

Distributed Nonsmooth Optimization with Coupled Inequality Constraints via Modified Lagrangian Function

Shu Liang, Xianlin Zeng, and Yiguang Hong

Abstract

This technical note considers a distributed convex optimization problem with nonsmooth cost functions and coupled nonlinear inequality constraints. To solve the problem, we first propose a modified Lagrangian function containing local multipliers and a nonsmooth penalty function. Then we construct a distributed continuous-time algorithm by virtue of a projected primal-dual subgradient dynamics. Based on the nonsmooth analysis and Lyapunov function, we obtain the existence of the solution to the nonsmooth algorithm and its convergence.

Index Terms

Distributed optimization, coupled constraint, modified Lagrangian function, primal-dual dynamics, nonsmooth analysis.

I. INTRODUCTION

Distributed convex optimization has attracted intense research attention in recent years, due to its theoretic significance and broad applications in many research fields such as sensor networks, smart grids and social networks. Various models of distributed optimization have been proposed and studied in the literature. Most works have focused on consensus-based formulations, where each agent estimates the entire optimal solution via plentiful discrete-time algorithms (e.g., see [1], [2] and the references therein). Recently, more and more effort has also been done for distributed continuous-time algorithms

S. Liang, X. Zeng and Y. Hong are with the Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190, China. e-mails: shuliang@amss.ac.cn (S. Liang), xianlin.zeng@amss.ac.cn (X. Zeng), yghong@iss.ac.cn (Y. Hong).

(see [3]–[6] for instance), partly due to the development of its hardware implementation [7] and flexible application in continuous-time physical systems [8].

Here we consider distributed optimizations with separable cost functions and coupled constraints. In the presence of a coupled constraint, the feasible region of one agent’s decision variable is influenced by some other agents’ decision variables. If such a constraint is known by all the related agents, various algorithms were obtained, such as dual gradient algorithms [9], [10], primal-dual algorithms [11]–[13], the saddle-point-like algorithm [14], and the distributed Newton-type algorithm [15]. However, coupled constraints may not be available to each agent in practice, and then the aforementioned algorithms may not work if there is no central coordinator in the network. To deal with the challenges, [16], [17] developed distributed initialization-free algorithms for the optimal resource allocation, while [18] proposed a distributed algorithm for the extended monotropic optimization. Note that [16]–[18] considered coupled equality constraints. Moreover, [19] proposed a distributed algorithm for coupled inequality constraints, based on the average consensus technique to estimate the constraint functions along with a local primal-dual perturbed subgradient method. These distributed algorithms adopted local dynamics to evaluate the optimal dual solution instead of the original centralized one. On the other hand, all of them have to further employ auxiliary dynamics in order to guarantee the correctness and convergence, whereas the distributed design may become quite complicated in the case with coupled inequality constraints.

The objective of this note is to develop a distributed algorithm for nonsmooth convex optimization with coupled inequality constraints. We propose a modified Lagrangian function such that not only its saddle point yields the correct optimal solution to the original problem, but also its primal-dual subgradient dynamics is fully distributed. Particularly, we introduce local multipliers to decouple the constraints and employ a nonsmooth penalty function for the correctness. Based on this modified Lagrangian function, we propose a continuous-time projected subgradient algorithm for saddle-point computation. Our algorithm is fully distributed since each agent updates its local variables according to its local data and the information of its neighbors, without requiring any center in the network. Moreover, our algorithm only involves the primal variables and local multipliers, which yields a lower order dynamics than those in existing algorithms.

The rest organization is as follows: Section 2 provides necessary preliminaries, while Section 3 formulates the problem. Then Section 4 presents the main results to prove the convergence of our nonsmooth algorithm, and Section 5 gives two numerical examples. Finally, Section 6 gives some concluding remarks.

Notations: Denote \mathbb{R}^n as the n -dimensional real vector space and \mathbb{R}_+^n as the positive orthant in \mathbb{R}^n . Denote $\mathbf{0}$ as a vector with each component being zero. For a vector $a \in \mathbb{R}^n$, $a \leq \mathbf{0}$ (or $a < \mathbf{0}$) means

that each component of a is less than or equal to zero (or less than zero). Denote $\|\cdot\|$ and $|\cdot|$ as the ℓ_2 -norm and ℓ_1 -norm for vectors, respectively. Denote $\text{col}(x_1, \dots, x_N) = (x_1^T, \dots, x_N^T)^T$ as the column vector stacked with column vectors x_1, \dots, x_N . For a set $\Omega \subset \mathbb{R}^n$, $\text{rint}(\Omega)$ is the relative interior and $d_\Omega(x) \triangleq \inf_{y \in \Omega} \|y - x\|$ is the distance function between point x and set Ω .

II. PRELIMINARIES

In this section, we introduce relevant preliminary knowledge about convex analysis, differential inclusions, and graph theory.

A set $C \subseteq \mathbb{R}^n$ is *convex* if $\lambda z_1 + (1 - \lambda)z_2 \in C$ for any $z_1, z_2 \in C$ and $\lambda \in [0, 1]$. For $x \in C$, the *tangent cone* to C at x , denoted by $\mathcal{T}_C(x)$, is defined as

$$\mathcal{T}_C(x) \triangleq \left\{ \lim_{k \rightarrow \infty} \frac{x_k - x}{t_k} \mid x_k \in C, t_k > 0, \text{ and } x_k \rightarrow x, t_k \rightarrow 0 \right\},$$

while the *normal cone* to C at x , denoted by $\mathcal{N}_C(x)$, is defined as

$$\mathcal{N}_C(x) \triangleq \{v \in \mathbb{R}^n \mid v^T(y - x) \leq 0, \text{ for all } y \in C\}.$$

A projection operator is defined as $P_C(z) \triangleq \text{argmin}_{x \in C} \|x - z\|$, and an operator that projects a point $z \in \mathbb{R}^n$ (or a set) onto the tangent cone $\mathcal{T}_C(x)$ is $\Pi_C(x; z) \triangleq P_{\mathcal{T}_C(x)}(z)$.

A function $f : C \rightarrow \mathbb{R}$ is said to be *convex* (or *strictly convex*) if $f(\lambda z_1 + (1 - \lambda)z_2) \leq$ (or $<$) $\lambda f(z_1) + (1 - \lambda)f(z_2)$ for any $z_1, z_2 \in C, z_1 \neq z_2$ and $\lambda \in (0, 1)$. A function g is said to be a *concave function* if $-g$ is a convex function.

A set-valued map \mathcal{F} from \mathbb{R}^n to \mathbb{R}^n is a map associated with any $x \in \mathbb{R}^n$ a subset $\mathcal{F}(x)$ of \mathbb{R}^n . \mathcal{F} is said to be *upper semicontinuous* at $x_0 \in \mathbb{R}^n$ if for any open set E containing $\mathcal{F}(x_0)$, there exists a neighborhood D of x_0 such that $\mathcal{F}(D) \subset E$. We say that \mathcal{F} is upper semicontinuous if it is so at every $x_0 \in \mathbb{R}^n$. The graph of \mathcal{F} , denoted by $\text{gph } \mathcal{F}$, is the set consists of all pairs (x, y) satisfying $y \in \mathcal{F}(x)$.

A differential inclusion can be expressed as follows:

$$\dot{x} \in \mathcal{F}(x), \quad x(0) = x_0. \tag{1}$$

A map $x(t) : [0, +\infty) \rightarrow \mathbb{R}^n$ is said to be a *solution* to (1) if it is absolutely continuous and satisfies the inclusion for almost all $t \in [0, +\infty)$.

A graph of a network is denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \dots, N\}$ is a set of nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of edges. Node j is said to be a *neighbor* of node i if $\{i, j\} \in \mathcal{E}$. The set of all the neighbors of node i is denoted by \mathcal{N}_i . \mathcal{G} is said to be *undirected* if $(i, j) \in \mathcal{E} \Leftrightarrow (j, i) \in \mathcal{E}$. A path of \mathcal{G} is a sequence of distinct nodes where any pair of consecutive nodes in the sequence has an edge of \mathcal{G} . Node j is said

to be *connected* to node i if there is a path from j to i . \mathcal{G} is said to be connected if any two nodes are connected. The detailed knowledge about graph theory can be found in [20].

