WILEY | Hindawi

Research Article

Distributed Optimization Methods for Nonconvex Problems with Inequality Constraints over Time-Varying Networks

Jueyou Li,[1] Chuanye Gu,[1] Zhiyou Wu,[1] and Changzhi Wu[2]

[1]*School of Mathematical Sciences, Chongqing Normal University, Chongqing 400047, China*
[2]*Australasian Joint Research Centre for Building Information Modelling, School of Built Environment, Curtin University, Bentley, WA 6102, Australia*

Correspondence should be addressed to Jueyou Li; lijueyou@163.com

Network-structured optimization problems are found widely in engineering applications. In this paper, we investigate a nonconvex distributed optimization problem with inequality constraints associated with a time-varying multiagent network, in which each agent is allowed to locally access its own cost function and collaboratively minimize a sum of nonconvex cost functions for all the agents in the network. Based on successive convex approximation techniques, we first approximate locally the nonconvex problem by a sequence of strongly convex constrained subproblems. In order to realize distributed computation, we then exploit the exact penalty function method to transform the sequence of convex constrained subproblems into unconstrained ones. Finally, a fully distributed method is designed to solve the unconstrained subproblems. The convergence of the proposed algorithm is rigorously established, which shows that the algorithm can converge asymptotically to a stationary solution of the problem under consideration. Several simulation results are illustrated to show the performance of the proposed method.

1. Introduction

Network-structured problems have drawn recently considerable attention in various applications, such as mobile ad hoc networks, wireless sensor networks, and Internet networks [1–3]. The absence of centralized access to information and time-varying network connectivity are common features of network-structured problems. For this reason, distributed optimization methods associated with multiagent networks should be designed on the base of local communication and computation and the change of network topology. Distributed computing allows each agent to only utilize its own cost function and communicate with its direct neighbors, which has the potential advantage on protecting agents' privacy. In recent years, there is an increasing trend to develop distributed optimization by integrating the communication and computation cooperation (see, e.g., [4, 5] and references therein).

The methods designed to solve distributed convex problems have been widely studied in the literature. Based on the consensus averaging mechanism, there are several useful approaches, including primal consensus distributed methods, dual consensus distributed methods, and primal-dual consensus distributed methods. Nedic and Ozdaglar in [5] originally developed a distributed subgradient-based algorithm, in which every agent optimizes its own objective and locally exchanges information with neighboring agents in a network. The convergence rate of their proposed algorithm has been obtained. But, only unconstrained distributed problems were investigated. In [6], Nedic et al. generalized the distributed method in [5] to solve constrained convex optimization. Later, many researchers proposed various extensions based on primal (sub)gradient-based methods in, for example, [7–12]. In [13], Duchi et al. extended the centralized dual averaging algorithm to the distributed setting and then proposed a distributed dual averaging algorithm. The convergence result of their algorithm shows that the number of iterations is dependent on the underlying network sizes and spectral gaps. The primal-dual consensus distributed method was designed for solving distributed convex problems

with global inequality constraints under the framework of Lagrangian duality. By finding the corresponding saddle-points of the Lagrangian function, Zhu and Martinez [14] first proposed a primal-dual consensus distributed algorithm and established the convergence results. The authors in [15] obtained the explicit convergence rate of the algorithm under the strong connectivity of networks.

However, the methods mentioned above are not suitable for the general nonconvex distributed optimization problems. Until recently, there are some references designing algorithms for distributed nonconvex problems, such as [16, 17]. In [16], Lorenzo and Scutari developed a novel distributed algorithm for solving unconstrained nonconvex optimization problems which associated a multiagent network with time-varying connectivity. They combined a successive convex approximation technique and a dynamic consensus mechanism to realize distributed computation as well as local information exchange in the network. For nonconvex optimization problems with constraints, Scutari et al. [17] proposed a successive convex approximation method by solving a sequence of strongly convex subproblems while maintaining feasibility. They have shown that their proposed method can converge to a stationary solution of the original constrained nonconvex problem under consideration. However, the method proposed in [17] is not suitable in the distributed setting. Additionally, the Lagrange dual method is utilized to solve the sequence of strongly convex subproblems, which may largely enlarge the dimension of the problem due to introducing the dual variables. Thus, the computational difficulty and cost may be increased.

In this paper, we investigate a distributed nonconvex problem with inequality constraints. The main contributions of this paper are twofold: (i) based on the penalty function method, a distributed algorithm for solving the nonconvex problem with global inequality constraints is proposed; (ii) the convergence of the proposed algorithm is rigorously proved. More specifically, we first transform the nonconvex problem to a sequence of strongly convex subproblems based on successive convex approximation techniques. In order to realize distributed computation, we then exploit the exact penalty function method to transform the sequence of strongly convex constrained subproblems into unconstrained ones. Finally, we propose a fully distributed method to solve these unconstrained subproblems. We obtain the convergence results of the proposed algorithm and demonstrate several numerical simulations.

The work in this paper is closely related to the previous works [16–19]. Our method is based on the algorithm proposed in [16], but the problem we consider in this paper is different from the one in [16], since we investigate a nonconvex problem with global inequality constraints. The proposed algorithm is different from [17], since our algorithm exploits the exact penalty function method to solve the related subproblems possibly reducing the dimensionality of the problem. However, the algorithm in [17] utilized the Lagrange dual method to solve the subproblems, in which the computation is not implemented in a distributed manner. Our algorithm is an extension of the algorithm in [18] to handle with the nonconvex case. By comparison, we solve a constrained

distributed nonconvex optimization in this paper, while the authors in [19] solved the unconstrained one.

The remainder of this paper is organized as follows. Section 2 provides the problem statement and related preparations. Section 3 gives the algorithm development. Section 4 proposes the distributed algorithm and establishes the results of convergence. Numerical simulations are given in Section 5. Finally, the conclusions are obtained in Section 6.

2. Problem Statement and Preparations

In the section, we state the optimization problem under consideration and give the assumptions and definitions that will be used in the sequel.

2.1. Problem Statement. Consider the following distributed nonconvex optimization problem:

$$\min_{\mathbf{x} \in \mathscr{K}} \quad F(\mathbf{x}) := \sum_{i=1}^{N} f_i(\mathbf{x}) \tag{1}$$

$$\text{s.t.} \quad \mathbf{g}(\mathbf{x}) \leq \mathbf{0},$$

where $f_i(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ is a nonconvex smooth cost function, only known by agent i for $i = 1, \ldots, N$; $\mathbf{g}(\mathbf{x}) := (g_1(\mathbf{x}), \ldots, g_p(\mathbf{x}))^T$, $g_s(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ is a convex smooth function for $s = 1, \ldots, p$; and $\mathscr{K} \subseteq \mathbb{R}^d$ is a closed, convex set. The constraint $\mathbf{g}(\mathbf{x})$ and the set $\mathscr{K}$ are known by all the agents. Let

$$\chi := \left\{ \mathbf{x} \in \mathbb{R}^d \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \ \mathbf{x} \in \mathscr{K} \right\} \tag{2}$$

be the feasible set of problem (1). We assume that Slater's condition [20] is satisfied for problem (1).

Problem (1) is ubiquitous that arises in many applications, such as networking, wireless communications, and machine learning. Therefore, it is meaningful to solve the problem.

2.2. Assumption and Definition. We first give the description of the network topology. Time is assumed to be discrete. At each time slot k, the network is modeled as a directed graph $\mathscr{G}[k] = (\mathscr{V}, \mathscr{E}[k])$, where $\mathscr{V} = \{1, \ldots, N\}$ is the set of nodes with N agents, and $\mathscr{E}[k]$ represents the set of time-varying directed edges. The neighborhood of agent i at time k is defined as $\mathscr{N}_i^{in}[k] = \{j \mid (j, i) \in \mathscr{E}[k]\} \cup \{i\}$. The communication pattern between neighbors is set as follows: agent $j \neq i$ in $\mathscr{N}_i^{in}[k]$ can communicate with node i at time k. We assign time-varying weights $(a_{ij}[k])$ to match the digraph $\mathscr{G}[k]$ and define the weight matrix $A[k] = (a_{ij}[k])_{i,j=1}^N$.

