2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Distributed nonsmooth optimization with different local constraints via exact penalty

Shuyu Liu, Shu Liang, Yiguang Hong, *Fellow, IEEE*

***Abstract*— In this paper, we study a distributed optimization problem, where decision variables of agents need to satisfy different local constraints and the consensus constraint to minimize the sum of local cost functions. We propose a novel method to remove all these constraints by employing the exact penalty so that the derived equivalent unconstrained problem can be directly solved by subgradient descent type differential inclusions. The algorithm achieves $\mathcal{O}(\frac{1}{t})$ rate of convergence when the cost functions are convex. Moreover, it achieves exponential convergence when the cost functions are strongly convex. Our method needs no dual variable to deal with the constraints so that computation and communication resources are saved in comparison with primal-dual methods. In addition, the method overcomes a divergence problem arising from an existing exponentially convergent distributed algorithm based on the exact penalty when the local constraints are different.**

I. Introduction

In recent years, distributed optimization has been widely studied and applied in many fields such as machine learning [1], localization problems [2], resource allocation [3] and sensor networks [4]. Most works have focused on distributed consensus optimization problems, where multiple agents only have access to their own information to minimize the cost function subject to the consensus constraint by exchanging information through a network. To solve such problems, many algorithms have been proposed, including subgradient dynamics [5]–[7] and projected algorithms [8]–[11].

Constrained distributed optimization problems are widely used in modeling practical problems, where agents have partial information. For example, in a power system, plants usually have different local generation capacities, acting as different local constraints. To solve distributed optimization problems with different local constraints, some distributed algorithms have been proposed in [7], [8], [12]. However, [7], [8] provide no explicit convergence rate, and [12] is only for equality constraints. Moreover, a counterexample given in [13] showed that even if local decision variables were initialized at some optimum, the dual averaging algorithm could not converge to the optimal solution. Also, another counterexample given in [14] showed that many algorithms fail to converge exponentially for strongly convex objectives and different local nonsmooth terms.

Exponential convergence is important for algorithm evaluation, since such algorithms can reach the equilibrium fast. Some exponential convergence results have been obtained in the distributed optimization works [15], [16] in the absence of local constraints. Moreover, distributed algorithms designed in [10], [17] have achieved exponential convergence for local equality constraints. Local constraints can be satisfied by local projection, whereas exponential convergence cannot be guaranteed with the projection terms. A recent work [11] has achieved exponential convergence by using projected dynamics in the presence of local constraints for strongly convex optimization. A drawback of [11] is that agents have to know the global constraint set for the local projections and cannot be adapted to account for different local constraint sets. A counterexample can be given to show that the algorithm given in [11] diverges if projections with respect to different constraints are used.

In this paper, we consider a distributed consensus optimization problem and focus on different local constraints. We design a distributed algorithm to solve this problem and achieve exponential convergence when the cost functions are strongly convex. The main contributions are listed as follows.

1) We overcome the difficulties caused by different local constraints and provide a distributed algorithm that achieves exponential convergence. The idea is to remove all local constraints and the consensus constraint by using the exact penalty method and then solve the equivalently transformed unconstrained optimization problem so as to remove the projection term. In [7], [10], [12], [17] on similar problems, exponential convergence is achieved only under equality constraints.
2) We design a distributed method for calculating parameters in multiple exact penalty terms. In [7], [18], such parameters are not obtained so they introduce additional adaptive dynamics for the estimation. Also, the method proposed in [5], [11] cannot be directly extended for inequality constraints. As a result, in comparisons with [5], [7], [11], [18], our method not only calculates the penalty parameters but also avoids the introduction of auxiliary variables.

The rest is organized as follows. Section II introduces the necessary definitions and concepts, and Section III formulates our problem. Section IV presents the main results, including the application of the exact penalty method, al-

This research was supported in part by the National Key Research and Development Program of China under grant 2022YFA1004700, and in part by the Natural Science Foundation of China under grant 72171171, and in part by Shanghai Municipal Science and Technology Major Project under grant 2021SHZDZX0100.

S. Liu, S. Liang, and Y. Hong are with the Department of Control Science and Engineering, Tongji University, Shanghai, 201804, China, and are also with Shanghai Research Institute for Intelligent Autonomous Systems, Shanghai, 201210, China shuyuliu@tongji.edu.cn (S. Liu), sliang@tongji.edu.cn (S. Liang), yghong@iss.ac.cn (Y. Hong)

gorithm design, and algorithm analysis. Then Section V gives two numerical simulation examples. Finally, Section VI concludes this work.

II. Preliminaries

In this section, we introduce some useful notations, definitions and results.