The following lemma collects some results given in [21] that will be used in our analysis.

Lemma 2.1: Let $f : X \rightarrow \mathbb{R}$ be locally Lipschitz continuous and let $\Omega \subset X$ be a closed convex subset, where $X \subset \mathbb{R}^n$. Then the following statements hold.

- 1) $\text{gph } \partial f$ is closed.
- 2) If f is convex, then $x^* \in \text{argmin}_{x \in \Omega} f(x)$ if and only if $0 \in \partial f(x^*) + \mathcal{N}_\Omega(x^*)$.
- 3) Suppose f has a Lipschitz constant K_0 on an open set that contains Ω . When $K > K_0$, $x^* \in \text{argmin}_{x \in \Omega} f(x)$ if and only if $x^* \in \text{argmin}_{x \in X} f(x) + K d_\Omega(x)$.

A collection of results in [22] with respect to set-valued maps and differential inclusions are given below.

Lemma 2.2: The following statements hold.

- 1) A set-valued map \mathcal{F} from \mathbb{R}^n to \mathbb{R}^n is upper semicontinuous if it has compact values and $\text{gph } \mathcal{F}$ is closed.
- 2) Let \mathcal{F}_0 from \mathbb{R}^n to \mathbb{R}^n be a set-valued map and $C \subset \mathbb{R}^n$ be a closed convex subset. Consider the following two differential inclusions

$$\dot{x} \in \mathcal{F}_0(x) - \mathcal{N}_C(x), \quad x(0) = x_0 \in C, \quad (2)$$

$$\dot{x} \in \Pi_C(x, \mathcal{F}_0(x)), \quad x(0) = x_0 \in C. \quad (3)$$

Then $x(\cdot)$ is a solution to (2) if and only if it is a solution to (3).

- 3) For any $x_0 \in C$, there is a solution to the differential inclusion (2) if \mathcal{F}_0 is upper semicontinuous and C is compact and convex.

Moreover, we introduce a lemma from [6, Lemma 4.2] which will be used in the convergence analysis.

Lemma 2.3: Let $x(t)$ be a solution to the differential inclusion (1). If z is a Lyapunov stable equilibrium of (1) and is also a cluster point of $x(\cdot)$, then $\lim_{t \rightarrow +\infty} x(t) = z$.

III. PROBLEM FORMULATION

Consider a multi-agent network with N agents, whose label set is denoted as $\mathcal{V} = \{1, \dots, N\}$, cooperating over a graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. For each agent i , there are a local decision variable $x_i \in \mathbb{R}^{n_i}$ and a local constraint set $\Omega_i \subset \mathbb{R}^{n_i}$ for $i \in \mathcal{V}$. Define $\mathbf{x} \triangleq \text{col}(x_1, \dots, x_N)$, and define the total cost function of the network as $f(\mathbf{x}) \triangleq \sum_{i \in \mathcal{V}} f_i(x_i)$, where $f_i : \Omega_i \rightarrow \mathbb{R}$ is a (nonsmooth) local cost function of agent i . In addition, the agents are subject to coupled inequality constraints in the form of $\mathbf{g}(\mathbf{x}) \triangleq \sum_{i \in \mathcal{V}} \mathbf{g}_i(x_i) \leq \mathbf{0}$,

where $\mathbf{g}_i : \Omega_i \rightarrow \mathbb{R}^M$ are continuous mappings for $i \in \mathcal{V}$ (that is, $\mathbf{g}_i = (g_{i1}, \dots, g_{iM})^T$ and $g_{ik} : \Omega_i \rightarrow \mathbb{R}$ are continuous functions for all i and k). To be strict, the optimization problem can be formulated as:

$$\min_{\mathbf{x} \in \Omega} f(\mathbf{x}), \text{ s.t. } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \quad (4)$$

where $\Omega \triangleq \prod_{i \in \mathcal{V}} \Omega_i \subset \mathbb{R}^{n_1 + \dots + n_N}$ denotes the local constraints of N agents.

The following assumption is needed to ensure the well-posedness of problem (4).

Assumption 3.1:

- 1) (Convexity and continuity) For all $i \in \mathcal{V}$, Ω_i is compact and convex. On an open set containing Ω_i , f_i is strictly convex and \mathbf{g}_i is convex, and f_i and \mathbf{g}_i are locally Lipschitz continuous.
- 2) (Slater's constraint qualification) There exists $\bar{\mathbf{x}} \in \text{rint}(\Omega)$ such that $\mathbf{g}(\bar{\mathbf{x}}) < \mathbf{0}$.
- 3) (Communication topology) The graph \mathcal{G} is connected and undirected.

This assumption is quite mild and similar ones are widely used in the literature (e.g., [17]).

The goal of this note is to develop a *distributed continuous-time algorithm* for solving (4) with each agent communicating with their neighbors. Moreover, for every $i \in \mathcal{V}$, agent i can only access $\mathbf{g}_i(x_i)$ rather than $\mathbf{g}(\mathbf{x})$.

The differences between our problem and those in existing literature are as follows.

- The decision variables can be heterogeneous with possibly different dimensions, in contrast to those consensus-based models.
- The cost and constraint functions can be nonsmooth, while some projected dynamics [12], [13], [17] and Newton type method [15] depend on smoothness.
- The coupled constraints may be unavailable to local agents, different from [14]. Also, coupled inequality constraints are considered as in [19], different from the coupled (affine) equality ones studied in [16]–[18].

IV. MAIN RESULTS

In this section, we first propose a modified Lagrangian function and then propose a distributed continuous-time algorithm for the considered optimization problem. Moreover, we prove the existence of the solution to the nonsmooth algorithm along with the discussions on its convergence and the rate.

A. Lagrangian Function and Distributed Algorithm Design

Consider the following dual problem with respect to the primal one (4),

$$\max_{\lambda \geq \mathbf{0}} q(\lambda), \quad q(\lambda) \triangleq \min_{\mathbf{x} \in \Omega} \mathcal{L}(\mathbf{x}, \lambda), \quad (5)$$

where $\mathcal{L} : \Omega \times \mathbb{R}_+^M \rightarrow \mathbb{R}$ is the Lagrangian function defined as

$$\mathcal{L}(\mathbf{x}, \lambda) \triangleq f(\mathbf{x}) + \lambda^T \mathbf{g}(\mathbf{x}). \quad (6)$$

It has been shown in [9] that the optimal dual solution λ^* of (5) lies in a compact set $\mathcal{D} \subset \mathbb{R}_+^M$, given by

$$\mathcal{D} \triangleq \{\lambda \in \mathbb{R}_+^M \mid \|\lambda\| \leq \frac{f(\bar{\mathbf{x}}) - \tilde{q}}{\gamma}\}, \quad (7)$$

where $\bar{\mathbf{x}}$ is a Slater point of (4), $\tilde{q} = \min_{\mathbf{x} \in \Omega} \mathcal{L}(\mathbf{x}, \tilde{\lambda})$ is a dual function value for an arbitrary $\tilde{\lambda} \geq 0$, $\gamma = \min_{k=1, \dots, M} \{-\sum_{i \in \mathcal{V}} g_{ik}(\bar{x}_i)\}$.

We present the following lemma, which is a well-known convex optimization result [23].