Assumption 1 (see [5]). (A1) The graph sequence $\{\mathscr{G}[k]\}$ is B-*strongly* connected; that is, there is an integer $B > 0$ such that the graph $\mathscr{G}[n] = (\mathscr{V}, \mathscr{E}[n])$, with the edge set $\mathscr{E}[n] = \cup_{k=nB}^{(n+1)B-1} \mathscr{E}[k]$, is strongly connected, for all $n \geq 0$.

(A2) There is a scalar ϑ with $0 < \vartheta < 1$, for all $i \in \{1, 2, \ldots, N\}$: (i) $a_{ii}[k] \geq \vartheta$ for all $k \geq 0$; (ii) $a_{ij}[k] \geq \vartheta$ for all $k \geq 0$ and all agents j communicate directly with agent i; (iii) otherwise, all agents j cannot communicate directly with agent i, $a_{ij}[k] = 0$ for all $k \geq 0$.

(A3) Each weight matrix $A[k]$ satisfies

$$A[k]\mathbf{1} = \mathbf{1},$$

$$\mathbf{1}^T A[k] = \mathbf{1}^T, \tag{3}$$

$$\forall k \geq 0.$$

Assumption 1 shows that, at any time k, any agent j will receive the information from agent i within the next B time slots. Moreover, the weigh matrix $A[k]$ is double-stochastic.

Assumption 2. (B1) The set $\mathscr{K}$ is nonempty, closed, and convex.

(B2) For $i \in \{1, \ldots, N\}$, each local function f_i is continuously differentiable on $\mathscr{K}$, and ∇f_i is Lipschitz continuous on $\mathscr{K}$ with constant L_i; let $L_{\max} = \max_i\{L_i\}$.

(B3) ∇F is bounded on $\mathscr{K}$; that is, there is a finite number $L_F > 0$ such that $\|\nabla F(\mathbf{x})\| \leq L_F$, for all $\mathbf{x} \in \mathscr{K}$.

(B4) For $s \in \{1, \ldots, p\}$, each function ∇g_s is bounded on $\mathscr{K}$; that is, there is a finite number L_{g_s} such that $\|\nabla g_s\| \leq L_{g_s}$; let $L_G = \max_s\{L_{g_s}\}$.

(B5) F is coercive on $\mathscr{K}$; that is, $\lim_{\mathbf{x}\in\mathscr{K}, \|\mathbf{x}\|\to\infty} F(\mathbf{x}) = +\infty$.

The above assumptions are quite general, which can be satisfied by a large class of problems in practical applications. Assumption (B5) can make sure that problem (1) has a solution.

The goal in this paper is to design a method that can find a stationary solution of problem (1). Moreover, the method is implemented in the distributed scenario satisfying Assumptions 1 and 2.

Next we introduce several definitions, which will be used in the convergence analysis of our method.

Definition 3 (regularity [21]). A point $\overline{\mathbf{x}} \in \chi$ is called *regular* for problem (1) if the Mangasarian-Fromovitz Constraint Qualification (MFCQ) holds at $\overline{\mathbf{x}}$, that is, if the following implication is satisfied:

$$0 \in \sum_{s\in\overline{I}}\rho_s\nabla g_s(\overline{\mathbf{x}}) + N_{\mathscr{K}}(\overline{\mathbf{x}}),$$

$$\rho_s \geq 0, \ \forall s \in \overline{I} \implies \rho_s = 0, \ \forall s \in \overline{I}, \tag{4}$$

where

$$N_{\mathscr{K}}(\overline{\mathbf{x}}) = \left\{\mathbf{d} \in \mathbb{R}^d : \mathbf{d}^T(\mathbf{y} - \overline{\mathbf{x}}) \leq 0, \ \forall \mathbf{y} \in \mathscr{K}\right\} \tag{5}$$

is the normal cone to $\mathscr{K}$ at $\overline{\mathbf{x}}$ and $\overline{I} = \{s \in \{1, 2, \ldots, p\} : g_s(\overline{\mathbf{x}}) = 0\}$ is the index set of convex constraints that are active at $\overline{\mathbf{x}}$.

Definition 4 (stationary point [17]). A point $\overline{\mathbf{x}} \in \chi$ is a *stationary point* of problem (1), if it satisfies the following KKT system:

$$0 \in \nabla F(\overline{\mathbf{x}}) + \sum_{s\in\overline{I}}\rho_s\nabla g_s(\overline{\mathbf{x}}) + N_{\mathscr{K}}(\overline{\mathbf{x}}), \quad \rho_s \geq 0, \ \forall s \in \overline{I}, \tag{6}$$

where $\rho_s \in \overline{I}$ are Lagrange multipliers chosen suitably.

As pointed out by [17], a regular (local) minimum point of problem (1) is also a stationary point. It is well known that the traditional goal for solving nonconvex problems is actually to find stationary points. To simplify the discussion, we assume that all feasible points of problem (1) are regular throughout the rest of this paper.

3. Development of Algorithm

Designing an effective distributed method for problem (1), we had to face three main challenges: (i) the nonconvexity of the objective function F; (ii) the unavailability of global knowledge on F; and (iii) the presence of inequality constraints **g**. In order to deal with the difficulties, we utilize successive convex approximation (SCA) techniques, exact penalty function methods, and dynamic consensus mechanisms to develop our algorithm.

3.1. Local SCA Approximation. In a distributed setting, the computational cost for directly solving problem (1) is considerably high and even infeasible. We would prefer to suitably approximate problem (1), in the sense of local convex approximation.

By copying the global variable **x**, each agent i maintains a local estimate $\mathbf{x}_i$ that needs to be updated at each iteration. We rewrite $F(\mathbf{x}_i) = f_i(\mathbf{x}_i) + \sum_{j\neq i} f_j(\mathbf{x}_i)$ and consider a convexification of F as follows: at each iteration, we use a strongly convex function $\widetilde{f}_i(\cdot; \mathbf{x}_i[k])$ to replace the nonconvex function f_i and linearize the $\sum_{j\neq i} f_j(\mathbf{x}_i)$ at $\mathbf{x}_i[k]$; that is,

$$\widehat{F}_i(\mathbf{x}_i; \mathbf{x}_i[k]) = \widetilde{f}_i(\mathbf{x}_i; \mathbf{x}_i[k]) + \pi_i(\mathbf{x}_i[k])^T(\mathbf{x}_i - \mathbf{x}_i[k]), \tag{7}$$

where $\widetilde{f}_i(\cdot; \mathbf{x}_i[k]) : \mathscr{K} \to \mathbb{R}$ is a strongly convex surrogate of the nonconvex f_i and $\pi_i(\mathbf{x}_i[k])$ is the gradient for the term $\sum_{j\neq i} f_j(\mathbf{x}_i)$ at $\mathbf{x}_i[k]$; that is,

$$\pi_i(\mathbf{x}_i[k]) = \sum_{j\neq i}\nabla_{\mathbf{x}}f_j(\mathbf{x}_i[k]). \tag{8}$$

At each iteration k, a strongly convex problem is solved by agent i

$$\widehat{\mathbf{x}}_i(\mathbf{x}_i[k]) = \arg\min_{\mathbf{x}_i\in\chi} \widehat{F}_i(\mathbf{x}_i; \mathbf{x}_i[k]). \tag{9}$$

Note that $\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])$ in (9) is well defined, since subproblem (9) has a unique solution.

We give the following assumptions on the approximation of f_i.

Assumption 5. Each $\widetilde{f}_i$ satisfies the following:

(C1) $\widetilde{f}_i(\cdot; \mathbf{x})$ is uniformly strongly convex on $\mathscr{K} \times \chi$ with a strongly convex parameter $\tau_i > 0$.

(C2) $\widetilde{f}_i(\mathbf{x}; \cdot)$ is uniformly Lipschitz continuous on $\mathscr{K} \times \chi$.

(C3) $\nabla\widetilde{f}_i(\mathbf{x}; \mathbf{x}) = \nabla f_i(\mathbf{x})$, for all $\mathbf{x} \in \chi$.