Let $\mathbb{R}$, $\mathbb{R}^n$, and $\mathbb{R}^{n\times m}$ be sets of real numbers, n dimensional real column vectors, and $n\times m$ matrices, respectively. Let $\mathbf{0}$ be a vector with all zero entries. 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). Let $(\cdot)^{\mathrm{T}}$ be the transpose of a vector. Let $|\cdot|$ and $\|\cdot\|$ be the l_1-norm and l_2-norm, respectively. Let $col(x_1,x_2,\cdots,x_n)$ be the column vector stacked with $x_1,x_2,\cdots,x_n$.

A function $f:\mathbb{R}^n\to\mathbb{R}$ is convex if for all $x_1,x_2\in\mathbb{R}^n$ and $0\leq\theta\leq 1$, $f(\theta x_1+(1-\theta)x_2)\leq\theta f(x_1)+(1-\theta)f(x_2)$. Moreover, it is α-strongly convex for some constant $\alpha>0$ if for all $x_1,x_2\in\mathbb{R}^n$ and $0\leq\theta\leq 1$, $f(\theta x_1+(1-\theta)x_2)\leq\theta f(x_1)+(1-\theta)f(x_2)-\frac{\theta(1-\theta)\alpha}{2}\|x_1-x_2\|^2$. The subdifferential of a convex function f at y_0 is denoted by $\partial f(y_0)$. For any $z\in\partial f(y_0)$, there holds $f(y)\geq f(y_0)+\langle z,y-y_0\rangle$, $\forall\, y\in\mathbb{R}^n$. Also, if f is α-strongly convex, then

$$\langle x-y,g_f(x)-g_f(y)\rangle\geq\alpha\|x-y\|^2,\ \forall\, x,y\in\mathbb{R}^n. \tag{1}$$

where $g_f(x)\in\partial f(x)$ and $g_f(y)\in\partial f(y)$.

A set-valued map $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^n$ associates with any $x\in\mathbb{R}^n$ a subset $F(x)\subset\mathbb{R}^n$. F is said to be upper semicontinuous if for any $x\in\mathbb{R}^n$ and any open set Q satisfying $F(x)\subset Q$, there exists a neighborhood O of x such that $F(O)\subset Q$. In particular, for any continuous convex function $f:\mathbb{R}^n\to\mathbb{R}$, the subdifferential map ∂f is upper semicontinuous.

The following lemmas will be used in our analysis.

Lemma 2.1 ([19]): Let $\Omega\subset\mathbb{R}^n$ be an open subset containing a point $x_0\in\mathbb{R}^n$. Also, let $F:\Omega\rightrightarrows\mathbb{R}^n$ be an upper semicontinuous map with nonempty convex compact values. Then there exists an absolutely continuous function $x(\cdot)$ defined on $[0,T]$ for some $T>0$ such that it is a solution to the differential inclusion $\dot{x}(t)\in F(x(t))$ with $x(0)=x_0$.

Lemma 2.2 (Barbalat's lemma, [20]): Let $\sigma:\mathbb{R}_{\geq 0}\to\mathbb{R}$ be a uniformly continuous function. If $\lim_{t\to\infty}\int_0^t\sigma(s)ds$ exists and is finite, then $\lim_{t\to\infty}\sigma(t)=0$.

III. Problem Formulation

Consider a multi-agent network with n agents interacting over an undirected network graph $\mathcal{G}\triangleq\{\mathcal{V},\mathcal{E}\}$, where $\mathcal{V}=\{1,\cdots,n\}$ is the node set and $\mathcal{E}\subset\mathcal{V}\times\mathcal{V}$ is the edge set. Each agent has a local cost function $f_i(x):\mathbb{R}^m\to\mathbb{R}$ and a local constraint $g_i(x)\leq\mathbf{0}$ with $g_i(x)=col(g_{i1}(x),g_{i2}(x),\cdots,g_{il_i}(x)):\mathbb{R}^m\to\mathbb{R}^{l_i}$. Denote $l=\sum_{i=1}^n l_i$. The overall cost function of the system is $\sum_{i\in\mathcal{V}}f_i(x)$. This optimization problem is formulated as

$$\begin{aligned}\min_{x\in\mathbb{R}^m}\ &\sum_{i\in\mathcal{V}}f_i(x),\\ \text{s.t.}\ \ &g_i(x)\leq\mathbf{0}, i\in\mathcal{V}.\end{aligned} \tag{2}$$

For $i\in\mathcal{V}$, the ith agent can only access $f_i(x),g_i(x)$ and communicate with its neighbors, whose index set is denoted by $\mathcal{N}_i=\{j\in\mathcal{V}\,|\,(i,j)\in\mathcal{E}\}$. Also, let $S_i,S\subset\mathbb{R}^m$ be the sets of local and overall constraints as $S_i\triangleq\{x\in\mathbb{R}^m\,|\,g_i(x)\leq\mathbf{0}\}$, and $S\triangleq\cap_{i\in\mathcal{V}}S_i$. Denote $\tilde{S}=\cup_{i\in\mathcal{V}}S_i$. The following assumptions are given to ensure the well-posedness of problem (2).