Lemma 4.1: Under Assumption 3.1, the following statements are equivalent:

- 1) (Primal-dual characterization) $(\mathbf{x}^*, \lambda^*)$ is a primal-dual solution pair of problems (4) and (5).
- 2) (Saddle-point characterization) $(\mathbf{x}^*, \lambda^*)$ is a *saddle point* of Lagrangian function (6), that is,

$$\mathcal{L}(\mathbf{x}^*, \lambda) \leq \mathcal{L}(\mathbf{x}^*, \lambda^*) \leq \mathcal{L}(\mathbf{x}, \lambda^*), \quad \forall \mathbf{x} \in \Omega, \lambda \geq \mathbf{0}.$$

- 3) (KKT characterization) $(\mathbf{x}^*, \lambda^*)$ satisfies

$$\mathbf{0} \in \partial f(\mathbf{x}^*) + \partial \mathbf{g}(\mathbf{x}^*) \lambda^* + \mathcal{N}_\Omega(\mathbf{x}^*), \quad \mathbf{0} \leq \lambda^* \perp -\mathbf{g}(\mathbf{x}^*) \geq \mathbf{0}.$$

- 4) (Minimax characterization) $(\mathbf{x}^*, \lambda^*)$ is a solution of the minimax problem

$$\min_{\mathbf{x} \in \Omega} \{\max_{\lambda \in \mathcal{D}} \mathcal{L}(\mathbf{x}, \lambda)\}.$$

Moreover, since $f(\mathbf{x})$ is strictly convex, such \mathbf{x}^* is unique while λ^* may not be unique.

A centralized projected primal-dual algorithm with respect to $\mathcal{L}(\mathbf{x}, \lambda)$ can be written as (referring to [13]):

$$\begin{cases} \dot{\mathbf{x}} \in \Pi_\Omega(\mathbf{x}, -\partial f(\mathbf{x}) - \partial \mathbf{g}(\mathbf{x}) \lambda), & \mathbf{x}(0) \in \Omega, \\ \dot{\lambda} \in \Pi_{\mathbb{R}_+^M}(\lambda, \mathbf{g}_1(x_1) + \dots + \mathbf{g}_N(x_N)), & \lambda(0) \in \mathbb{R}_+^M, \end{cases}$$

which needs a center to broadcast λ and gather $\mathbf{g}_1, \dots, \mathbf{g}_N$ for the update. In order to develop fully distributed algorithms without a center, we employ local multipliers and a nonsmooth penalty function to construct a modified Lagrangian function. To be specific, define

$$\boldsymbol{\lambda} \triangleq \text{col}(\lambda_1, \dots, \lambda_N) \in \mathbb{R}_+^{MN}, \quad (8a)$$

$$\mathcal{S} \triangleq \{\boldsymbol{\lambda} \in \mathbb{R}_+^{MN} \mid \lambda_1 = \dots = \lambda_N\}, \quad (8b)$$

$$\phi(\boldsymbol{\lambda}) \triangleq \frac{1}{2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} |\lambda_i - \lambda_j|, \quad (8c)$$

$$\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) \triangleq \sum_{i \in \mathcal{V}} f_i(x_i) + \lambda_i^T \mathbf{g}_i(x_i) - K \phi(\boldsymbol{\lambda}), \quad (8d)$$

where λ is a collection of *local multipliers* employed for distributed design and \mathcal{S} is a cone for the local multipliers to reach a consensus there. Moreover, $\phi(\lambda)$ serves as a *metric of consensus* for the multipliers and $\tilde{\mathcal{L}}(\mathbf{x}, \lambda)$ is a modified Lagrangian function with a constant $K > 0$.

The following lemma reveals that the nonsmooth $K\phi(\lambda)$ plays a role as an exact penalty function for the consensus of multipliers.

Lemma 4.2: Under Assumption 3.1, for any $\mathbf{x} \in \Omega$, there holds

$$\operatorname{argmax}_{\lambda \in \mathbb{R}_+^{MN}} \tilde{\mathcal{L}}(\mathbf{x}, \lambda) = \operatorname{argmax}_{\lambda \in \mathcal{S}} \tilde{\mathcal{L}}(\mathbf{x}, \lambda),$$

provided $K > \sqrt{N}K_0$, where $K_0 \triangleq \max_{\mathbf{x} \in \Omega} \|\operatorname{col}(\mathbf{g}_1(x_1), \dots, \mathbf{g}_N(x_N))\|$.

Proof: It follows from part 3) in Lemma 2.1 that, for any $K_d > K_0$,

$$\operatorname{argmax}_{\lambda \in \mathcal{S}} \sum_{i \in \mathcal{V}} f_i(x_i) + \lambda_i^T \mathbf{g}_i(x_i) = \operatorname{argmax}_{\lambda \in \mathbb{R}_+^{MN}} \sum_{i \in \mathcal{V}} f_i(x_i) + \lambda_i^T \mathbf{g}_i(x_i) - K_d d_{\mathcal{S}}(\lambda).$$

Since $\phi(\lambda) = d_{\mathcal{S}}(\lambda) = 0, \forall \lambda \in \mathcal{S}$, it suffices to prove $\sqrt{N}\phi(\lambda) > d_{\mathcal{S}}(\lambda)$ for all $\lambda \in \mathbb{R}_+^{MN} \setminus \mathcal{S}$.

On one hand,

$$d_{\mathcal{S}}^2(\lambda) = \min_{\tilde{\lambda} \in \mathcal{S}} \|\tilde{\lambda} - \lambda\|^2 = \sum_{k=1}^N \left\| \lambda_k - \frac{\lambda_1 + \dots + \lambda_N}{N} \right\|^2 \leq \frac{1}{N} \sum_{k=1}^N \sum_{l=1}^N \|\lambda_k - \lambda_l\|^2 \leq \frac{1}{N} \sum_{k=1}^N \sum_{l=1}^N |\lambda_k - \lambda_l|^2.$$

On the other hand, since the graph is connected and undirected, there is a path $\mathcal{P}_{kl} \subset \mathcal{E}$ connecting nodes k and l for any $k, l \in \mathcal{V}$. Then

$$\phi(\lambda) = \frac{1}{2} \sum_{(i,j) \in \mathcal{E}} |\lambda_i - \lambda_j| \geq \frac{1}{2} \sum_{(i,j) \in \mathcal{P}_{kl}} |\lambda_i - \lambda_j| \geq |\lambda_k - \lambda_l|.$$

Thus, $d_{\mathcal{S}}^2(\lambda) \leq N\phi^2(\lambda)$ and the equality holds if and only if $\lambda \in \mathcal{S}$, which implies the conclusion. ■

The correctness of the Lagrangian function $\tilde{\mathcal{L}}(\mathbf{x}, \lambda)$ to problem (4) is indicated in the following result.

Theorem 4.1: Under Assumption 3.1, the following statements are equivalent:

1) $(\mathbf{x}^*, \lambda^*) \in \Omega \times \mathbb{R}_+^{MN}$ renders the following equations

$$0 \in \Pi_{\Omega}(\mathbf{x}^*, -\partial_{\mathbf{x}} \tilde{\mathcal{L}}(\mathbf{x}^*, \lambda^*)), \quad (9a)$$

$$0 \in \Pi_{\mathbb{R}_+^{MN}}(\lambda^*, -\partial_{\lambda}(-\tilde{\mathcal{L}})(\mathbf{x}^*, \lambda^*)). \quad (9b)$$

2) $(\mathbf{x}^*, \lambda^*)$ is a saddle point of $\tilde{\mathcal{L}}(\mathbf{x}, \lambda)$ in $\Omega \times \mathbb{R}_+^{MN}$.

3) $\lambda^* = \operatorname{col}(\lambda^*, \dots, \lambda^*)$ and $(\mathbf{x}^*, \lambda^*)$ is a saddle point of $\mathcal{L}(\mathbf{x}, \lambda)$ in $\Omega \times \mathbb{R}_+^M$.