Assumption 5 is quite natural. The function $\widetilde{f}_i$ is viewed as a strongly convex, local approximation of f_i around $\mathbf{x}$ that inherits the first-order properties of f_i. Assumption 5(C2) is the requirement of Lipschitzianity that is surely satisfied, for example, if the set $\mathscr{K}$ is bounded. For a given f_i, several feasible choices are provided in [16, 22, 23].

3.2. Exact Penalty Function Method. Subproblem (9) is not easy to solve due to the presence of the global constraints $\mathbf{g}$ and the local accessibility of f_i; thus, the exact penalty function method is utilized to transform subproblem (9) into an unconstrained problem, given by

$$\widehat{\mathbf{x}}_i\left(\mathbf{x}_i\left[k\right]\right) = \arg\min_{\mathbf{x}_i \in \mathscr{K}} \left\{\widehat{F}_i\left(\mathbf{x}_i; \mathbf{x}_i\left[k\right]\right) + c_i p\left(\mathbf{x}_i\right)\right\}, \quad (10)$$

where $p(\mathbf{x}_i) = \max\{0, g_1(\mathbf{x}_i), \ldots, g_p(\mathbf{x}_i)\}$, $c_i > 0$ is a penalty parameter. We can obtain that $p(\mathbf{x}_i)$ is convex on $\mathscr{K}$, and $\|S_p(\mathbf{x}_i)\| \leq L_G$ for all $S_p(\mathbf{x}_i) \in \partial p(\mathbf{x}_i)$, where $S_p(\mathbf{x}_i)$ stands for the subgradient of p at $\mathbf{x}_i$.

Under suitable conditions [24], *the solution set of the penalized problem (10) coincides with the solution set of the constrained problem (9)*. In order to explain the fact in detail, we introduce the Lagrangian function of problem (9):

$$\mathscr{L}_i\left(\mathbf{x}_i; \lambda_i\right) = \widehat{F}_i\left(\mathbf{x}_i; \mathbf{x}_i\left[k\right]\right) + \lambda_i^T \mathbf{g}\left(\mathbf{x}_i\right), \quad (11)$$

where $\lambda_i = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{ip})^T \in \mathbb{R}_+^p$ is the vector of dual variable corresponding to the constraints $\mathbf{g}(\mathbf{x}_i) \leq \mathbf{0}$. The dual problem of problem (9) is

$$\max_{\lambda_i \in \mathbb{R}_+^p} d\left(\lambda_i\right) \quad \text{with } d\left(\lambda_i\right) = \min_{\mathbf{x}_i \in \mathscr{K}} \mathscr{L}_i\left(\mathbf{x}_i; \lambda_i\right). \quad (12)$$

It can be proved that no duality gap exists between subproblem (9) and its dual problem (12) if Slater's condition is satisfied (see Proposition 5.3.1 in [20]). In addition, the set of dual optimal solutions is nonempty bounded. Thus, based on Proposition 1 in [24], there is a penalty parameter satisfied $c_i > \sum_{s=1}^{p} \lambda_{i,s}$ such that the solutions of penalized problem (10) are the same as those of subproblem (9). Thus, throughout the rest of this paper, we can always select a finite penalty parameter $\bar{c}_i$ such that $\bar{c}_i \geq c_i$.

3.3. Consensus Update. We now introduce a consensus mechanism to ensure that each local estimate $\mathbf{x}_i$ gradually coincides among all agents. A consensus-based step is used on $\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])$ and each agent i updates its state as follows:

$$\mathbf{x}_i\left[k+1\right] = \sum_{j \in \mathscr{N}_i^{in}[k]} a_{ij}\left[k\right] \widehat{\mathbf{x}}_j\left(\mathbf{x}_j\left[k\right]\right), \quad (13)$$

where $a_{ij}[k]$ is the weight satisfying Assumption 1.

Since the evaluation of $\pi_i(\mathbf{x}_i[k])$ in (8) requires the quantity of all $\nabla f_j(\mathbf{x}_i[k])$, $j \neq i$, which is not feasible at agent i. In order to deal with the issue, we need a local estimation of $\pi_i(\mathbf{x}_i[k])$ in (8), eventually converging to $\sum_{j \neq i} \nabla f_j(\mathbf{x}_i[k])$. We rewrite $\pi_i(\mathbf{x}_i[k])$ in (8) as follows:

$$\pi_i\left(\mathbf{x}_i\left[k\right]\right) = N \cdot \left(\frac{1}{N}\sum_{j=1}^{N} \nabla f_j\left(\mathbf{x}_i\left[k\right]\right)\right) - \nabla f_i\left(\mathbf{x}_i\left[k\right]\right), \quad (14)$$

and let $\overline{\nabla f}(\mathbf{x}_i[k]) = (1/N)\sum_{j=1}^{N} \nabla f_j(\mathbf{x}_i[k])$. Then, we replace $\pi_i(\mathbf{x}_i[k])$ in (14) by $\widetilde{\pi}_i(\mathbf{x}_i[k])$

$$\widetilde{\pi}_i\left(\mathbf{x}_i\left[k\right]\right) = N \cdot \mathbf{y}_i\left[k\right] - \nabla f_i\left(\mathbf{x}_i\left[k\right]\right), \quad (15)$$

where $\mathbf{y}_i[k]$ is a locally auxiliary variable updated by agent i, asymptotically tracking $\overline{\nabla f}(\mathbf{x}_i[k])$. By using the dynamic averaging consensus strategy [25], we can update $\mathbf{y}_i[k]$ in (15) via the following formula:

$$\mathbf{y}_i\left[k+1\right] = \sum_{j=1}^{N} a_{ij}\left[k\right] \mathbf{y}_j\left[k\right] \\ + \left(\nabla f_i\left(\mathbf{x}_i\left[k+1\right]\right) - \nabla f_i\left(\mathbf{x}_i\left[k\right]\right)\right), \quad (16)$$

where $\mathbf{y}_i[0] = \nabla f_i(\mathbf{x}_i[0])$.

Note that the update of $\mathbf{y}_i[k]$ in (16), and thus $\widetilde{\pi}_i(\mathbf{x}_i[k])$) in (15), can be now performed locally with message exchanges with the agents in the neighborhood $\mathscr{N}_i^{in}[k]$. By the above description, problem (10) can be converted into the following problem:

$$\widetilde{\mathbf{x}}_i\left[k\right] = \arg\min_{\mathbf{x}_i \in \mathscr{K}} \widetilde{U}_i\left(\mathbf{x}_i; \mathbf{x}_i\left[k\right], \widetilde{\pi}_i\left(\mathbf{x}_i\left[k\right]\right)\right), \quad (17)$$

where $\widetilde{U}_i(\mathbf{x}_i; \mathbf{x}_i[k], \widetilde{\pi}_i(\mathbf{x}_i[k])) = \widetilde{f}_i(\mathbf{x}_i; \mathbf{x}_i[k]) + \widetilde{\pi}_i(\mathbf{x}_i[k])^T(\mathbf{x}_i - \mathbf{x}_i[k]) + c_i p(\mathbf{x}_i)$.

4. Algorithm and Convergence Results

Based on the previous algorithm development, we propose an exact penalty function based distributed algorithm (EPDA, for short) to solve problem (1), presented in Algorithm 1.

In order to obtain the convergence results of the proposed algorithm, we first give Lemma 6, which shows the relationship between the solutions of problem (1) and the solutions of subproblem (9). Note that the following conditions (18) and (19) can be satisfied; please see the proof of relation (A.32).

Lemma 6. *Suppose that Assumptions 1, 2, and 5 hold. Let $\{\mathbf{x}_i[k]\}$ be the sequence generated by Algorithm 1; the following results hold:*
(i) If

$$\liminf_{k\to\infty} \left\|\widehat{\mathbf{x}}_i\left(\mathbf{x}_i\left[k\right]\right) - \mathbf{x}_i\left[k\right]\right\| = 0, \quad \forall i = 1, \ldots, N, \quad (18)$$

then at least one regular limit point of $\{\mathbf{x}_i[k]\}$ is a stationary solution of problem (1).
(ii) If

$$\lim_{k\to\infty} \left\|\widehat{\mathbf{x}}_i\left(\mathbf{x}_i\left[k\right]\right) - \mathbf{x}_i\left[k\right]\right\| = 0, \quad \forall i = 1, \ldots, N, \quad (19)$$

then every regular limit point of $\{\mathbf{x}_i[k]\}$ is a stationary solution of problem (1).