Assumption 1: For $i\in\mathcal{V}$, $f_i(x)$ and the components of $g_i(x)$ are convex, not necessarily smooth.

Assumption 2: For $i\in\mathcal{V}$, $f_i(x)$ is c_i-Lipschitz continuous on $\tilde{S}$ for some constant $c_i>0$, i.e., $|f_i(x)-f_i(y)|\leq c_i\|x-y\|$, $\forall\, x,y\in\tilde{S}$. Also, $g_i(x)$ is s_i-Lipschitz continuous on $\tilde{S}$ for some constant $s_i>0$.

Assumption 3: The undirected graph $\mathcal{G}$ is connected.

Assumption 4: The constraint set S is internally nonempty and bounded.

Assumption 5: For $i\in\mathcal{V}$, f_i is α-strongly convex.

In Assumption 1, the convexity is to ensure the solvability of the problem (2). The local Lipschitz continuity in Assumption 2 has a necessary role in reformulating (2) by the exact penalty method. The setting of the communication topology in Assumption 3 is widely used to ensure that all agents can reach a consensus by communicating with their neighbors. Assumption 4 is for the existence of a Slater's point, i.e., a point $\tilde{x}\in\mathbb{R}^m$ satisfying $g_i(\tilde{x})<0$, $\forall\, i\in\mathcal{V}$.

Remark 1: The problem (2) is a distributed consensus optimization with different local constraints. Similar problems have been studied in [7], [8], [11], [21]. In particular, [7], [8], [21] solve their problems without guaranteeing the exponential convergence. Although [11] achieves exponential convergence, it requires that the overall constraint set S is available to all agents.

IV. Main Results

In this section, we present the problem transformation, design a distributed algorithm, and analyze its convergence.

A. Problem Transformation

In this part, we aim to transform (2) into an equivalent unconstrained problem, which is easy to solve.

By introducing the consensus constraint, (2) is rewritten as

$$\begin{aligned}\min\ \ &\sum_{i\in\mathcal{V}}f_i(x_i),\\ \text{s.t.}\ \ &g_i(x_i)\leq\mathbf{0},\, i\in\mathcal{V},\\ &x_i=x_j,\, i\in\mathcal{V}, j\in\mathcal{N}_i.\end{aligned} \tag{3}$$

Define the consensus constraint set as $\hat{S}\triangleq\{\mathbf{x}:x_i=x_j,i\in\mathcal{V},j\in\mathcal{N}_i\}$ and the Lagrangian function $\mathcal{L}$ as $\mathcal{L}(\mathbf{x},\lambda)=f(\mathbf{x})+\lambda^{\mathrm{T}}g(\mathbf{x})=\sum_{i\in\mathcal{V}}f_i(x_i)+\sum_{i\in\mathcal{V}}\sum_{k=1}^{l_i}\lambda_{ik}g_{ik}(x_i)$, where $\mathbf{x}=col(x_1,x_2,\cdots,x_n)$, $f(\mathbf{x})=\sum_{i\in\mathcal{V}}f_i(x_i)$, $\lambda_i=col(\lambda_{i1},\lambda_{i2},\cdots,\lambda_{il_i})$, $\lambda=col(\lambda_1,\lambda_2,\cdots,\lambda_n)$, and $g(\mathbf{x})=col(g_1(x_1),g_2(x_2),\cdots,g_n(x_n))$.

Then the dual problem of (3) is

$$\max_{\lambda\geq 0}q(\lambda),\quad q(\lambda)\triangleq\min_{\mathbf{x}\in\hat{S}}\mathcal{L}(\mathbf{x},\lambda). \tag{4}$$

We present two lemmas for the value of exact penalty parameters and dual optimal solutions that are used in the subsequent analysis.

Lemma 4.1 ([22]): The dual optimal solution λ^* to (4) lies in a compact set $\mathcal{D} \subset \mathbb{R}_+^l$, given by

$$\mathcal{D} \triangleq \left\{ \lambda \in \mathbb{R}_+^l \,|\, \|\lambda\| \le \frac{f(\bar{\mathbf{x}}) - \widetilde{q}}{-\hat{g}} \right\},$$

where $\bar{\mathbf{x}}$ is a Slater's point of (3), $\widetilde{q} = \min_{\mathbf{x} \in \hat{S}} \mathcal{L}(\mathbf{x}, \widetilde{\lambda})$ is a dual function value for an arbitrary $\widetilde{\lambda} \ge 0$, and $\hat{g} = \max_{i=1,\cdots,n,k=1,\cdots,l_i}\{g_{ik}(\bar{\mathbf{x}})\}$.