Proof: 1) \Rightarrow 2): Let $(\mathbf{x}^*, \lambda^*) \in \Omega \times \mathbb{R}_+^{MN}$ satisfying (9). Then

$$0 \in -\partial_{\mathbf{x}} \tilde{\mathcal{L}}(\mathbf{x}^*, \lambda^*) - \mathcal{N}_{\Omega}(\mathbf{x}^*), \quad (10a)$$

$$0 \in -\partial_{\lambda}(-\tilde{\mathcal{L}})(\mathbf{x}^*, \lambda^*) - \mathcal{N}_{\mathbb{R}_+^{MN}}(\lambda^*). \quad (10b)$$

Since $\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$ is convex in \mathbf{x} and concave in $\boldsymbol{\lambda}$ (or equivalently, $-\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$ is convex in $\boldsymbol{\lambda}$), it follows from part 2) in Lemma 2.1 that \mathbf{x}^* is the minimum point of $\tilde{\mathcal{L}}(\cdot, \boldsymbol{\lambda}^*)$ in Ω and $\boldsymbol{\lambda}^*$ is a maximum point of $\tilde{\mathcal{L}}(\mathbf{x}^*, \cdot)$ in \mathbb{R}_+^{MN} , which implies statement 2).

2) \Rightarrow 3): Let $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ be a saddle point of $\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$ in $\Omega \times \mathbb{R}_+^{MN}$. Then

$$\boldsymbol{\lambda}^* \in \operatorname{argmax}_{\boldsymbol{\lambda} \geq \mathbf{0}} \tilde{\mathcal{L}}(\mathbf{x}^*, \boldsymbol{\lambda}). \quad (11)$$

It follows from Lemma 4.2 that $\boldsymbol{\lambda}^* = \operatorname{col}(\lambda^*, \dots, \lambda^*)$. Substituting this $\boldsymbol{\lambda}^*$ into the saddle point inequalities with respect to $\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$ yields

$$\mathbf{x}^* \in \operatorname{argmin}_{\mathbf{x} \in \Omega} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*) \text{ and } \lambda^* \in \operatorname{argmax}_{\lambda \geq 0} \mathcal{L}(\mathbf{x}^*, \lambda), \quad (12)$$

because of the identity $\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}^*) = \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*)$. Therefore, the conclusion follows.

3) \Rightarrow 1): Suppose $\boldsymbol{\lambda}^* = \operatorname{col}(\lambda^*, \dots, \lambda^*)$ and $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a saddle point of $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$. According to Lemma 4.2, condition (11) holds. Again, from part 2) in Lemma 2.1, condition (10) holds, which implies statement 1). \blacksquare

By Theorem 4.1 and Lemma 4.1, the saddle points of $\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$ match exactly the saddle points of $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$, which are in accordance with the optimal primal-dual solutions.

Based on $\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$, we present a distributed continuous-time algorithm to solve (4) as follows:

$$\forall i \in \mathcal{V} : \begin{cases} \dot{x}_i \in \Pi_{\Omega_i}(x_i, -\partial_{x_i} \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})), & x_i(0) \in \Omega_i \\ \dot{\lambda}_i \in \Pi_{\mathbb{R}_+^M}(\lambda_i, -\partial_{\lambda_i} (-\tilde{\mathcal{L}})(\mathbf{x}, \boldsymbol{\lambda})), & \lambda_i(0) \in \mathbb{R}_+^M, \end{cases} \quad (13)$$

where

$$\partial_{x_i} \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) = \partial f_i(x_i) + \partial g_i(x_i) \lambda_i, \quad -\partial_{\lambda_i} (-\tilde{\mathcal{L}})(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{g}_i(x_i) - K \sum_{j \in \mathcal{N}_i} \operatorname{Sgn}(\lambda_i - \lambda_j),$$

(the second equality follows because graph \mathcal{G} is undirected) and $\operatorname{Sgn}(\cdot)$ is the set-valued sign function with each component defined as

$$\operatorname{Sgn}(y) \triangleq \partial|y| = \begin{cases} \{1\} & \text{if } y > 0 \\ \{-1\} & \text{if } y < 0 \\ [-1, 1] & \text{if } y = 0 \end{cases}.$$

For simplicity, we rewrite algorithm (13) in a compact form

$$\begin{cases} \dot{\mathbf{x}} \in \Pi_{\Omega}(\mathbf{x}, -\partial_{\mathbf{x}} \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})), & \mathbf{x}(0) \in \Omega \\ \dot{\boldsymbol{\lambda}} \in \Pi_{\mathbb{R}_+^{MN}}(\boldsymbol{\lambda}, -\partial_{\boldsymbol{\lambda}} (-\tilde{\mathcal{L}})(\mathbf{x}, \boldsymbol{\lambda})), & \boldsymbol{\lambda}(0) \in \mathbb{R}_+^{MN}. \end{cases} \quad (14)$$

Algorithm (13) is fully distributed since each agent $i \in \mathcal{V}$ only updates its local variables x_i and λ_i according to its local functions f_i, g_i and the information of its neighbors $\lambda_j, j \in \mathcal{N}_i$.

Remark 4.1: Some discussions about our method are given below.

- In the original $\mathcal{L}(\mathbf{x}, \lambda)$, each $\mathbf{g}_i(\mathbf{x}_i)$ shares a common λ and the multiplier λ performs on the coupled $\mathbf{g}(\mathbf{x})$, while, in the modified one $\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$, all the local parts of cost and constraint functions are gathered in a decoupled way.
- Our distributed algorithm involves only primal variables and local multipliers without auxiliary dynamics, while some existing distributed algorithms such as those given in [16]–[19] employed auxiliary dynamics for the convergence.
- From the estimation $K_0 \leq \sum_{i=1}^N \max_{\mathbf{x}_i \in \Omega_i} \{\|\mathbf{g}_i(\mathbf{x}_i)\|\}$ for Lemma 4.2, parameter K can be assigned via local estimation of each $\max_{\mathbf{x}_i \in \Omega_i} \|\mathbf{g}_i(\mathbf{x}_i)\|$ and calculating the sum in a distributed manner.

B. Existence and Convergence

Dynamics (14) is nonsmooth due to the projection operator and subgradients of the nonsmooth Lagrangian function. Thus, we need to check the existence of its solution (trajectory).

Theorem 4.2: Under Assumption 3.1, for any initial value $\mathbf{x}(0) \in \Omega, \boldsymbol{\lambda}(0) \in \mathbb{R}_+^{MN}$, there exists a solution to (14).

Proof: Let $\tilde{\mathcal{D}}$ be the convex hull of $\boldsymbol{\lambda}(0)$ and $\prod_{i=1}^N \mathcal{D}$, where \mathcal{D} is in (7). Then $\tilde{\mathcal{D}}$ is compact and convex. Consider the following differential inclusion

$$\begin{cases} \dot{\mathbf{x}} \in \Pi_{\Omega}(\mathbf{x}, -\partial_{\mathbf{x}} \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})) \\ \dot{\boldsymbol{\lambda}} \in \Pi_{\tilde{\mathcal{D}}}(\boldsymbol{\lambda}, -\partial_{\boldsymbol{\lambda}}(-\tilde{\mathcal{L}})(\mathbf{x}, \boldsymbol{\lambda})) \end{cases}. \quad (15)$$

Since $\mathcal{T}_{\tilde{\mathcal{D}}}(\boldsymbol{\lambda}) \subset \mathcal{T}_{\mathbb{R}_+^{MN}}(\boldsymbol{\lambda}), \forall \boldsymbol{\lambda} \in \tilde{\mathcal{D}}$, any solution to (15) is also a solution to (14). Thus, it suffices to prove the existence of solution for (15).