Proof. See Appendix (2). □

We are now in the position to give the convergence properties of Algorithm 1.

(1) Initialization: $\mathbf{x}_i[0] \in \mathcal{K}$, $\mathbf{y}_i[0] = \nabla f_i[0]$, $\widetilde{\pi}_i(\mathbf{x}_i[0]) = N \cdot \mathbf{y}_i[0] - \nabla f_i[0]$, for all $i = 1, \ldots, N$.
      Choose the weigh matrix $(a_{ij}[k])$ satisfied Assumption 1 and the step-size $\alpha[k]$ satisfied (20).
      Select a penalty parameter $c_i > 0$. Set $k := 0$;
(2) **repeat**
(3)   **for** each agent $i = 1, \ldots, N$ **do**
(4)   Local SCA optimization: compute $\widetilde{\mathbf{x}}_i[k]$ by solving problem (17);
(5)   Update $\mathbf{z}_i[k] = \mathbf{x}_i[k] + \alpha[k](\widetilde{\mathbf{x}}_i[k] - \mathbf{x}_i[k])$;
(6)   Consensus update:
        (a) $\mathbf{x}_i[k + 1] = \sum_{j=1}^{N} a_{ij}[k]\mathbf{z}_j[k]$;
        (b) $\mathbf{y}_i[k + 1] = \sum_{j=1}^{N} a_{ij}[k]\mathbf{y}_j[k] + (\nabla f_i(\mathbf{x}[k + 1]) - \nabla f_i(\mathbf{x}[k]))$;
        (c) $\widetilde{\pi}_i(\mathbf{x}_i[k + 1]) = N \cdot \mathbf{y}_i[k + 1] - \nabla f_i(\mathbf{x}[k + 1])$;
(7)   **end for**
(8)   Set $k := k + 1$;
(9) **until** $\mathbf{x}_i[k]$ satisfies a termination criterion.

ALGORITHM 1: Exact penalty function based distributed algorithm.

Theorem 7. *Let* $\{\mathbf{x}[k]\} = \{(\mathbf{x}_i[k])_{i=1}^{N}\}$ *be the sequence generated by Algorithm 1 and let* $\{\overline{\mathbf{x}}[k]\} = \{(1/N) \sum_{i=1}^{N} \mathbf{x}_i[k]\}$ *be the average. Suppose that (a) Assumptions 1, 2, and 5 hold; and (b) the step-size sequence* $\{\alpha[k]\}$ *is selected such that* $\alpha[k] \in (0, 1]$, *for all* k,

$$\sum_{k=0}^{\infty} \alpha[k] = \infty,$$

$$\sum_{k=0}^{\infty} \alpha[k]^2 < \infty. \tag{20}$$

Then, (i) the sequence $\{\overline{\mathbf{x}}[k]\}$ *is bounded and all its limit points are stationary solutions of problem (1); (ii) all the sequences* $\{\mathbf{x}_i[k]\}$ *asymptotically agree; that is,* $\|\mathbf{x}_i[k] - \overline{\mathbf{x}}[k]\| \rightarrow 0$ *as* $k \rightarrow \infty$, *for all* $i = 1, \ldots, N$.

Proof. See Appendix (3). □

Remark 8. (1) On the choice of surrogate functions, we only present several instances to show how to choose the surrogate function $\widetilde{f}_i$; also see [16, 19, 22, 23].

(i) If any convex structure of f_i is unavailable, the linearization of f_i at $\mathbf{x}_i[k]$ is the most plain choice; that is,

$$\widetilde{f}_i(\mathbf{x}_i; \mathbf{x}_i[k]) = f_i(\mathbf{x}_i[k]) + \nabla f_i(\mathbf{x}_i[k])^T (\mathbf{x}_i - \mathbf{x}_i[k]) + \frac{\tau_i}{2} \|\mathbf{x}_i - \mathbf{x}_i[k]\|^2. \tag{21}$$

(ii) If f_i is convex, one can just set

$$\widetilde{f}_i(\mathbf{x}_i; \mathbf{x}_i[k]) = f_i(\mathbf{x}_i[k]) + \frac{\tau_i}{2} \|\mathbf{x}_i - \mathbf{x}_i[k]\|^2. \tag{22}$$

(iii) Consider the case where f_i can be decomposed as $f_i(\mathbf{x}_i) = f_i^{(1)}(\mathbf{x}_i) + f_i^{(2)}(\mathbf{x}_i)$, where $f_i^{(1)}$ is convex and $f_i^{(2)}$ is

nonconvex. We can only linearize $f_i^{(2)}$ and preserve the convex $f_i^{(1)}$ as follows:

$$\begin{aligned} \widetilde{f}_i(\mathbf{x}_i; \mathbf{x}_i[k]) &= f_i^{(1)}(\mathbf{x}_i) + f_i^{(2)}(\mathbf{x}_i[k]) \\ &\quad + \frac{\tau_i}{2} \|\mathbf{x}_i - \mathbf{x}_i[k]\|^2 \\ &\quad + \nabla f_i^{(2)}(\mathbf{x}_i[k])^T (\mathbf{x}_i - \mathbf{x}_i[k]). \end{aligned} \tag{23}$$

(2) On the choice of step-sizes, condition (20) in Theorem 7 requires that the step-size sequence reduces to zero, but not too fast. The choice of the step-size $\alpha[k]$ meeting (20) is quite flexible [1, 23]. The following two choices of step-size are very effective in our simulations:

(i) $\alpha[k] = \alpha_0/(k + 1)^\beta$, $\alpha_0 > 0$, $0.5 < \beta \leq 1$;
(ii) $\alpha[k] = \alpha[k - 1](1 - \mu\alpha[k - 1])$, $\alpha[0] \in (0, 1]$, $\mu \in (0, 1)$.

(3) On the choice of weigh matrices, Assumption 1 requires that each communication weigh matrix $A[k]$ is doubly stochastic. References [4, 5] provided several choices of weigh matrices, such as the maximum degree weigh matrix, the Metropolis-Hastings weigh matrix, and the least-mean square consensus weight matrix.

5. Numerical Simulations

We consider a distributed nonconvex quadratic problem with quadratic inequality constraints (also see Example C in [22]):

$$\min_{\mathbf{x} \in \mathcal{K}} \quad F(\mathbf{x}) := \sum_{i=1}^{N} \left(\|\mathbf{A}_i\mathbf{x} - \mathbf{a}_i\|^2 - q_i \|\mathbf{x}\|^2 \right) \tag{24}$$

$$\text{s.t.} \quad \mathbf{x}^T \mathbf{B}\mathbf{x} - b \leq 0,$$

where $\mathbf{A}_i \in \mathbb{R}^{d \times d}$, $\mathbf{a}_i \in \mathbb{R}^d$, $\mathbf{B} \in \mathbb{R}^{d \times d}$, q_i and b are nonnegative constants, and $\mathcal{K}$ is a box constraint in $\mathbb{R}^d$. Note that the function $f_i(\mathbf{x}) := \|\mathbf{A}_i\mathbf{x} - \mathbf{a}_i\|^2 - q_i \|\mathbf{x}\|^2$ is locally accessible, only known by agent i. The local function f_i is differentiable, and its gradient ∇f_i is L_i-Lipschitz continuous with $L_i = \|(\mathbf{A}_i^T\mathbf{A}_i - q_i \cdot \mathbf{I})^T(\mathbf{A}_i^T\mathbf{A}_i - q_i \cdot \mathbf{I})\|$, where $\mathbf{I}$ is a unit matrix.