Lemma 4.2 ([23]): Consider a convex function $P: \mathbb{R}^l \to \mathbb{R}$ with $P(u) = 0$, for $\forall u \le 0$, $P(u) > 0$, if $u_j > 0$ for some $j = 1, \cdots, l$, where $u = col(u_1, u_2, \cdots, u_l)$ and the following penalized problem

$$\begin{aligned} \min\ & \sum_{i \in \mathcal{V}} f_i(x_i) + P(g(\mathbf{x})), \\ \text{s.t.}\ & x_i = x_j,\, i \in \mathcal{V}, j \in \mathcal{N}_i. \end{aligned} \tag{5}$$

In order for the penalized problem (5) and the constrained problem (3) to have the same set of optimal solutions, it is sufficient that there exists a dual optimal solution λ^* such that $u^{\mathrm{T}}\lambda^* < P(u)$, $\forall u$ with $u_j > 0$ for some j.

The following theorem gives an equivalent and unconstrained optimization model for the original problem.

Theorem 4.1: Under Assumptions 1—4, the constrained optimization problem (3) shares the same optimal solutions with the following unconstrained optimization problem:

$$\begin{aligned} \min h(\mathbf{x}),\ h(\mathbf{x}) = & \sum_{i \in \mathcal{V}} f_i(x_i) + K_1 \sum_{i \in \mathcal{V}} \sum_{k=1}^{l_i} \max(0, g_{ik}(x_i)) \\ & + K_2 \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} |x_i - x_j|, \end{aligned} \tag{6}$$

where $K_1 > \frac{f(\bar{\mathbf{x}}) - \widetilde{q}}{-\hat{g}}$ and $K_2 > \frac{n\xi}{2}$ are constants. Here, $\xi = \max_{i \in \mathcal{V}} \xi_i$, where ξ_i is the local Lipschitz constant of $f_i(x) + K_1 \sum_{k=1}^{l_i} \max(0, g_{ik}(x))$ on S_i. The definitions of $\bar{\mathbf{x}}$, $\widetilde{q}$ and $\hat{g}$ are the same as given in Lemma 4.1.

Proof: At first, to deal with different local constraints in (3), we introduce the penalty function $P(g(\mathbf{x})) = K_1 \sum_{i \in \mathcal{V}} \sum_{k=1}^{l_i} \max(0, g_{ik}(x_i))$ with the penalty parameter $K_1 \ge 0$. Then (3) can be transformed into

$$\begin{aligned} \min\ & \sum_{i \in \mathcal{V}} f_i(x_i) + K_1 \sum_{i \in \mathcal{V}} \sum_{k=1}^{l_i} \max(0, g_{ik}(x_i)), \\ \text{s.t.}\ & x_i = x_j,\ i \in \mathcal{V}, j \in \mathcal{N}_i. \end{aligned} \tag{7}$$

To make (3) and (7) equivalent, we need to analyze the value range of K_1. By Lemma 4.2 and the form of the penalty function $P(g(\mathbf{x}))$, if the dual optimal solution λ^* to (3) satisfies $u^{\mathrm{T}}\lambda^* < K_1 \sum_{j=1}^{l} \max(0, u_j)$ for all u with some $u_j > 0$, which implies $K_1 > |\lambda^*|$, then (3) and (7) have the same optimal solution set. It follows from Lemma 4.1 that the upper bound of $\|\lambda^*\|$ is $\frac{f(\bar{\mathbf{x}}) - \widetilde{q}}{-\hat{g}}$, where $\bar{\mathbf{x}}$ is a Slater's point of (3), $\widetilde{q} = \min_{\mathbf{x} \in \hat{S}} \mathcal{L}(\mathbf{x}, \widetilde{\lambda})$ is a dual function value for arbitrary $\widetilde{\lambda} \ge 0$ and $\hat{g} = \max_{i=1,\cdots,n,k=1,\cdots,l_i}\{g_{ik}(\bar{x})\}$. By Assumption 4, there exists a Slater's point $\bar{\mathbf{x}} = col(\bar{x}, \bar{x}, \cdots, \bar{x})$ of (3) satisfying $g_i(\bar{x}) < 0, \forall i \in \mathcal{V}$. Thus, the optimal solution set of (3) is the same as the optimal solution set of (7) when $K_1 > \frac{f(\bar{\mathbf{x}}) - \widetilde{q}}{-\hat{g}}$.