Let $\mathcal{F}(\mathbf{x}, \boldsymbol{\lambda}) \triangleq \text{col}(-\partial_{\mathbf{x}} \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}), -\partial_{\boldsymbol{\lambda}}(-\tilde{\mathcal{L}})(\mathbf{x}, \boldsymbol{\lambda}))$, and $C \triangleq \Omega \times \tilde{\mathcal{D}}$. We claim that \mathcal{F} is upper semicontinuous over C . The locally Lipschitz continuity of $f(\mathbf{x}), \mathbf{g}(\mathbf{x}), \phi(\boldsymbol{\lambda})$ implies that \mathcal{F} has compact values over the compact set C . Then it suffices to prove $\text{gph } \mathcal{F}$ is closed due to part 1) in Lemma 2.2. Let $\{\mathbf{x}_k, \boldsymbol{\lambda}_k\}$ and $\{\zeta_k, \eta_k\}$ be sequences in C and $\mathbb{R}^{n_1+\dots+n_N} \times \mathbb{R}^{MN}$ such that (a) $\text{col}(\zeta_k, \eta_k) \in \mathcal{F}(\mathbf{x}_k, \boldsymbol{\lambda}_k)$, (b) $(\mathbf{x}_k, \boldsymbol{\lambda}_k)$ converges to $(\mathbf{x}, \boldsymbol{\lambda})$, and (c) (ζ, η) is a cluster point of the sequence (ζ_k, η_k) . We can extract a subsequence of (ζ_k, η_k) (without relabeling) such that $\lim_{k \rightarrow +\infty} (\zeta_k, \eta_k) = (\zeta, \eta)$. Since $-\partial_{\mathbf{x}} \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) = -\partial f(\mathbf{x}) - \text{diag}\{\partial \mathbf{g}_1(\mathbf{x}), \dots, \partial \mathbf{g}_N(\mathbf{x})\} \boldsymbol{\lambda}$, $\zeta_k = -\alpha_k - \beta_k \boldsymbol{\lambda}_k$, where $\alpha_k \in \partial f(\mathbf{x}_k)$ and $\beta_k \in \text{diag}\{\partial \mathbf{g}_1(\mathbf{x}_k), \dots, \partial \mathbf{g}_N(\mathbf{x}_k)\}$. It follows from part 1) in Lemma 2.1 that $\lim_{k \rightarrow +\infty} \alpha_k = \alpha \in \partial f(\mathbf{x})$ and $\lim_{k \rightarrow +\infty} \beta_k = \beta \in \text{diag}\{\partial \mathbf{g}_1(\mathbf{x}), \dots, \partial \mathbf{g}_N(\mathbf{x})\}$ after extracting subsequences of $\{\alpha_k\}$ and $\{\beta_k\}$ without relabeling. Therefore, $\zeta \in -\partial_{\mathbf{x}} \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$. Similarly, $\eta \in -\partial_{\boldsymbol{\lambda}}(-\tilde{\mathcal{L}})(\mathbf{x}, \boldsymbol{\lambda})$. Thus, $(\zeta, \eta) \in \mathcal{F}(\mathbf{x}, \boldsymbol{\lambda})$, i.e., $\text{gph } \mathcal{F}$ is closed.

Finally, according to part 2) and part 3) in Lemma 2.2, there exists a solution to system (15), which is also a solution to (14). \blacksquare

Then it is time to show the convergence of our algorithm.

Theorem 4.3: Under Assumption 3.1, algorithm (14) is stable and any of its solutions converges to the set of saddle points of $\tilde{\mathcal{L}}$. Moreover, for any solution $(\mathbf{x}(t), \boldsymbol{\lambda}(t))$, there exists a saddle point $(\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*)$ of $\tilde{\mathcal{L}}$ such that

$$\lim_{t \rightarrow +\infty} (\mathbf{x}(t), \boldsymbol{\lambda}(t)) = (\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*). \quad (16)$$

Proof: For all $\mathbf{x}, \underline{\mathbf{x}} \in \Omega, \boldsymbol{\lambda}, \underline{\boldsymbol{\lambda}} \in \mathbb{R}_+^{MN}$, the following basic conditions hold, according to the definitions of projection, normal cone and the convexity-concavity of $\tilde{\mathcal{L}}$.

- (projection)

$$\begin{aligned} \Pi_{\Omega}(\mathbf{x}, -\partial_{\mathbf{x}}\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})) &\subset -\partial_{\mathbf{x}}\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) - \mathcal{N}_{\Omega}(\mathbf{x}), \\ \Pi_{\mathbb{R}_+^{MN}}(\boldsymbol{\lambda}, -\partial_{\boldsymbol{\lambda}}(-\tilde{\mathcal{L}})(\mathbf{x}, \boldsymbol{\lambda})) &\subset -\partial_{\boldsymbol{\lambda}}(-\tilde{\mathcal{L}})(\mathbf{x}, \boldsymbol{\lambda}) - \mathcal{N}_{\mathbb{R}_+^{MN}}(\boldsymbol{\lambda}), \end{aligned}$$

- (normal cone)

$$(\underline{\mathbf{x}} - \mathbf{x})^T u_x \leq 0, \forall u_x \in \mathcal{N}_{\Omega}(\mathbf{x}), \quad (\underline{\boldsymbol{\lambda}} - \boldsymbol{\lambda})^T u_{\lambda} \leq 0, \forall u_{\lambda} \in \mathcal{N}_{\mathbb{R}_+^{MN}}(\boldsymbol{\lambda}),$$

- (convexity-concavity)

$$\begin{aligned} (\underline{\mathbf{x}} - \mathbf{x})^T v_x &\leq \tilde{\mathcal{L}}(\underline{\mathbf{x}}, \boldsymbol{\lambda}) - \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}), \quad \forall v_x \in \partial_{\mathbf{x}}\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}), \\ (\underline{\boldsymbol{\lambda}} - \boldsymbol{\lambda})^T v_{\lambda} &\leq \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) - \tilde{\mathcal{L}}(\mathbf{x}, \underline{\boldsymbol{\lambda}}), \quad \forall v_{\lambda} \in \partial_{\boldsymbol{\lambda}}(-\tilde{\mathcal{L}})(\mathbf{x}, \boldsymbol{\lambda}). \end{aligned}$$

Therefore,

$$(\mathbf{x} - \underline{\mathbf{x}})^T w_x \leq \tilde{\mathcal{L}}(\underline{\mathbf{x}}, \boldsymbol{\lambda}) - \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}), \quad \forall w_x \in \Pi_{\Omega}(\mathbf{x}, -\partial_{\mathbf{x}}\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})), \quad (17a)$$

$$(\boldsymbol{\lambda} - \underline{\boldsymbol{\lambda}})^T w_{\lambda} \leq \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}) - \tilde{\mathcal{L}}(\mathbf{x}, \underline{\boldsymbol{\lambda}}), \quad \forall w_{\lambda} \in \Pi_{\mathbb{R}_+^{MN}}(\boldsymbol{\lambda}, -\partial_{\boldsymbol{\lambda}}(-\tilde{\mathcal{L}})(\mathbf{x}, \boldsymbol{\lambda})). \quad (17b)$$

Let $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ be an equilibrium point of (14), which satisfies (9). From Lemma 4.1 and Theorem 4.1, \mathbf{x}^* coincides with the unique solution of the primal problem and $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a saddle point of $\tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda})$. Define

$$W(\mathbf{x}, \boldsymbol{\lambda}) \triangleq \tilde{\mathcal{L}}(\mathbf{x}, \boldsymbol{\lambda}^*) - \tilde{\mathcal{L}}(\mathbf{x}^*, \boldsymbol{\lambda}), \quad \forall \mathbf{x} \in \Omega, \boldsymbol{\lambda} \geq \mathbf{0}.$$

Obviously, $W(\cdot)$ is a locally Lipschitz continuous function. Moreover, since $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$ is a saddle point of $\tilde{\mathcal{L}}$, $W(\mathbf{x}, \boldsymbol{\lambda}) \geq 0$ and $W(\mathbf{x}, \boldsymbol{\lambda}) = 0$ if and only if $(\mathbf{x}, \boldsymbol{\lambda}) = (\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*)$ for some saddle point $(\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*)$ of $\tilde{\mathcal{L}}$.