The corresponding penalized problem for problem (24) is given by

$$
\min_{\mathbf{x} \in \mathscr{K}} \quad F_c(\mathbf{x})
$$

$$
:= \sum_{i=1}^{N} \left(\|\mathbf{A}_i \mathbf{x} - \mathbf{a}_i\|^2 - q_i \|\mathbf{x}\|^2 \right) + c \tag{25}
$$

$$
\cdot \max \left\{ 0, \mathbf{x}^T \mathbf{B} \mathbf{x} - b \right\},
$$

where the penalty parameter $c > 0$ is chosen suitably. In general, there are two ways to choose c such that it satisfies the requirement. One way is by solving the unconstrained optimization problem defined in (12), and the other one is by heuristic. In order to imitate the time-varying weight matrix, a pool of 50 weight matrices from connected random graphs are generated, in which each weight matrix is satisfied (Assumption 1). The time-varying weight matrices required in Steps 6(a) and 6(b) of Algorithm 1 are randomly drawn from the above pool. To compare, we use the distributed Lagrangian primal-dual subgradient (DLPDS, for short) algorithm proposed in [14] to solve the corresponding sub-problem (9).

For simplicity, we assume that each $\mathbf{A}_i$ is a diagonal matrix with elements generated randomly in $(0, 1]$, $\mathbf{a}_i$ is a vector with elements generated randomly in $(0, 1]$, q_i is generated randomly in $[1, 2]$, $\mathbf{B}$ is a diagonal matrix with elements generated randomly in $(0, 1]$, b is generated randomly in $(0, 10]$, and the initial points $\mathbf{x}_i[0]$ are generated randomly in $\mathscr{K}$, where the box constraint is set as $\mathscr{K} = [-2, 2]^d$. In the numerical experiments, we heuristically select the penalty parameter $c = 10$.

Some experimental results are presented to illustrate the convergence behavior of the proposed Algorithm EPDA. All the curves are averaged over 20 independent realizations. Some comparisons with existing Algorithm DLPDS are also given.

Figures 1(a) and 2(a) depict the value of max error $\max_i \{\|\mathbf{x}_i[k] - \widetilde{\mathbf{x}}_i[k]\|/\|\widetilde{\mathbf{x}}_i[k]\|\}$ versus the number of iterations with different nodes and dimensions. It can be observed that both algorithms have the potential to converge to the same stationary solution. However, our Algorithm EPDA is much faster than Algorithm DLPDS [14]. As can be seen from Figure 1(a), Algorithm EPDA can reach higher precision than Algorithm DLPDS after 400 iterations. Figure 2(a) shows the similar results when $N = 100$ and $d = 10$.

Figures 1(b) and 2(b) depict the value of function $F(\mathbf{x}[k])$ versus the number of iterations with different nodes and dimensions. For two tested cases, the value of the objective function gradually decreases when the number of iterations increases, but the value of the objective function for Algorithm EPDA reduces faster than that of Algorithm DLPDS.

6. Conclusions

In this paper, a distributed algorithm was proposed to solve the nonconvex distributed optimization with global inequality constraints over time-varying multiagent networks. The proposed algorithm was based on the successive convex approximation technique, exact penalty function method, and dynamic averaging consensus. The convergence of the proposed algorithm was proved. Several numerical results showed the effectiveness of the proposed method.

Appendix

(1) Preliminaries. We present some preliminary results that will be used to prove our main results. For ease of notation, we need to give the following:

$$
\overline{\mathbf{x}}[k] = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i[k], \tag{A.1}
$$

$$
\nabla f_i^{av}[k] = \nabla f_i(\overline{\mathbf{x}}[k]), \tag{A.2}
$$

$$
\mathbf{y}_i^{av}[k+1] = \sum_{j \in \mathcal{N}_i^{in}[k]} a_{ij}[k] \mathbf{y}_i^{av}[k]
$$
$$
+ \left(\nabla f_i^{av}[k+1] - \nabla f_i^{av}[k] \right), \tag{A.3}
$$

$$
\widetilde{\pi}_i^{av}[k] = N \cdot \mathbf{y}_i^{av}[k] - \nabla f_i^{av}[k], \tag{A.4}
$$

$$
\widetilde{\mathbf{x}}_i^{av}[k] = \arg \min_{\mathbf{x}_i \in \mathscr{K}} \widetilde{U}_i \left(\mathbf{x}_i; \overline{\mathbf{x}}[k]; \widetilde{\pi}_i^{av}[k] \right), \tag{A.5}
$$

where $\mathbf{y}_i^{av}[0] = \nabla f_i^{av}[0]$.

Lemma A.1 (see [6]). *Let $\{Y[k]\}$, $\{X[k]\}$, and $\{Z[k]\}$ be three sequences of numbers such that $X[k] \geq 0$ for all $k > 0$. Suppose that*

$$
Y[k+1] \leq Y[k] - X[k] + Z[k], \quad \forall k = 1, 2, \ldots, \tag{A.6}
$$

and $\sum_{k=1}^{\infty} Z[k] < \infty$. Then, either $Y[k] \to -\infty$ or $\{Y[k]\}$ converges to a finite value and $\sum_{k=1}^{\infty} X[k] < \infty$.

Similar to the proof of Proposition 9 in [16], we can obtain the following lemma.

Lemma A.2. *Let $\{(\mathbf{x}_i[k])_{i=1}^{N}\}$ be the sequence generated by Algorithm 1. Then, for all $k > 0$ and for all $i = 1, \ldots, N$, the following results hold:*
   (i) Bounded disagreement:

$$
\|\widetilde{\mathbf{x}}_i[k] - \mathbf{x}_i[k]\| \leq m, \tag{A.7}
$$

where $m > 0$ is a finite constant.
   (ii) Asymptotic agreement on $\mathbf{x}_i[k]$:

$$
\lim_{k \to \infty} \|\mathbf{x}_i[k] - \overline{\mathbf{x}}[k]\| = 0. \tag{A.8}
$$

(iii) Asymptotically vanishing tracking error:

$$
\sum_{k=1}^{\infty} \alpha[k] \|\widetilde{\mathbf{x}}_i^{av}[k] - \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k])\| < \infty. \tag{A.9}
$$

(iv) Asymptotic agreement on best-responses:

$$
\sum_{k=1}^{\infty} \alpha[k] \|\widetilde{\mathbf{x}}_i[k] - \widetilde{\mathbf{x}}_i^{av}[k]\| < \infty. \tag{A.10}
$$

(a) Max error versus number of iterations

(b) Function value versus number of iterations
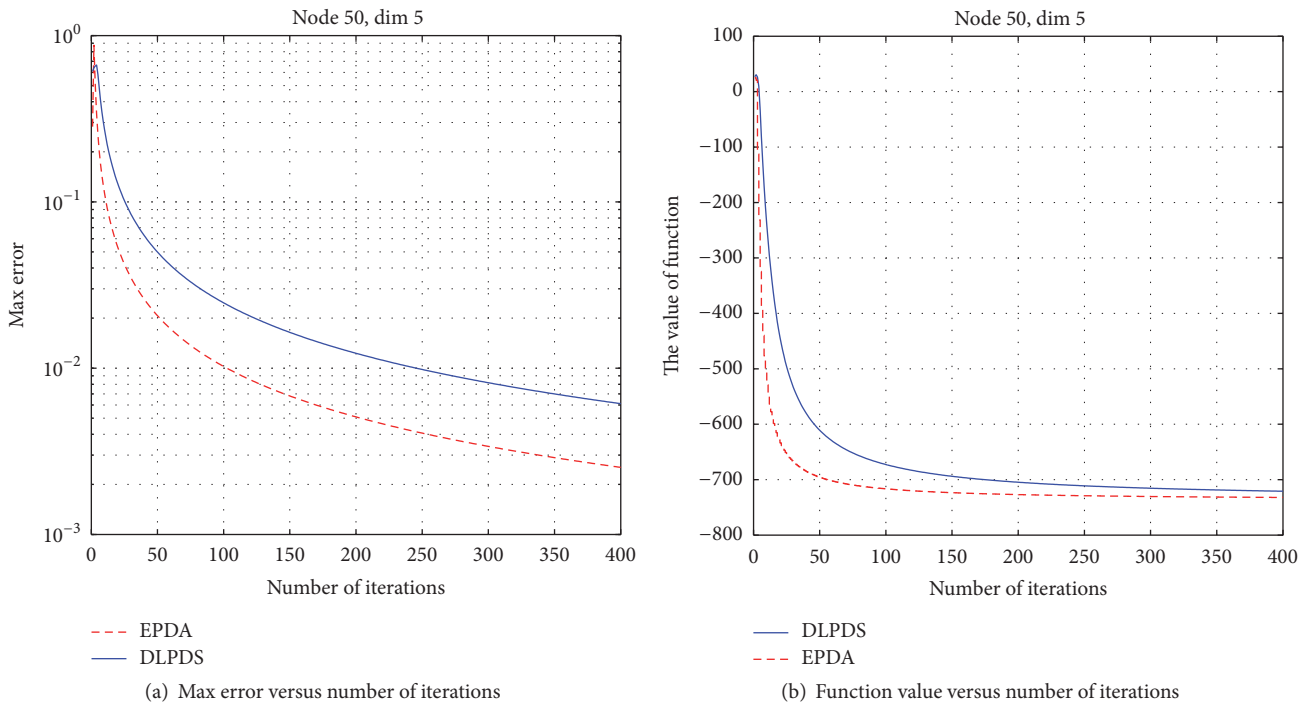
FIGURE 1: Numerical results for nodes of the network $N = 50$ and dimensions of the problem $d = 5$.