To remove the consensus constraint in (7), we consider the unconstrained optimization problem (6) with the penalty parameters $K_1 > \frac{f(\bar{\mathbf{x}}) - \widetilde{q}}{-\hat{g}}$ and $K_2 \ge 0$. To make (6) and (7) equivalent, we need to analyze the value range of K_2. Let $\mathbf{x}^* = col(x_1^*, x_2^*, \cdots, x_n^*)$ be the solution to (6), $\bar{x}^* = \frac{1}{n}\sum_{i \in \mathcal{V}} x_i^*$ and $\bar{\mathbf{x}}^* = 1_n \otimes \bar{x}^*$. By calculation, $h(\mathbf{x}^*) - h(\bar{\mathbf{x}}^*) = \sum_{i \in \mathcal{V}} \Big(f_i(x_i^*) - f_i(\bar{x}^*) + K_1 \sum_{k=1}^{l_i} \max(0, g_{ik}(x_i^*)) - K_1 \sum_{k=1}^{l_i} \max(0, g_{ik}(\bar{x}^*)) + K_2 \sum_{j \in \mathcal{N}_i} |x_i^* - x_j^*| \Big)$.

Denote $\mathcal{M}_i(x) = f_i(x) + K_1 \sum_{k=1}^{l_i} \max(0, g_{ik}(x))$, for $i \in \mathcal{V}$. By Assumption 2, $\mathcal{M}_i(x)$ is locally Lipschitz continuous on $\tilde{S}$ and its local Lipschitz constant is ξ_i. It follows that $\mathcal{M}_i(x_i^*) - \mathcal{M}_i(\bar{x}^*) + \xi_i \|x_i^* - \bar{x}^*\| \ge 0$. Let $\xi = \max_{i \in \mathcal{V}} \xi_i$. Since $|x_i^* - x_j^*| \ge \|x_i^* - x_j^*\|$, $h(\mathbf{x}^*) - h(\bar{\mathbf{x}}^*) \ge -\xi \sum_{i \in \mathcal{V}} \|x_i^* - \bar{x}^*\| + K_2 \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \|x_i^* - x_j^*\|$.

It follows from $\bar{x}^* = \frac{1}{n}\sum_{i \in \mathcal{V}} x_i^*$ that $\sum_{i \in \mathcal{V}} \|x_i^* - \bar{x}^*\| \le \sum_{i \in \mathcal{V}} \sum_{k=1}^{n} \frac{\|x_i^* - x_k^*\|}{n}$. By Assumption 3, there must be a path $\mathcal{P}_{ik} \subset \mathcal{E}$ connecting agent i and k for any $i, k \in \mathcal{V}$. Thus,

$$\begin{aligned} \sum_{i \in \mathcal{V}} \|x_i^* - \bar{x}^*\| \le & \frac{1}{2n} \sum_{i \in \mathcal{V}} \sum_{k \in \mathcal{V}} \sum_{(r,z) \in \mathcal{P}_{ik}} \|x_r^* - x_z^*\| \\ \le & \frac{1}{2n} \sum_{i \in \mathcal{V}} \sum_{k \in \mathcal{V}} \left(\sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \|x_i^* - x_j^*\| \right) \\ \le & \frac{n}{2} \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \|x_i^* - x_j^*\|, \end{aligned}$$

which implies

$$h(\mathbf{x}^*) - h(\bar{\mathbf{x}}^*) \ge (K_2 - \frac{n\xi}{2}) \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{N}_i} \|x_i^* - x_j^*\|.$$

By $K_2 > \frac{n\xi}{2}$, $h(\mathbf{x}^*) - h(\bar{\mathbf{x}}^*) \ge 0$ and the equality holds if and only if $x_i^* = \bar{x}^*$. Therefore, if $\mathbf{x}^*$ is a solution to (6), then it is also a solution to (7). Conversely, it can be shown that a solution to (7) is also a solution to (6).

Consequently, when $K_1 > \frac{f(\bar{\mathbf{x}}) - \widetilde{q}}{-\hat{g}}$ and $K_2 > \frac{n\xi}{2}$, (3) can be equivalently transformed into (6), and their optimal solution sets are the same. This completes the proof. ■

By Theorem 4.1, when the penalty parameters K_1 and K_2 are selected within the given range, we can obtain an optimal solution to (3) by solving the unconstrained problem (6).

B. Algorithm Design

For (6), we design a distributed differential inclusion algorithm as follows. For each $i \in \mathcal{V}$,

$$\begin{aligned} \dot{x}_i(t) \in & -\partial f_i(x_i(t)) - K_1 \sum_{k=1}^{l_i} \partial \max(0, g_{ik}(x_i(t))) \\ & - K_2 \sum_{j \in \mathcal{N}_i} \partial |x_i(t) - x_j(t)| \end{aligned} \tag{8}$$

where $K_1 > \frac{f(\bar{\mathbf{x}})-\widetilde{q}}{-\hat{g}}$ and $K_2 > \frac{n\xi}{2}$ are constants.