Consider a Lyapunov function

$$V(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2} \|\mathbf{x} - \mathbf{x}^*\|^2 + \frac{1}{2} \|\boldsymbol{\lambda} - \boldsymbol{\lambda}^*\|^2.$$

Let $(\mathbf{x}(t), \boldsymbol{\lambda}(t))$ be any solution to (14). Since $\dot{\mathbf{x}} \in \mathcal{T}_\Omega(\mathbf{x})$ and $\dot{\boldsymbol{\lambda}} \in \mathcal{T}_{\mathbb{R}_+^{MN}}(\boldsymbol{\lambda})$, we have $\mathbf{x}(t) \in \Omega$, $\boldsymbol{\lambda}(t) \geq \mathbf{0}$, $\forall t \geq 0$. Moreover, it follows from (17) that

$$\frac{d}{dt}V(\mathbf{x}(t), \boldsymbol{\lambda}(t)) = (\mathbf{x}(t) - \mathbf{x}^*)^T \dot{\mathbf{x}}(t) + (\boldsymbol{\lambda}(t) - \boldsymbol{\lambda}^*)^T \dot{\boldsymbol{\lambda}}(t) \leq -W(\mathbf{x}(t), \boldsymbol{\lambda}(t)) \leq 0, \quad (18)$$

for almost all $t \geq 0$. Therefore, (14) is stable.

Furthermore, since $W(\cdot)$ is locally Lipschitz continuous and $\mathbf{x}(t), \boldsymbol{\lambda}(t)$ are absolutely continuous, $W(t)$ (shorthand for $W(\mathbf{x}(t), \boldsymbol{\lambda}(t))$) is uniformly continuous in t . We claim that $W(t)$ is Riemann integrable over the infinite interval $[0, +\infty)$. In fact, the Riemann integral of the continuous function W over any finite interval $[0, t]$ equals to the corresponding Lebesgue integral. Moreover, $\int_0^t W(\tau) d\tau$ is monotonically increasing since W is nonnegative, and it follows from (18) that the Lebesgue integral of W over the infinite interval $[0, +\infty)$, if exists, must be bounded.

As a result, $\int_0^{+\infty} W(\tau) d\tau$ exists and is finite. Then, by the Barbalat's lemma, $(\mathbf{x}(t), \boldsymbol{\lambda}(t))$ converges to the zeros set of W , which is exactly the set of saddle points of $\tilde{\mathcal{L}}$.

Let $(\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*)$ be a cluster point of $(\mathbf{x}(t), \boldsymbol{\lambda}(t))$ as $t \rightarrow +\infty$. Then $(\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*)$ is a saddle point of $\tilde{\mathcal{L}}$. Define

$$\tilde{V}(\mathbf{x}, \boldsymbol{\lambda}) = \frac{1}{2} \|\mathbf{x} - \mathbf{x}^*\|^2 + \frac{1}{2} \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}^*\|^2.$$

It follows from similar arguments that $\dot{\tilde{V}}(t) \leq -\tilde{\mathcal{L}}(\mathbf{x}(t), \tilde{\boldsymbol{\lambda}}^*) + \tilde{\mathcal{L}}(\mathbf{x}^*, \boldsymbol{\lambda}(t)) \leq 0$ for almost all $t > 0$. Hence, $(\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*)$ is Lyapunov stable. It follows from Lemma 2.3 that (16) holds. ■

Finally, we discuss the convergence rate. Define

$$\hat{\mathbf{x}}(t) \triangleq \frac{1}{t} \int_0^t \mathbf{x}(\tau) d\tau, \quad \hat{\boldsymbol{\lambda}}(t) \triangleq \frac{1}{t} \int_0^t \boldsymbol{\lambda}(\tau) d\tau, \quad (19)$$

where trajectories $\mathbf{x}(\cdot), \boldsymbol{\lambda}(\cdot)$ are in Theorem 4.3.

Theorem 4.4: Under Assumption 3.1, there exists a constant $\theta_0 > 0$ such that

$$\|\tilde{\mathcal{L}}(\hat{\mathbf{x}}(t), \hat{\boldsymbol{\lambda}}(t)) - \tilde{\mathcal{L}}(\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*)\| \leq \frac{\theta_0}{t}, \quad \forall t > 0. \quad (20)$$

Proof: Since Ω is convex and $\mathbf{x}(\cdot) \in \Omega, \boldsymbol{\lambda}(\cdot) \in \mathbb{R}_+^{MN}$, $\hat{\mathbf{x}}(t) \in \Omega, \hat{\boldsymbol{\lambda}}(t) \in \mathbb{R}_+^{MN}$, $\forall t > 0$. It follows from the Jensen's inequality for the convex-concave $\tilde{\mathcal{L}}$ that, for any $\underline{\mathbf{x}} \in \Omega, \underline{\boldsymbol{\lambda}} \in \mathbb{R}_+^{MN}$,

$$\frac{1}{t} \int_0^t \tilde{\mathcal{L}}(\mathbf{x}(\tau), \underline{\boldsymbol{\lambda}}) d\tau \geq \tilde{\mathcal{L}}(\hat{\mathbf{x}}(t), \underline{\boldsymbol{\lambda}}), \quad \frac{1}{t} \int_0^t \tilde{\mathcal{L}}(\underline{\mathbf{x}}, \boldsymbol{\lambda}(\tau)) d\tau \leq \tilde{\mathcal{L}}(\underline{\mathbf{x}}, \hat{\boldsymbol{\lambda}}(t)). \quad (21)$$

Moreover, it follows from (17) that, for almost all $\tau > 0$,

$$\begin{aligned} \frac{d}{d\tau} \left(\frac{1}{2} \|\mathbf{x}(\tau) - \underline{\mathbf{x}}\|^2 \right) &\leq \tilde{\mathcal{L}}(\underline{\mathbf{x}}, \boldsymbol{\lambda}(\tau)) - \tilde{\mathcal{L}}(\mathbf{x}(\tau), \boldsymbol{\lambda}(\tau)), \\ \frac{d}{d\tau} \left(\frac{1}{2} \|\boldsymbol{\lambda}(\tau) - \underline{\boldsymbol{\lambda}}\|^2 \right) &\leq \tilde{\mathcal{L}}(\mathbf{x}(\tau), \boldsymbol{\lambda}(\tau)) - \tilde{\mathcal{L}}(\mathbf{x}(\tau), \underline{\boldsymbol{\lambda}}). \end{aligned} \quad (22)$$

For any fixed $t > 0$, the time average of (22) over integral interval $[0, t]$ with relaxation (21) indicates

$$\begin{aligned} -\frac{\|\mathbf{x}(0) - \underline{\mathbf{x}}\|^2}{2t} &\leq \tilde{\mathcal{L}}(\underline{\mathbf{x}}, \hat{\boldsymbol{\lambda}}(t)) - \frac{1}{t} \int_0^t \tilde{\mathcal{L}}(\mathbf{x}(\tau), \boldsymbol{\lambda}(\tau)) d\tau, \\ -\frac{\|\boldsymbol{\lambda}(0) - \underline{\boldsymbol{\lambda}}\|^2}{2t} &\leq \frac{1}{t} \int_0^t \tilde{\mathcal{L}}(\mathbf{x}(\tau), \boldsymbol{\lambda}(\tau)) d\tau - \tilde{\mathcal{L}}(\hat{\mathbf{x}}(t), \underline{\boldsymbol{\lambda}}), \end{aligned} \quad (23)$$

Replacing $(\underline{\mathbf{x}}, \underline{\boldsymbol{\lambda}})$ by $(\hat{\mathbf{x}}(t), \hat{\boldsymbol{\lambda}}(t))$ in (23) yields

$$\left\| \frac{1}{t} \int_0^t \tilde{\mathcal{L}}(\mathbf{x}(\tau), \boldsymbol{\lambda}(\tau)) d\tau - \tilde{\mathcal{L}}(\hat{\mathbf{x}}(t), \hat{\boldsymbol{\lambda}}(t)) \right\| \leq \frac{\theta_1}{2t},$$

where $\theta_1 \triangleq \max_{t>0} \{\|\mathbf{x}(0) - \hat{\mathbf{x}}(t)\|^2, \|\boldsymbol{\lambda}(0) - \hat{\boldsymbol{\lambda}}(t)\|^2\}$. Note that $(\hat{\mathbf{x}}(t), \hat{\boldsymbol{\lambda}}(t))$ are uniformly bounded for $t \in [0, +\infty)$ due to (16). Similarly, since $\tilde{\mathcal{L}}(\mathbf{x}^*, \hat{\boldsymbol{\lambda}}(t)) \leq \tilde{\mathcal{L}}(\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*) \leq \tilde{\mathcal{L}}(\hat{\mathbf{x}}(t), \tilde{\boldsymbol{\lambda}}^*)$, we have from replacing $(\underline{\mathbf{x}}, \underline{\boldsymbol{\lambda}})$ by $(\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*)$ in (23) that

$$\left\| \frac{1}{t} \int_0^t \tilde{\mathcal{L}}(\mathbf{x}(\tau), \boldsymbol{\lambda}(\tau)) d\tau - \tilde{\mathcal{L}}(\mathbf{x}^*, \tilde{\boldsymbol{\lambda}}^*) \right\| \leq \frac{\theta_2}{2t},$$

where $\theta_2 \triangleq \max\{\|\mathbf{x}(0) - \mathbf{x}^*\|^2, \|\boldsymbol{\lambda}(0) - \tilde{\boldsymbol{\lambda}}^*\|^2\}$. Thus, (20) holds with $\theta_0 \triangleq (\theta_1 + \theta_2)/2$. \blacksquare