(a) Max error versus number of iterations

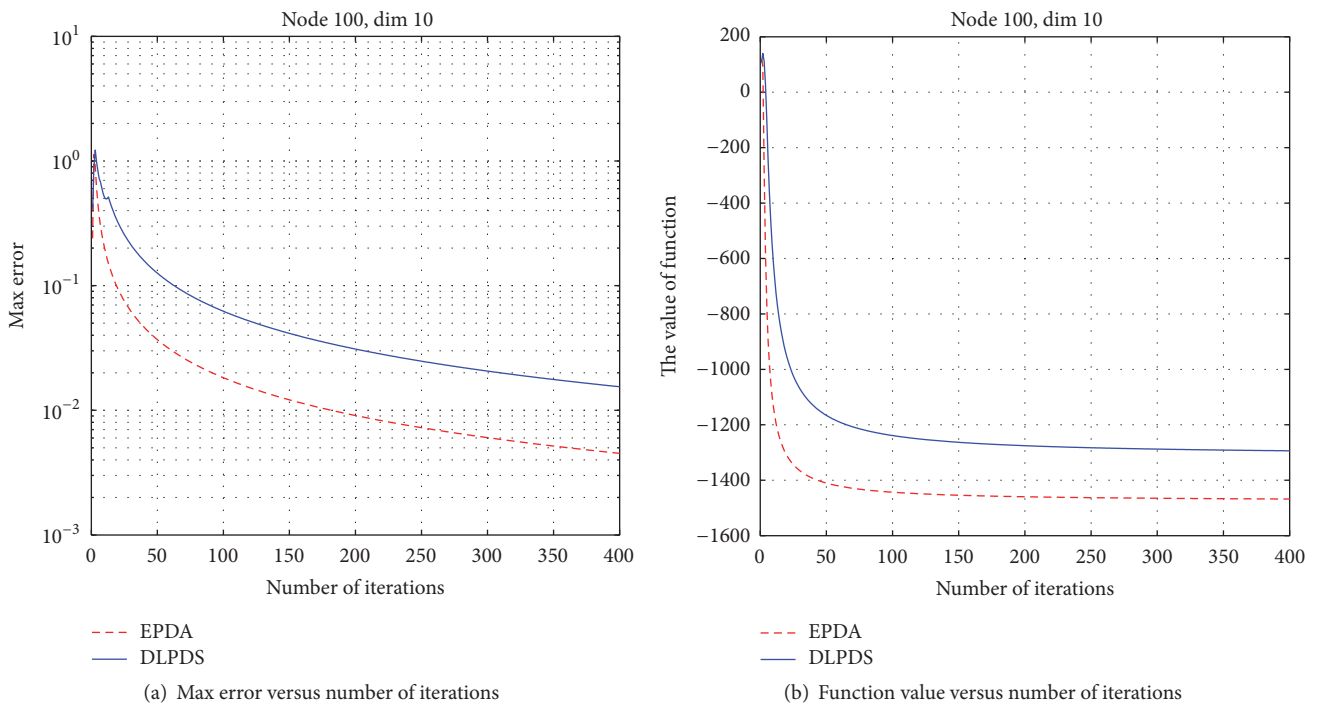(b) Function value versus number of iterations

FIGURE 2: Numerical results for nodes of the network $N = 100$ and dimensions of problem $d = 10$.

(2) Proof of Lemma 6. We only prove statement (i) of Lemma 6; then statement (ii) of Lemma 6 is proved by applying statement (i) to every convergence subsequence of $\{\mathbf{x}_i[k]\}$.

Assuming that $\overline{\mathbf{x}}_i$ is a regular accumulation point of subsequence $\{\mathbf{x}_i[k]\}_M$ of $\{\mathbf{x}_i[k]\}$ satisfied (18); therefore, there is $M' \subseteq M$ such that $\lim_{M' \ni k \to \infty} \mathbf{x}_i[k] = \overline{\mathbf{x}}_i$. Next, we prove that $\overline{\mathbf{x}}_i$ is a KKT point of problem (1). Let

$$\overline{I} = \left\{ s \in \{1, \ldots, p\} : g_s(\overline{\mathbf{x}}_i) = 0 \right\},$$

$$I^k = \left\{ s \in \{1, \ldots, p\} : g_s(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])) = 0 \right\}, \quad \text{(A.11)}$$

$$\text{for } k \in M'.$$

Using $\lim_{M' \ni k \to \infty} \|\widehat{\mathbf{x}}_i(\mathbf{x}_i[k]) - \mathbf{x}_i[k]\| = 0$ together with the continuity of g_s, we have

$$\lim_{M' \ni k \to \infty} g_s(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])) = g_s(\overline{\mathbf{x}}_i), \quad s = 1, \ldots, p. \quad \text{(A.12)}$$

The limit in (A.12) means that there is a positive integer $\widetilde{k} \in M'$ such that

$$I^k \subseteq \overline{I}, \quad \forall k \geq \widetilde{k}, \ k \in M'. \quad \text{(A.13)}$$

Since the functions $\nabla \widehat{F}_i$ and ∇g_s are continuous, by Assumption 5(C3), we obtain

$$\lim_{M' \ni k \to \infty} \nabla \widehat{F}_i(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k]); \mathbf{x}_i[k]) = \nabla \widehat{F}_i(\overline{\mathbf{x}}_i; \overline{\mathbf{x}}_i)$$

$$= \nabla F(\overline{\mathbf{x}}_i), \quad \text{(A.14)}$$

and, for $s = 1, \ldots, p$,

$$\lim_{M' \ni k \to \infty} \nabla g_s(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])) = \nabla g_s(\overline{\mathbf{x}}_i). \quad \text{(A.15)}$$

We now prove that, for sufficiently large $k \in M'$, the MFCQ holds at $\widehat{\mathbf{x}}_i(\mathbf{x}_i[k]) \in \chi$. By contradiction, suppose that the following relation does not hold for infinitely many $k \in M'$:

$$-\sum_{s \in I^k} \rho_s^k \nabla g_s(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])) \in N_{\mathscr{K}}(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])),$$

$$\rho_s^k \geq 0, \ \forall s \in I^k \Longrightarrow \rho_s^k = 0, \ \forall s \in I^k. \quad \text{(A.16)}$$

