumption 3 we have that \(z_i(t)\) is uniformly bounded with respect to \(t\), which implies the uniform boundedness of \(v - z_i(t)\). So we also assume that \(\|v - z_i(t)\| \leq M\), where \(M\) is as given in Remark 2. Then, under Assumption 2 it follows from (64) that
\[
2y_i^T(t)(v - z_i(t))
\]
\[
\leq 2\alpha(t)\mu^i(t-1)(f(v) - f(\bar{x}(t))) + 2MN \sum_{j=1}^{N} \alpha(t)\|x_j(t) - \bar{x}(t)\|
\]
\[
+ 2LM \sum_{j=1}^{N} \alpha(t)\|x_j(t) - \bar{x}(t)\| + 2M\|y_i(t) - \mu^i(t-1)\tilde{y}(t)\|.
\]
(65)
Moreover, noticing Proposition 1, we have that
\[
\text{dist}(x, X_0) \leq R \max_{1 \leq i \leq N} \text{dist}(x, X_i) + R \max_{1 \leq i \leq N} g_i^+(x),
\]
which further implies that
\[
\|s(t) - \bar{x}(t)\| \leq R \sum_{i=1}^{N} \text{dist}(\bar{x}, X_i) + R \sum_{i=1}^{N} g_i^+(\bar{x})
\]
\[
\leq R \sum_{i=1}^{N} \|x_i(t) - \bar{x}(t)\| + R \sum_{i=1}^{N} g_i^+(z_i(t)) + R \sum_{i=1}^{N} g_i^+(\bar{x}) - R \sum_{i=1}^{N} g_i^+(z_i(t))
\]
\[
\leq R(MN + 1) \sum_{i=1}^{N} \|x_i(t) - \bar{x}(t)\| + R \sum_{i=1}^{N} g_i^+(z_i(t)).
\]  \hfill (66)

Observing that
\[
f(v) - f(\bar{x}(t)) \leq f(v) - f(s(t)) + MN\|s(t) - \bar{x}(t)\|,
\]
we can get from (65) and (66) that
\[
2y_i^T(t)(v - z_i(t)) \leq 2\alpha(t)\mu^i(t - 1)(f(v) - f(s(t))) + 2D_1 \sum_{j=1}^{N} \alpha(t)\|x_j(t) - \bar{x}(t)\|
\]
\[
+ 2MNR\alpha(t) \sum_{i=1}^{N} g_i^+(z_i(t)) + 2M\|y_i(t) - \mu^i(t - 1)\bar{y}(t)\|. \hfill (67)
\]

Now, for the third term in (63), we can obtain directly from (12) and Lemma 1 that \( \lim_{t \to \infty} (y_i(t) - \mu^i(t - 1)\bar{y}(t)) = 0 \). So we can also assume that \( \|y_i(t) - \mu^i(t - 1)\bar{y}(t)\| \leq M \) always holds. Then, it follows from Lemma 7 that
\[
\|y_i(t)\|^2 \leq 2\|y_i(t) - \mu^i(t - 1)\bar{y}(t)\|^2 + 2\|\mu^i(t - 1)\bar{y}(t)\|^2
\]
\[
\leq 2M\|y_i(t) - \mu^i(t - 1)\bar{y}(t)\| + 2M^2N^2\alpha^2(t). \hfill (68)
\]
Noting from the convex property of the norm function that
\[
\|z_i(t) - v\|^2 \leq \sum_{j=1}^{N} A_{ij}(t)\|x_i(t) - v\|^2,
\]  \hfill (69)
we can derive from (66), (67) and (68) that
\[
\|z_i(t) - y_i(t) - v\|^2
\]
\[
\leq \sum_{j=1}^{N} A_{ij}(t)\|x_i(t) - v\|^2 + 2\alpha(t)\mu^i(t - 1)(f(v) - f(s(t)))
\]
\[
+ 2D_1\alpha(t) \sum_{j=1}^{N} ||x_i(t) - \bar{x}(t)|| + 4M^2N^2\alpha^2(t) + \frac{N}{\tau_1} \sum_{i=1}^{N} (g_i^+(z_i(t)))^2
\]
\[
+ 2M^2N^2\alpha^2(t) + 4M\|y_i(t) - \mu^i(t - 1)\bar{y}(t)\|. \hfill (70)
\]
with \( \tau_1 > 0 \) being a constant defined in the following. We next relate \( (g_i^+(v_i(t)))^2 \) to \( (g_i^+(z_i(t)))^2 \). Clearly,
\[
g_i^+(v_i(t)) = (g_i^+(v_i(t)) - g_i^+(z_i(t))) + g_i^+(z_i(t)).