Theorems 4.1–4.4 provide a complete procedure to prove that algorithm (14) solves problem (4). In particular, Theorem 4.4 indicates that the value of the Lagrangian function with respect to time average trajectories converges to the value at saddle points with the convergence rate $O(\frac{1}{t})$.

V. NUMERICAL EXAMPLES

In this section, we first take a simple example for illustration and then consider a more practical example for the performance of our algorithm.

Example 5.1: Consider 4 agents for the optimization problem (4) with nonsmooth cost and constraint functions:

$$f_i(x_i) = (x_{i,1} + a_{i,1}x_{i,2})^2 + x_{i,1} + a_{i,2}x_{i,2} + \sqrt{x_{i,1}^2 + x_{i,2}^2}, \quad i = 1, 2, 3, 4$$

and

$$g_{i,1}(x_i) = \sqrt{x_{i,1}^2 + x_{i,2}^2} - d_{i,1}, \quad g_{i,2}(x_i) = -x_{i,1} - x_{i,2} + d_{i,2},$$

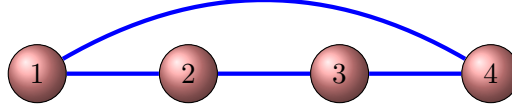


Fig. 1. The communication graph of the four agents.

TABLE I
PARAMETERS SETTING

	d_i	a_i	$x_i(0)$
$i = 1$	(6, 2)	(8, 2)	(2, 6)
$i = 2$	(6, 3)	(4, 7)	(1, 1)
$i = 3$	(6, 4)	(0.13, 8)	(5, 4)
$i = 4$	(6, 5)	(4, 20)	(10, 5)

where $x_i = (x_{i,1}, x_{i,2}) \in \mathbb{R}^2$, $d_i = (d_{i,1}, d_{i,2}) \in \mathbb{R}^2$ and $a_i = (a_{i,1}, a_{i,2}) \in \mathbb{R}^2$ for $i = 1, 2, 3, 4$. The local constraint sets of the four agents are

$$\Omega_1 = \{x_1 \in \mathbb{R}^2 \mid (x_{1,1} - 2)^2 + (x_{1,2} - 3)^2 \leq 25\},$$

$$\Omega_2 = \{x_2 \in \mathbb{R}^2 \mid x_{2,1} \geq 0, x_{2,1} \geq 0, x_{2,1} + 2x_{2,2} \leq 4\},$$

$$\Omega_3 = \{x_3 \in \mathbb{R}^2 \mid 4 \leq x_{3,1} \leq 6, 2 \leq x_{3,2} \leq 5\},$$

$$\Omega_4 = \{x_4 \in \mathbb{R}^2 \mid 0 \leq x_{4,1} \leq 15, 0 \leq x_{4,1} \leq 20\}.$$

The communication graph is shown in Fig. 1 and algorithm parameters are listed in Table I. Both centralized primal-dual algorithm and our distributed algorithm are utilized to solve this problem and the results are shown in Figs. 2–4. The trajectories of primal variables are both within their local constraint sets as shown in Figs. 2 and 3, while the Lyapunov functions of the algorithms decrease monotonically as shown in Fig. 4.

Example 5.2: Consider problem (4) with each local constraint as $x_i \in [0, 1]$, where each cost function is

$$f_i(x_i) = a_i x_i^2 + \ln(1 + b_i x_i) + c_i |x_i - d_i| + e_i x_i,$$

and the coupled inequality constraints are $\mathbf{g}(\mathbf{x}) = \mathbf{P}\mathbf{x} - \mathbf{q} \leq \mathbf{0}$. We randomly generate coefficients $a_i, b_i, c_i, d_i, e_i \in [0, 1]$, matrices $\mathbf{0}_{M \times N} \leq \mathbf{P} \leq \mathbf{1}_M \mathbf{1}_N^T$ and $\mathbf{q} \geq \mathbf{0}_M$ such that strictly feasible point exists. We choose the network size as $N = 10, 20, 50$ and the number of coupled constraints as $M = 5$. For each problem setting, we randomly generate 100 communication graphs. Over each graph, we conduct