In (8), $-\partial f_i(x_i(t))$ provides a possibly descending direction of f_i, and $-K_1 \sum_{k=1}^{l_i} \partial \max(0, g_{ik}(x_i(t)))$ and $-K_2 \sum_{j\in\mathcal{N}_i} \partial |x_i(t) - x_j(t)|$ are subdifferentials of exact penalty terms with penalty parameters K_1 and K_2 to deal with the consensus constraint and different local constraints. Clearly, (8) is distributed since the ith agent uses only the local data and requires only neighbors' decision variables.

The values of K_1 and K_2 can be derived in a distributed manner, provided that the values of the selected $\bar{x}$ and $\widetilde{\lambda}$ are known to each agent. Here we give a distributed process to obtain K_1.

1) Each agent calculates its own $f_i(\bar{x})$ and $g_{ik}(\bar{x})$, $k = 1, 2, \cdots, l_i$.
2) Then $f(\bar{\mathbf{x}})$ and $\hat{g}$ can be calculated by summing and maximizing operations, respectively. Since the network topology graph $\mathcal{G}$ is connected, agents i can communicate with their neighbors on the values of $f_j(\bar{x})$ and $g_{jk}(\bar{x})$, $k = 1, 2, \cdots, l_i$, $j \in \mathcal{N}_i$.
3) We can use the algorithm in [24] to calculate $\widetilde{q}$, which is the optimal value of an unconstrained optimization problem.

Then we can obtain K_1. Note that the Lipschitz constant ξ_i of $\mathcal{M}_i$ can be obtained since the local Lipschitz constants of f_i and g_i are available. We can refer to the method in [11] to calculate ξ and K_2 in a distributed manner.

Remark 2: Compared with [7], [8], the dual and auxiliary variables in their algorithms are not needed in our algorithm, since our method is based on the exact penalty to remove the constraints. As a result, the cost of computation and communication with respect to those variables is reduced in our algorithm.

C. Existence and Convergence Analysis

The compact form of (8) can be written as

$$\dot{\mathbf{x}}(t) \in H(\mathbf{x}(t)), \tag{9}$$

where $H(\mathbf{x}(t)) = -\partial h(\mathbf{x}(t))$ and h is defined in (6). The existence of a trajectory solution to (9) is presented as follows.

Theorem 4.2: Under Assumptions 1—4, for any initial value $\mathbf{x}(0) \in \mathbb{R}^{nm}$, a solution $\mathbf{x}(t)$ to (9) exists globally and is bounded.

Proof: By Assumption 1 and Lemma 2.1, the right-hand side of (9) is an upper semicontinuous set-valued map with nonempty convex compact values, and (9) has at least one solution on $[0, T]$ with $T > 0$ for any initial value $\mathbf{x}(0) \in \mathbb{R}^{nm}$. Let $\mathbf{x}^* = col(x_1^*, x_2^*, \cdots, x_n^*)$ be an optimal solution to (6). By the first-order optimality condition, for each $i \in \mathcal{V}$,

$$\mathbf{0} \in \partial f_i(x_i^*) + K_1 \sum_{k=1}^{l_i} \partial \max(0, g_{ik}(x_i)) + K_2 \sum_{j\in\mathcal{N}_i} \partial |x_i^* - x_j^*| \tag{10}$$

Define

$$V(t) \triangleq \frac{1}{2} \sum_{i\in\mathcal{V}} \|x_i(t) - x_i^*\|^2. \tag{11}$$

The set-valued Lie derivative of $V(t)$ with respect to (9) is

$$\begin{aligned} \mathcal{L}_H V(t) = \Big\{ & a(t) \in \mathbb{R} : a(t) = \sum_{i\in\mathcal{V}} \Big\langle x_i(t) - x_i^*, -\eta_i \\ & -K_1 \sum_{k=1}^{l_i} \gamma_{ik}(t) - K_2 \sum_{j\in\mathcal{N}_i} \zeta_{ij}(t) \Big\rangle, \eta_i(t) \in \partial f_i(x_i(t)), \\ & \gamma_{ik}(t) \in \partial \max(0, g_{ik}(x_i(t))), \zeta_{ij}(t) \in \partial |x_i(t) - x_j(t)| \Big\}. \end{aligned} \tag{12}$$