Thus, we can obtain
\[
(g_i^+(v_i(t)))^2 \geq 2(g_i^+(v_i(t)) - g_i^+(z_i(t)))g_i^+(z_i(t)) + (g_i^+(z_i(t)))^2
\]
\[
\geq 2(\partial g_i^+(z_i(t)))^T(v_i(t) - z_i(t))g_i^+(z_i(t)) + (g_i^+(z_i(t)))^2
\]
\[
\geq -2M\|y_i(t)\|g_i^+(z_i(t)) + (g_i^+(z_i(t)))^2, \tag{71}
\]
where we have also assumed that \(\|\partial g_i^+(z_i(t))\| \leq M\). Additionally, observing that
\[
-2M\|y_i(t)\|g_i^+(z_i(t))
\]
\[
\geq -\tau_2 M^2\|y_i(t)\|^2 - \frac{1}{\tau_2}(g_i^+(z_i(t)))^2
\]
\[
\geq -2\tau_2 M^3\|y_i(t) - \mu_i(t-1)\hat{y}(t)\| - 2\tau_2 M^4 N^2 \alpha^2(t) - \frac{1}{\tau_2}(g_i^+(z_i(t)))^2, \tag{72}
\]
with \(\tau_2 > 0\) being a constant defined in the following, we further have
\[
(g_i^+(v_i(t)))^2
\]
\[
\geq -2\tau_2 M^3\|y_i(t) - \mu_i(t-1)\hat{y}(t)\| - 2\tau_2 M^4 N^2 \alpha^2(t) + \left(1 - \frac{1}{\tau_2}\right)(g_i^+(z_i(t)))^2. \tag{73}
\]
Selecting \(\tau_2 = 4\), we can get from (62), (70) and (72) that
\[
\|x_i(t+1) - v\|^2
\]
\[
\leq \sum_{j=1}^{N} A_{ij}(t)\|x_i(t) - v\|^2 + 2\theta_2 \alpha(t)(f(v) - f(s(t))) + 2D_1 \sum_{j=1}^{N} \alpha(t)\|x_i(t) - \bar{x}(t)\| + 4M^2 N^2 R^2 \alpha^2(t)\tau_1 + \frac{N}{4\tau_1} \sum_{i=1}^{N} (g_i^+(z_i(t)))^2 + 8M^2 N^2 \beta(2 - \beta)\alpha^2(t) + 2M^2 N^2 \beta^2(t)
\]
\[
+ (2M + 8M \beta(2 - \beta))\|y_i(t) - \mu_i(t-1)\hat{y}(t)\| - \frac{3\beta(2 - \beta)}{4M^2}(g_i^+(z_i(t)))^2 - \|\phi_i(t)\|^2, \tag{74}
\]
where the first inequality is obtained from the facts that \(\theta_2 \leq \mu_i(t) < 1\) and \(f(v) - f(s(t)) \leq 0\) for all \(i = 1, 2, \cdots N\) and \(t \geq 0\). Additionally, from part (c) of Lemma 4 we can attain
\[
\sum_{i=1}^{N} \pi_i^i(t+1)\sum_{j=1}^{N} A_{ij}(t)\|x_i(t) - v\|^2 = \sum_{i=1}^{N} \pi_i^i(t)\|x_i(t) - v\|^2.