Then, there is a nonempty index set $\overline{\overline{I}} \subseteq \overline{I}$ such that, after a suitable remuneration, for each $k \in M'$, we have

$$-\sum_{s \in \overline{\overline{I}}} \rho_s^k \nabla g_s(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])) \in N_{\mathscr{K}}(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k]));$$

$$\rho_s^k \geq 0, \ \forall s \in \overline{\overline{I}}; \quad \text{(A.17)}$$

$$\sum_{s \in \overline{\overline{I}}} \rho_s^k = 1.$$

Without loss of generality, we assume that, for every $s \in \overline{\overline{I}}$, the sequence $\{\rho_s^k\}$ converges to a limit $\overline{\rho}_s$ such that $\sum_{s \in \overline{\overline{I}}} \overline{\rho}_s = 1$. By $\overline{\overline{I}} \subseteq \overline{I}$ and taking the limit $M' \ni k \to \infty$ in (A.17), and invoking (A.15) along with the outer semicontinuity of the mapping $N_{\mathscr{K}}(\cdot)$ (see Proposition 6.6 in [21]), we obtain

$$-\sum_{s \in \overline{\overline{I}}} \overline{\rho}_s \nabla g_s(\overline{\mathbf{x}}_i) \in N_{\mathscr{K}}(\overline{\mathbf{x}}_i); \quad \overline{\rho}_s \geq 0, \ \forall s \in \overline{\overline{I}};$$

$$\sum_{s \in \overline{\overline{I}}} \overline{\rho}_s = 1, \quad \text{(A.18)}$$

which is contradicted with the regularity of $\overline{\mathbf{x}}_i$. Thus, (A.16) must hold for sufficiently large $k \in M'$, implying that the KKT system of subproblem (9) has a solution for each sufficiently large $k \in M'$. Therefore, there exist $(\rho_s^k)_{s=1}^p$ such that

$$-\left[\nabla \widehat{F}_i(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k]); \mathbf{x}_i[k]) + \sum_{s=1}^p \rho_s^k \nabla g_s(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])) \right]$$

$$\in N_{\mathscr{K}}(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])) \quad \text{(A.19)}$$

$$0 \leq \rho_s^k \perp g_s(\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])) \leq 0, \ s = 1, \ldots, p.$$

From (A.13) and the complementarity slackness in (A.19), we have $\rho_s^k = 0$ for all $s \notin \overline{I}$ and large $k \in M'$. Moreover, the sequence of nonnegative multipliers $\{\rho^k = (\rho_s^k)_{s \in \overline{I}}\}_{k \in M'}$ must be bounded. By contradiction, assuming that $\lim_{M' \ni k \to +\infty} \|\rho^k\| = +\infty$ for some $\{\widehat{\mathbf{x}}_i(\mathbf{x}_i[k])\}_{M'}$, dividing both sides of (A.19) by $\|\rho^k\|$, and taking the limit $M' \ni k \to \infty$, we can obtain

$$-\sum_{s \in \overline{I}} \overline{\overline{\rho}}_s \nabla g_s(\overline{\mathbf{x}}_i) \in N_{\mathscr{K}}(\overline{\mathbf{x}}_i);$$

$$0 \leq \overline{\overline{\rho}}_s \perp g_s(\overline{\mathbf{x}}_i) \leq 0, \ s \in \overline{I}, \quad \text{(A.20)}$$

for some $\overline{\overline{\rho}} = (\overline{\overline{\rho}}_s)_{s \in \overline{I}} \neq \mathbf{0}$, in contradiction with Definition 3.

Therefore, the sequence $\{\rho^k = (\rho_s^k)_{s \in \overline{I}}\}_{k \in M'}$ must have a limit. Let $(\overline{\rho}_s)_{s \in \overline{I}}$ be such a limit (after a suitable remuneration). Taking the limit $M' \ni k \to \infty$ in (A.19), and combining (A.14) and (A.15) along with the outer semicontinuity of the mapping $N_{\mathscr{K}}(\cdot)$, we have

$$-\left[\nabla F(\overline{\mathbf{x}}_i) + \sum_{s \in \overline{I}} \overline{\rho}_s \nabla g_s(\overline{\mathbf{x}}_i) \right] \in N_{\mathscr{K}}(\overline{\mathbf{x}}_i);$$

$$0 \leq \overline{\rho}_s \perp g_s(\overline{\mathbf{x}}_i) \leq 0, \ s \in \overline{I}. \quad \text{(A.21)}$$

By (A.21), $\overline{\mathbf{x}}_i$ is a stationary solution of problem (1).

(3) Proof of Theorem 7. Using the descent lemma on F, Steps 5 and 6(a) of Algorithm 1 and (A.1), we obtain

$$F(\overline{\mathbf{x}}[k+1])$$

$$\leq F(\overline{\mathbf{x}}[k])$$

$$+ \frac{\alpha[k]}{N} \nabla F(\overline{\mathbf{x}}[k])^T \sum_{i=1}^N (\widetilde{\mathbf{x}}_i[k] - \overline{\mathbf{x}}[k]) \quad \text{(A.22)}$$

$$+ \frac{L_{\max}}{2N} (\alpha[k])^2 \sum_{i=1}^N \|\widetilde{\mathbf{x}}_i[k] - \mathbf{x}_i[k]\|^2.$$

Summing and subtracting $(\alpha[k]/N)\nabla F(\overline{\mathbf{x}}[k])^T \sum_{i=1}^N (\widetilde{\mathbf{x}}_i^{av}[k] + \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]))$ to the right side of inequality (A.22), we have

$$F(\overline{\mathbf{x}}[k+1]) \leq F(\overline{\mathbf{x}}[k])$$
$$+ \frac{\alpha[k]}{N} \nabla F(\overline{\mathbf{x}}[k])^T \sum_{i=1}^{N} \left(\widetilde{\mathbf{x}}_i^{av}[k] - \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) \right) + \frac{\alpha[k]}{N} \nabla F(\overline{\mathbf{x}}[k])^T \sum_{i=1}^{N} \left(\widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k] \right)$$
$$+ \frac{\alpha[k]}{N} \nabla F(\overline{\mathbf{x}}[k])^T \sum_{i=1}^{N} \left(\widetilde{\mathbf{x}}_i[k] - \widetilde{\mathbf{x}}_i^{av}[k] \right) + \frac{L_{\max}}{2N} (\alpha[k])^2 \sum_{i=1}^{N} \left\| \widetilde{\mathbf{x}}_i[k] - \mathbf{x}_i[k] \right\|^2. \tag{A.23}$$

Next we evaluate the third term on the right side of (A.23). For a given $\overline{\mathbf{x}}[k] \in \chi$, let $\widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k])$ be the unique solution of subproblem (9). By the optimal condition of subproblem (9), we have, for all $\mathbf{y}_i \in \chi$,

$$\left(\mathbf{y}_i - \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) \right)^T \nabla \widehat{F}_i \left(\widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]); \overline{\mathbf{x}}[k] \right) \geq 0. \tag{A.24}$$

Adding and subtracting $\nabla \widehat{F}_i(\mathbf{y}_i; \overline{\mathbf{x}}[k])$, and using Assumption 5(C3), we get

$$\left(\mathbf{y}_i - \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) \right)^T \nabla \widehat{F}_i \left(\mathbf{y}_i; \overline{\mathbf{x}}[k] \right) \geq \left(\widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \mathbf{y}_i \right)^T$$
$$\cdot \left(\nabla \widehat{F}_i \left(\widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]); \overline{\mathbf{x}}[k] \right) - \nabla \widehat{F}_i \left(\mathbf{y}_i; \overline{\mathbf{x}}[k] \right) \right) \tag{A.25}$$
$$\geq \tau_i \left\| \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \mathbf{y}_i \right\|^2,$$

where the second inequality of (A.25) is obtained by Assumption 5(C1) and Lemma 7 in [22].

Letting $\mathbf{y}_i = \overline{\mathbf{x}}[k]$ in (A.25) and noting that

$$\nabla \widehat{F}_i(\overline{\mathbf{x}}[k]; \overline{\mathbf{x}}[k]) = \nabla \widetilde{f}_i(\overline{\mathbf{x}}[k]) + \pi(\overline{\mathbf{x}}[k])$$
$$= \nabla f_i(\overline{\mathbf{x}}[k]) + \pi(\overline{\mathbf{x}}[k]) \tag{A.26}$$
$$= \nabla F(\overline{\mathbf{x}}[k]),$$

we have, for $i = 1, \ldots, N$,

$$\left(\overline{\mathbf{x}}[k] - \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) \right)^T \nabla F(\overline{\mathbf{x}}[k])$$
$$\geq \tau \left\| \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k] \right\|^2, \tag{A.27}$$

where $\tau = \min_i \tau_i$. Summing inequality (A.27) from $i = 1$ to N, we can obtain the following relation:

$$\frac{\alpha[k]}{N} \nabla F(\overline{\mathbf{x}}[k])^T \sum_{i=1}^{N} \left(\widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k] \right)$$
$$\leq -\frac{\tau \alpha[k]}{N} \sum_{i=1}^{N} \left\| \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k] \right\|^2. \tag{A.28}$$