Let $W(t) = h(\mathbf{x}(t)) - h(\mathbf{x}^*) \geq 0$. It follows from the basic property of the subdifferential that

$$\begin{aligned} a(t) \leq & \sum_{i\in\mathcal{V}} (f_i(x_i^*) - f_i(x_i(t))) \\ & + K_1 \sum_{i\in\mathcal{V}} \Big(\sum_{k=1}^{l_i} \max(0, g_{ik}(x_i^*)) - \sum_{k=1}^{l_i} \max(0, g_{ik}(x_i(t))) \Big) \\ & + K_2 \sum_{i\in\mathcal{V}} \Big(\sum_{j\in\mathcal{N}_i} |x_i^* - x_j^*| - \sum_{j\in\mathcal{N}_i} |x_i(t) - x_j(t)| \Big) \\ & \leq h(\mathbf{x}^*) - h(\mathbf{x}(t)) = -W(t) \leq 0. \end{aligned} \tag{13}$$

Therefore, $V(t)$ is monotonically decreasing and $V(t) \leq V(0)$, $\forall\, t \in [0, T]$. Thus, $\mathbf{x}(t)$ is bounded on $[0, T]$. By the extension theorem given in [19], $\mathbf{x}(t)$ exists globally and remains bounded. This completes the proof. ■

Theorem 4.2 proves the boundedness and global existence of a solution $\mathbf{x}(t)$ to (9). Next, we present the convergence analysis as follows.

Theorem 4.3: Under Assumptions 1—4, the trajectory $\mathbf{x}(t)$ converges to an equilibrium point $\hat{\mathbf{x}}^*$, which is also an optimal solution to (3) and (6).

Proof: Let $\mathbf{x}^* = col(x_1^*, x_2^*, \cdots, x_n^*)$ be an equilibrium point of (9). By (10), $\mathbf{x}^*$ is also the optimal solution to (6).

Consider the Lyapunov function (11). By (13), for any $a(t) \in \mathcal{L}_H V$, $a(t) \leq 0$. Recall that $W(t) = h(\mathbf{x}(t)) - h(\mathbf{x}^*) \geq 0$. Since $h(\cdot)$ is locally Lipschitz continuous and $\mathbf{x}(t)$ is absolutely continuous, $W(t)$ is uniformly continuous with respect to t. Therefore, $W(t)$ is Riemann integrable. It follows from (13) that $\int_0^\infty W(\mathbf{x}(\tau))d\tau \leq V(0) < \infty$. By Lemma 2.2, $\lim_{t\to\infty} W(\mathbf{x}(t)) = 0$, i.e., $\mathbf{x}(t)$ converges to the set $\Phi = \{\mathbf{x} \in \mathbb{R}^{nm} \,|\, W(\mathbf{x}) = 0\}$.

Next, we prove that $\mathbf{x}(t)$ converges to one point in Φ. By the convergence of $\mathbf{x}(t)$, there exists a strictly increasing sequence $\{t_k\}$ with $\lim_{k\to\infty} t_k = \infty$ such that $\lim_{k\to\infty} \mathbf{x}(t_k) = \hat{\mathbf{x}}^*$, $\hat{\mathbf{x}}^* \in \Phi$. Consider a new Lyapunov function $\widehat{V}(t) = \frac{1}{2} \sum_{i\in\mathcal{V}} \|x_i(t) - \hat{x}_i^*\|^2$. Then $\liminf_{t\to\infty} \widehat{V}(t) = 0$. Also, by similar analysis about V in (11), $\widehat{V}(t)$ is also monotonically decreasing. As a result, $\lim_{t\to\infty} \widehat{V}(t) = 0$, which implies $\lim_{t\to\infty} \mathbf{x}(t) = \hat{\mathbf{x}}^*$.

Since $\hat{\mathbf{x}}^* \in \Phi$, it is an optimal solution to (6). When K_1 and K_2 satisfy the value range given in Theorem 4.1, $\hat{\mathbf{x}}^*$ is also an optimal solution to (3). This completes the proof. ■

Remark 3: The convergence of (9) is proved under Assumptions 1—4 without Assumption 5, which is different from the strict convexity required in [5].

Further we analyze the convergence rate in the following.

Theorem 4.4: Under Assumptions 1—4, (9) converges to its equilibrium point with $\mathcal{O}(\frac{1}{t})$ rate as

$$0 \le h(\hat{\mathbf{x}}(t)) - h(\mathbf{x}^*) \le \frac{1}{t}V(0), \ \forall t \ge 0$$

where $\hat{\mathbf{x}}(t) \triangleq \frac{1}{t}\int_0^t \mathbf{x}(\tau)d\tau$.