\]
Moreover, from parts (a) and (b) of Lemma 4 we can deduce
\[
\theta_1 \leq \pi_i(t) < 1, \quad \sum_{i=1}^{N} \pi_i^i(t) = 1,
\]
for all \(i = 1, 2, \cdots N\) and \(t \geq 0\). Then, selecting \(\tau_1 = \frac{M^2 N}{\theta_1 \mu_i(x_i - \bar{x})}\), multiplying both sides of (74) by \(\pi_i^i(t+1)\), and taking summation from \(i = 1\) to \(i = N\) yield
\[
\sum_{i=1}^{N} \pi_i^i(t+1)\|x_i(t+1) - v\|^2
\]
\[
\leq \sum_{i=1}^{N} \pi_i^i(t)\|x_i(t) - v\|^2 + 2\theta_2 \alpha(t)(f(v) - f(s(t))) + 2D_1 \alpha(t) \sum_{j=1}^{N} \|x_j(t) - \bar{x}(t)\|
Furthermore, noting that
\[ D_2 \alpha^2(t) + D_3 \sum_{i=1}^{N} \|y_i(t) - \mu_i(t-1)\bar{y}(t)\| \]
\[ - \sum_{i=1}^{N} \theta_1 \|\phi_i(t)\|^2 - \sum_{i=1}^{N} \frac{\theta_1 \beta(2-\beta)}{2M^2} (g_i^+(z_i(t)))^2. \] 
(75)

Lastly, combining (14) and (15) with (75) together can directly complete the proof.

**Proof of Lemma 12** From (16) and Proposition 2, we arrive at
\[ \sum_{i=1}^{N} \pi^i(t+1)\|x_i(t+1) - v\|^2 + e(t+1) \]
\[ \leq \sum_{i=1}^{N} \pi^i(t)\|x_i(t) - v\|^2 + 2\theta_2 \alpha(t)(f(v) - f(s(t))) + D_4 \gamma_1(t) + D_5 \gamma_2(t) + D_6 \eta_1(t) \]
\[ - \theta_1 \sum_{i=1}^{N} \|\phi_i(t)\|^2 + D_7 \eta_2(t) + D_8 \eta_3(t) - \frac{\theta_1 \beta(2-\beta)}{2M^2} \sum_{i=1}^{N} (g_i^+(z_i(t)))^2 + \xi_3(t) \]
\[ + c(t) + a_1 b_1 (\lambda_1 - 1) \gamma_1(t) + a_2 b_2 (\lambda_1 - 1) \gamma_2(t) + c_1 (\lambda_2 - 1) \eta_1(t) \]
\[ + c_2 (\lambda_2 - 1) \eta_2(t) + a_3 b_3 (\lambda_2 - 1) \eta_3(t) + a_1 b_1 \alpha(t) w_1(t) + a_2 b_2 \alpha(t) w_2(t) \]
\[ + a_3 b_3 \alpha(t) w_2(t) + c_1 \gamma_1(t) + c_2 \gamma_2(t). \] 
(76)

Clearly, with the given \( a_i, \ i = 1, 2, 3, \) and \( c_i, \ i = 1, 2, \) we have
\[ D_4 \gamma_1(t) + a_1 b_1 (\lambda_1 - 1) \gamma_1(t) + c_1 \gamma_1(t) = 0, \]
\[ D_5 \gamma_2(t) + a_2 b_2 (\lambda_1 - 1) \gamma_2(t) + c_2 \gamma_2(t) = 0, \]
\[ D_6 \eta_1(t) + c_1 (\lambda_1 - 1) \eta_1(t) = 0, \]
\[ D_7 \eta_2(t) + c_2 (\lambda_2 - 1) \eta_2(t) = 0, \]
\[ D_8 \eta_3(t) + a_3 b_3 (\lambda_2 - 1) \eta_3(t) = 0. \]

Furthermore, noting that
\[ a_1 b_1 \alpha(t) w_1(t) \leq \frac{1}{2} a_1^2 \alpha^2 + \frac{1}{2} b_1^2 w_1^2(t) \leq \frac{1}{2} a_1^2 \alpha^2 + \frac{1}{2} b_1^2 N \sum_{i=1}^{N} \|\phi_i(t)\|^2, \]
\[ a_2 b_2 \alpha(t) w_2(t) \leq \frac{1}{2} a_2^2 \alpha^2 + \frac{1}{2} b_2^2 N \sum_{i=1}^{N} (g_i^+(z_i(t)))^2, \ j = 2, 3, \]
we can get from (76) that
\[ \sum_{i=1}^{N} \pi^i(t+1)\|x_i(t+1) - v\|^2 + e(t+1) \]
\[ \leq \sum_{i=1}^{N} \pi^i(t)\|x_i(t) - v\|^2 + e(t) + 2\theta_2 \alpha(t)(f(v) - f(s(t))) \]
\[ + \xi_4(t) - \frac{\theta_1}{2} \sum_{i=1}^{N} \|\phi_i(t)\|^2 - \frac{\theta_1 \beta(2-\beta)}{4M^2} \sum_{i=1}^{N} (g_i^+(z_i(t)))^2, \]
(77)

thereby completing the proof.