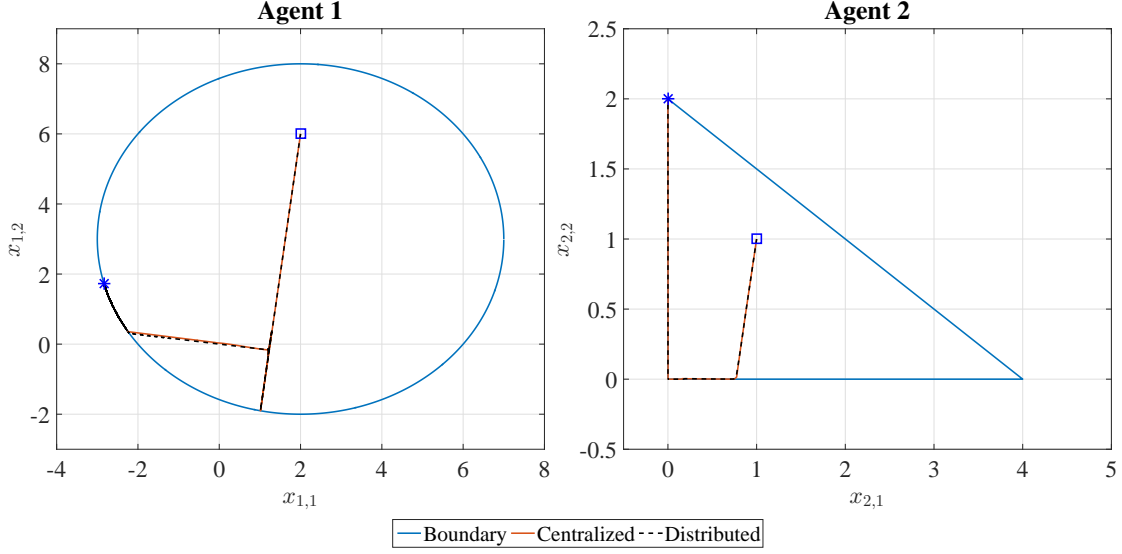


Fig. 2. The trajectories of agent 1 and agent 2

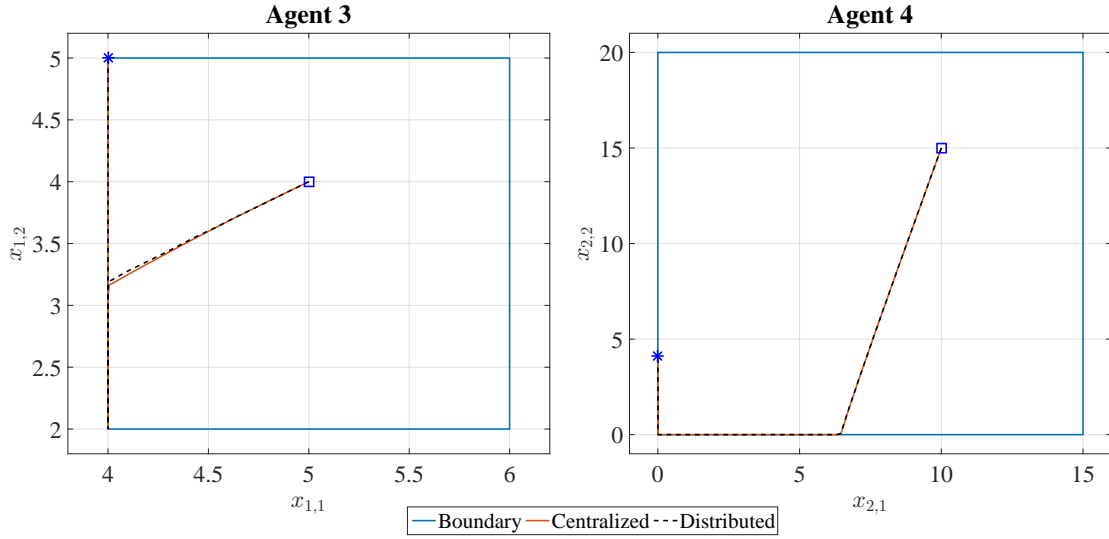


Fig. 3. The trajectories of agent 3 and agent 4

the numerical experiment and take the relative error $e(t) = \frac{\max_{i=1,\dots,N} |x_i(t) - x_i^*|}{\max_{i=1,\dots,N} |x_i^*|}$ for $t = 20, 60, 100$. The average results are shown in Table II, which indicates the effectiveness of our distributed algorithm.

VI. CONCLUSION

In this note, a distributed nonsmooth convex optimization problem with coupled inequality constraints has been studied. Based on a modified Lagrangian function constructed via local multipliers

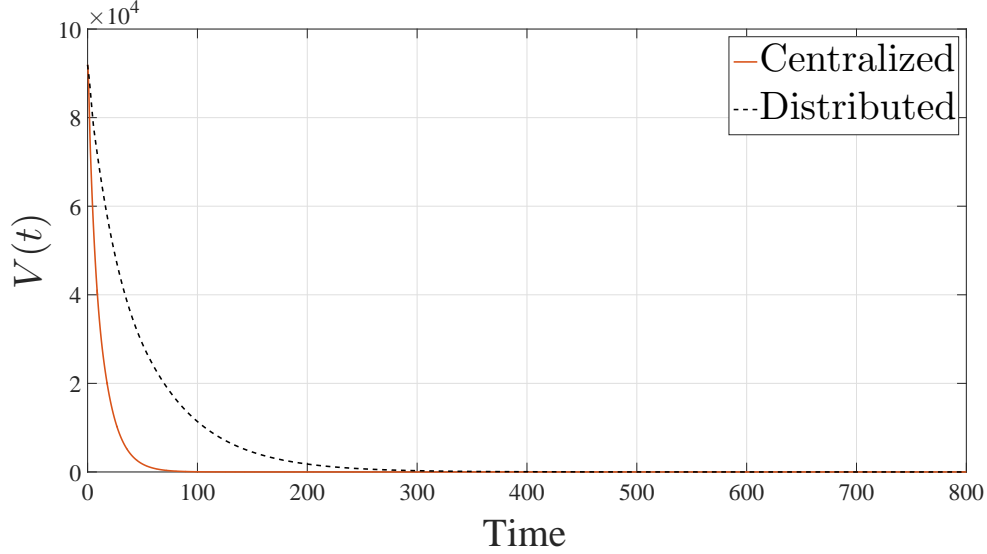


Fig. 4. The Lyapunov functions

TABLE II
RELATIVE ERROR VS. NETWORK SIZE

	$t = 20$	$t = 60$	$t = 100$
$N = 10$	0.1982	0.0711	0.0143
$N = 20$	0.5530	0.0290	0.0042
$N = 50$	0.1391	0.0170	0.0105

and nonsmooth penalty technique, a distributed continuous-time algorithm has been proposed. Also, the convergence of the nonsmooth dynamics has been proved and the convergence rate has been analyzed. Additionally, the effectiveness of the algorithm has been illustrated by two numerical examples.

REFERENCES

- [1] A. Nedić, A. Olshevsky, and W. Shi, “Achieving geometric convergence for distributed optimization over time-varying graphs,” *arXiv preprint arXiv:1607.03218*, 2016.
- [2] M. Zhu and S. Martínez, “On distributed convex optimization under inequality and equality constraints,” *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.
- [3] G. Shi, K. H. Johansson, and Y. Hong, “Reaching an optimal consensus: dynamical systems that compute intersections of convex sets,” *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 610–622, 2013.
- [4] B. Gharesifard and J. Cortés, “Distributed continuous-time convex optimization on weight-balanced digraphs,” *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2014.
- [5] Q. Liu and J. Wang, “A second-order multi-agent network for bound-constrained distributed optimization,” *IEEE Transactions on Automatic Control*, vol. 60, no. 12, pp. 3310–3315, 2015.

- [6] X. Zeng, P. Yi, and Y. Hong, “Distributed continuous-time algorithm for constrained convex optimizations via nonsmooth analysis approach,” *IEEE Transactions on Automatic Control*, 2017.
- [7] M. Forti, P. Nistri, and M. Quincampoix, “Generalized neural network for nonsmooth nonlinear programming problems,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 9, pp. 1741–1754, 2004.
- [8] Y. Zhang, Z. Deng, and Y. Hong, “Distributed optimal coordination for multiple heterogeneous Euler–Lagrangian systems,” *Automatica*, vol. 79, pp. 207–213, 2017.
- [9] A. Nedić and A. Ozdaglar, “Approximate primal solutions and rate analysis for dual subgradient methods,” *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1757–1780, 2009.
- [10] I. Necoara and V. Nedelcu, “On linear convergence of a distributed dual gradient algorithm for linearly constrained separable convex problems,” *Automatica*, vol. 55, pp. 209–216, 2015.
- [11] A. Nedić and A. Ozdaglar, “Subgradient methods for saddle-point problems,” *Journal of Optimization Theory and Applications*, vol. 142, no. 1, pp. 205–228, 2009.
- [12] D. Feijer and F. Paganini, “Stability of primal-dual gradient dynamics and applications to network optimization,” *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.
- [13] A. Cherukuri, E. Mallada, and J. Cortés, “Asymptotic convergence of constrained primal-dual dynamics,” *Systems & Control Letters*, vol. 87, pp. 10–15, 2016.
- [14] S. K. Niederländer and J. Cortés, “Distributed coordination for nonsmooth convex optimization via saddle-point dynamics,” *arXiv preprint arXiv:1606.09298*, 2016.
- [15] E. Wei, A. Ozdaglar, and A. Jadbabaie, “A distributed Newton method for network utility maximization—I: algorithm,” *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2162–2175, 2013.
- [16] A. Cherukuri and J. Cortés, “Initialization-free distributed coordination for economic dispatch under varying loads and generator commitment,” *Automatica*, vol. 74, no. 12, pp. 183–193, 2016.
- [17] P. Yi, Y. Hong, and F. Liu, “Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and its application to economic dispatch of power systems,” *Automatica*, vol. 74, no. 12, pp. 259–269, 2016.
- [18] X. Zeng, P. Yi, Y. Hong, and L. Xie, “Continuous-time distributed algorithms for extended monotropic optimization problems,” *arXiv:1608.01167*, 2016.
- [19] T. Chang, A. Nedić, and A. Scaglione, “Distributed constrained optimization by consensus-based primal-dual perturbation method,” *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1524–1538, 2014.
- [20] C. Godsil and G. F. Royle, *Algebraic Graph Theory*, ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 2001, vol. 207.
- [21] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 1998, vol. 178.
- [22] J. P. Aubin and A. Cellina, *Differential Inclusions*, ser. Grundlehren der mathematischen Wissenschaften. Berlin Heidelberg: Springer-Verlag, 1984, vol. 264.
- [23] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, ser. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, 1998, vol. 317.