By (A.23), (A.28), (A.7), Assumption 2(B4), and the triangle inequalities, we can get

$$F(\overline{\mathbf{x}}[k+1]) \leq F(\overline{\mathbf{x}}[k])$$
$$- \frac{\tau \alpha[k]}{N} \sum_{i=1}^{N} \left\| \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k] \right\|^2$$
$$+ \frac{m^2 L_{\max}}{2} \alpha[k]^2 \tag{A.29}$$
$$+ \frac{L_F}{N} \alpha[k] \sum_{i=1}^{N} \left\| \widetilde{\mathbf{x}}_i^{av}[k] - \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) \right\|$$
$$+ \frac{L_F}{N} \alpha[k] \sum_{i=1}^{N} \left\| \widetilde{\mathbf{x}}_i[k] - \widetilde{\mathbf{x}}_i^{av}[k] \right\|.$$

Let

$$Y[k] = F(\overline{\mathbf{x}}[k]),$$
$$X[k] = \frac{\tau \alpha[k]}{N} \sum_{i=1}^{N} \left\| \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k] \right\|^2,$$
$$Z[k] = \frac{L_F}{N} \alpha[k] \sum_{i=1}^{N} \left\| \widetilde{\mathbf{x}}_i^{av}[k] - \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) \right\| \tag{A.30}$$
$$+ \frac{L_F}{N} \alpha[k] \sum_{i=1}^{N} \left\| \widetilde{\mathbf{x}}_i[k] - \widetilde{\mathbf{x}}_i^{av}[k] \right\|$$
$$+ \frac{m^2 L_{\max}}{2} \alpha[k]^2.$$

By (A.9), (A.10), (20), and (A.7), we have $\sum_{k=1}^{\infty} Z[k] < \infty$. Since $F(\overline{\mathbf{x}}[k])$ is coercive, it follows from Lemma A.1 that $\{F(\overline{\mathbf{x}}[k])\}$ is convergent to a finite value, and

$$\sum_{k=1}^{\infty} \alpha[k] \left\| \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k] \right\|^2 < \infty, \quad \forall i = 1, \ldots, N. \tag{A.31}$$

Combining the inequality above and (20), we have

$$\liminf_{k \to \infty} \left\| \widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k] \right\| = 0, \quad \forall i = 1, \ldots, N. \tag{A.32}$$

By a similar argument as in [22], we can obtain that $\limsup_{k \to \infty} \|\widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k]\| = 0, \forall i = 1, \ldots, N$. Thus, we have $\lim_{k \to \infty} \|\widehat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k]\| = 0, \forall i = 1, \ldots, N$. By Assumption 2(B5) and the convergence of $\{F(\overline{\mathbf{x}}[k])\}$, the

sequence $\{\overline{\mathbf{x}}[k]\}$ is bounded. Furthermore, the sequence of $\{\overline{\mathbf{x}}[k]\}$ has at least one limit point $\overline{\mathbf{x}}^{\infty} \in \chi$. Owing to the continuity of any $\hat{\mathbf{x}}_i(\cdot)$ and $\lim_{k \to \infty}\|\hat{\mathbf{x}}_i(\overline{\mathbf{x}}[k]) - \overline{\mathbf{x}}[k]\| = 0$, $\hat{\mathbf{x}}_i(\overline{\mathbf{x}}^{\infty}) = \overline{\mathbf{x}}^{\infty}$ must exist for all i; therefore, $\overline{\mathbf{x}}^{\infty}$ is a stationary solution of problem (1) by (i) of Lemma 6. Thus, statement (i) of the theorem is proved, and statement (ii) readily follows from (A.8).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods; Athena Scientific*, MA, Belmont, 1997.

[2] M. J. Park, O. M. Kwon, S. G. Choi, and E. J. Cha, "Consensus protocol design for discrete-time networks of multiagent with time-varying delay via logarithmic quantizer," *Complexity*, vol. 21, no. 1, pp. 163–176, 2015.

[3] M. Sathishkumar, R. Sakthivel, P. Selvaraj, and S. Marshal anthoni, "Robust reliable dissipative control of nonlinear networked control systems," *Complexity*, vol. 21, pp. 427–437, 2016.

[4] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," *Journal of Parallel and Distributed Computing*, vol. 67, no. 1, pp. 33–46, 2007.

[5] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[6] Y. A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[7] S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, 2010.

[8] I. Lobel and A. Ozdaglar, "Distributed subgradient methods for convex optimization over random networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1291–1306, 2011.

[9] J. Li, C. Wu, Z. Wu, and Q. Long, "Gradient-free method for nonsmooth distributed optimization," *Journal of Global Optimization*, vol. 61, no. 2, pp. 325–340, 2014.

[10] J. Li, Z. Wu, C. Wu, Q. Long, and X. Wang, "An inexact dual fast gradient-projection method for separable convex optimization with linear coupled constraints," *Journal of Optimization Theory and Applications*, vol. 168, no. 1, pp. 153–171, 2016.

[11] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.

[12] J. Li, G. Chen, Z. Dong, Z. Wu, and M. Yao, "Distributed mirror descent method for saddle point problems over directed graphs," *Complexity*, vol. 21, pp. 178–190, 2016.

[13] J. C. Duchi, A. Agarwal, and M. J. Wainwright, "Dual averaging for distributed optimization: convergence analysis and network scaling," *IEEE Transactions on Automatic Control*, vol. 57, no. 3, pp. 592–606, 2012.

[14] M. Zhu and S. Martinez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.

[15] D. Yuan, S. Xu, and H. Zhao, "Distributed primal-dual subgradient method for multiagent optimization via consensus algorithms," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 6, pp. 1715–1724, 2011.

[16] P. D. Lorenzo and G. Scutari, "NEXT: In-Network Nonconvex Optimization," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 2, no. 2, pp. 120–136, 2016.

[17] G. Scutari, F. Facchinei, and L. Lampariello, "Parallel and Distributed Methods for Constrained Nonconvex Optimization - Part I: Theory," *IEEE Transactions on Signal Processing*, vol. 65, no. 8, pp. 1929–1944, 2017.

[18] J. Li, C. Wu, Z. Wu, Q. Long, and X. Wang, "Distributed proximal-gradient method for convex optimization with inequality constraints," *Anziam Journal*, vol. 56, no. 2, pp. 160–178, 2014.

[19] Y. Sun, G. Scutari, and D. Palomar, "Distributed nonconvex multiagent optimization over time-varying networks," in *Proceedings of the 50th Asilomar Conference on Signals, Systems and Computers, ACSSC 2016*, pp. 788–794, USA, November 2016.

[20] D. P. Bertsekas, *Nonlinear Programming*, Belmont, Mass., Athena Scientific, USA, 1995.

[21] R. T. Rockafellar and R. J. B. Wets, *Variational analysis*, Springer Science Business Media, 2009.

[22] F. Facchinei, G. Scutari, and S. Sagratella, "Parallel selective algorithms for nonconvex big data optimization," *IEEE Transactions on Signal Processing*, vol. 63, no. 7, pp. 1874–1889, 2015.

[23] G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 641–655, 2014.

[24] D. P. Bertsekas, "Necessary and sufficient condition for a penalty method to be exact," *Mathematical Programming*, vol. 9, no. 1, pp. 87–99, 1975.

[25] M. Zhu and S. Martnez, "Discrete-time dynamic average consensus," *Automatica*, vol. 46, no. 2, pp. 322–329, 2010.

Advances in
Operations Research

Advances in
Decision Sciences

Journal of
Applied Mathematics

Algebra

Journal of
Probability and Statistics

The Scientific
World Journal

International Journal of
Differential Equations

International
Journal of
Combinatorics

Hindawi

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

Advances in
Mathematical Physics

Journal of
Complex Analysis

Journal of
Mathematics

Mathematical Problems
in Engineering

Abstract and
Applied Analysis

Discrete Dynamics in
Nature and Society

International
Journal of
Mathematics and
Mathematical
Sciences

Journal of
Discrete Mathematics

Journal of
Function Spaces

International Journal of
Stochastic Analysis

Journal of
Optimization