REFERENCES

[1] J. Wang and N. Elia, “A control perspective for centralized and distributed convex optimization,” in Proc. 50th IEEE Conf. Decision Control, Orlando, FL, USA, Dec. 2011, pp. 3800–3805.
[2] B. Gharesifard and J. Cortés, “Distributed continuous-time convex optimization on weight-balanced digraphs,” IEEE Trans. Autom. Control, vol. 59, no. 3, pp. 781–786, 2014.
[3] S. S. Kia, J. Cortés, and S. Martínez, “Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication,” Automatica, vol. 55, no. 1, pp. 254–264, 2015.
[4] S. Liang, L. Y. Wang, and G. Yin, “Exponential convergence of distributed primal-dual convex optimization algorithm without strong convexity,” Automatica, vol. 105, no. 7, pp. 298–306, 2019.
[5] Q. Liu and J. Wang, “A second-order multi-agent network for bound-constrained distributed optimization,” IEEE Trans. Autom. Control, vol. 60, no. 12, pp. 3310–3315, 2015.
[6] G. Shi, K. H. Johansson, and Y. Hong, “Reaching an optimal consensus: Dynamical systems that compute intersections of convex sets,” IEEE Trans. Autom. Control, vol. 58, no. 3, pp. 610–622, 2013.
[7] Z. Qiu, S. Yang, and J. Wang, “Distributed constrained optimal consensus of multi-agent systems,” Automatica, vol. 68, pp. 209–215, 2016.
[8] P. Lin and W. Ren, “Distributed subgradient projection algorithm for multi-agent optimization with nonidentical constraints and switching topologies,” in Proc. 51st IEEE Conf. Decision Control, Maui, HI, USA, Dec. 2012, vol. 5, pp. 6813–6818.
[9] P. Lin, W. Ren, and J. A. Farrell, “Distributed continuous-time optimization: Nonuniform gradient gains, finite-time convergence, and convex constraint set,” IEEE Trans. Autom. Control, vol. 62, no. 5, pp. 2239–2253, 2017.
[10] Z. Yan, J. Wang, and G. Li, “A collective neurodynamic optimization approach to bound-constrained nonconvex optimization,” Neural Netw., vol. 55, pp. 20–29, 2014.
[11] P. Yi and Y. Hong, “Distributed continuous-time gradient-based algorithm for constrained optimization,” in Proc. 33rd Chinese Control Conf., Nanjing, China, Jul. 2014, pp. 1563–1567.
[12] P. Yi, Y. Hong, and F. Liu, “Distributed gradient algorithm for constrained optimization with application to load sharing in power systems,” Syst. Control Lett., vol. 83, pp. 45–52, 2015.
[13] Q. Liu, S. Yang, and J. Wang, “A collective neurodynamic approach to distributed constrained optimization,” IEEE Trans. Neural. Netw. Learn. Syst., vol. 28, no. 8, pp. 1747–1758, 2017.
[14] Q. Liu and J. Wang, “A one-layer projection neural network for nonsmooth optimization subject to linear equalities and bound constraints,” IEEE Trans. Neural Netw. Learn. Syst., vol. 24, no. 5, pp. 812–824, 2013.
[15] S. Yang, Q. Liu, and J. Wang, “A multi-agent system with a proportional-integral protocol for distributed constrained optimization,” IEEE Trans. Autom. Control, vol. 62, no. 7, pp. 3461–3467, 2017.
[16] A. Nedić and A. Ozdaglar, “Distributed subgradient methods for multi-agent optimization,” IEEE Trans. Autom. Control, vol. 54, no. 1, pp. 48–61, 2009.
[17] 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, 2010.
[18] D. Yuan, D. W. C. Ho, and Y. Hong, “On Convergence Rate of Distributed Stochastic Gradient Algorithm for Convex Optimization with Inequality Constraints,” SIAM J. Optim., vol. 54, no. 5, pp. 2872–2892, 2016.
[19] 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.
[20] 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.
[21] Q. Liu, S. Yang, and Y. Hong, “Constrained consensus algorithms with fixed step size for distributed convex optimization over multiagent networks,” IEEE Trans. Autom. Control, vol. 62, no. 8, pp. 4259–4265, 2017.
[22] A. Nedić and A. Ozlshesky, “Distributed optimization over time-varying directed graphs,” IEEE Trans. Autom. Control, vol. 60, no. 3, pp. 601–615, 2015.
[23] A. Nedić, A. Ozdaglar, and W. Shi, “Achieving geometric convergence for distributed optimization over time-varying graphs,” SIAM J. Optim., vol. 27, no. 4, pp. 2597–2633, 2017.
[24] S. Liang, L. Y. 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, 2020.
[25] C. Gu, Z. Wu, and J. Li, “Regularized dual gradient distributed method for constrained convex optimization over unbalanced directed graphs,” Numer. Algorithms, vol. 84, no. 1, pp. 91–115, 2020.
[26] S. Pu, W. Shi, J. Xu, and A. Nedić, “A push-pull gradient method for distributed optimization in networks,” in Proc. 57th IEEE Conf. Decision Control, Miami, FL, Dec. 2018, pp. 3385–3390.
[27] R. Xin and U. A. Khan, “A linear algorithm for optimization over directed graphs with geometric convergence,” Syst. Control Lett., vol. 2, no. 3, pp. 315–320, 2018.
[28] R. Xin and U. A. Khan, “Distributed heavy-ball: A generalization and acceleration of first-order methods with gradient tracking,” IEEE Trans. Autom. Control, vol. 65, no. 6, pp. 2627–2633, 2020.
[29] F. Saadatniaki, R. Xin, and U. A. Khan, “Decentralized optimization over time-varying directed graphs with row and column-stochastic matrices,” IEEE Trans. Autom. Control, vol. 65, no. 11, pp. 4769–4780, 2020.
[30] V. S. Mai and E. H. Abed, “Distributed optimization over weighted directed graphs using row stochastic matrix,” Proc. Amer. Control Conf., Boston, MA, Jul. 2016, pp. 7165–7170.
[31] 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.
[32] H. Li, Q. Lv, 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, 2019.
[33] B. T. Polyak, “Minimization of nonsmooth functionals,” U.S.S.R. Comput. Math. Math. Phys., vol. 9, no. 3, pp. 509–521, 1969.
[34] A. Nedić, “Random algorithms for convex minimization problems,” Math. Program., vol. 129, no. 2, pp. 225–253, 2011.
[35] G. Qu and N. Li, “Harnessing smoothness to accelerate distributed optimization,” IEEE Trans. Control Netw. Syst., vol. 5, no. 3, pp. 1245–1260, 2018.
[36] C. Xi, R. Xin, and U. A. Khan, “ADD-OPT: Accelerated distributed directed optimization,” IEEE Trans. Autom. Control, vol. 63, no. 5, pp. 1329–1339, 2018.
[37] R. Xin, S. Pu, A. Nedić, and U. A. Khan, “A general framework for decentralized optimization with first-order methods,” Proc. IEEE, vol. 108, no. 11, pp. 1869–1889, 2020.