Proof: Let $\mathbf{x}^*$ be an equilibrium point of (9). Moreover, $\mathbf{x}^*$ is also the minimum point of $h(\mathbf{x})$. Then $h(\hat{\mathbf{x}}(t)) - h(\mathbf{x}^*) \ge 0$. It follows from (13) that

$$a \le h(\mathbf{x}^*) - h(\mathbf{x}(t)). \tag{14}$$

Integrating both sides of (14) over the interval $[0, t)$ yields

$$-\frac{1}{t}V(0) \le \frac{1}{t}\big(V(t) - V(0)\big) \le \frac{1}{t}\int_0^t \big(h(\mathbf{x}^*) - h(\mathbf{x}(\tau))\big)d\tau.$$

Since $h(\mathbf{x})$ is convex with respect to $\mathbf{x}$, using the Jensen's inequality yields $h(\hat{\mathbf{x}}(t)) \le \frac{1}{t}\int_0^t h(\mathbf{x}(\tau))d\tau$. Therefore, $h(\hat{\mathbf{x}}(t)) \le \frac{1}{t}V(0) + \frac{1}{t}\int_0^t h(\mathbf{x}^*)d\tau$, i.e., $h(\hat{\mathbf{x}}(t)) - h(\mathbf{x}^*) \le \frac{1}{t}V(0)$. This completes the proof. ■

In Theorem 4.4, we prove that when the cost functions are convex, (9) achieves $\mathcal{O}(\frac{1}{t})$ rate of convergence. This result has also been obtained in [5], whereas their cost functions are required to be strictly convex. Moroever, our method can also guarantee the exponential convergence with strongly convex functions, as shown in the next theorem.

Theorem 4.5: Under Assumptions 1—5, (9) converges exponentially to its equilibrium point with

$$V(t) \le V(0)e^{-2\alpha t}, \ \forall t \ge 0$$

where the rate α is the strong convexity parameter.

Proof: By Assumption 5, (6) has only one optimal solution, and the equilibrium of (9) is also unique, denoted by $\mathbf{x}^* = col(x_1^*, x_2^*, \cdots, x_n^*)$. Consider the Lyapunov function (11). It follows from (12) that for any $a \in \mathcal{L}_H V$,

$$a = \sum_{i\in\mathcal{V}} \Big\langle x_i - x_i^*, -\eta_i - K_1\sum_{k=1}^{l_i}\gamma_{ik} - K_2\sum_{j\in\mathcal{N}_i}\zeta_{ij}\Big\rangle,$$
$$\eta_i \in \partial f_i(x_i), \gamma_{ik} \in \partial\max(0, g_{ik}(x_i)), \zeta_{ij} \in \partial|x_i - x_j|. \tag{15}$$

Moreover, $\mathbf{x}^*$ is the optimal solution to (6) if and only if

$$\Big\langle x_i - x_i^*, \eta_i^* + K_1\sum_{k=1}^{l_i}\gamma_{ik}^* + K_2\sum_{j\in\mathcal{N}_i}\zeta_{ij}^*\Big\rangle \ge 0, \forall x_i \in \mathbb{R}^m, \tag{16}$$

where $\eta_i^* \in \partial f_i(x_i^*), \gamma_i^* \in \partial\max(0, g_i(x_i^*)), \zeta_{ij}^* \in \partial|x_i^* - x_j^*|$. Substituting (16) into (15) yields

$$\begin{aligned} a \le &-\sum_{i\in\mathcal{V}}\Big\langle x_i - x_i^*, \eta_i - \eta_i^* + K_1\sum_{k=1}^{l_i}\gamma_{ik} - K_1\sum_{k=1}^{l_i}\gamma_{ik}^* \\ &+ K_2\sum_{j\in\mathcal{N}_i}\zeta_{ij} - K_2\sum_{j\in\mathcal{N}_i}\zeta_{ij}^*\rangle \\ \le &-\sum_{i\in\mathcal{V}}\langle x_i - x_i^*, \eta_i - \eta_i^*\rangle. \end{aligned}$$

By Assumption 5, $a \le -\sum_{i\in\mathcal{V}}\alpha\|x_i - x_i^*\|^2 = -2\alpha V$. Then $V(t) \le V(0)e^{-2\alpha t}$, which completes the proof. ■

Remark 4: The exponential convergence under the strong convexity condition has also been considered in [11], [15], [16] without local constraints or with global constraints. In addition, it has been achieved in [10] for optimization problems with local linear equality constraints, but the exponential convergence cannot be guaranteed in the presence of inequality constraints. In contrast, our algorithm achieves exponential convergence for optimization problems with different local inequality constraints.

V. Numerical Simulation

In this section, we give two numerical simulation examples to verify the effectiveness of (9).

Example 1: Consider two agents which can communicate with each other for the optimization problem (2). Their cost functions are $f_1(\mathbf{x}) = x_1^2 + x_2^2$, $f_2(\mathbf{x}) = x_1^2 + (x_2 - 0.19)^2$, where $\mathbf{x} \in \mathbb{R}^2$ is the decision variable. The local constraints $g_1(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}^4$, $g_2(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}^5$ are $g_1(\mathbf{x}) = col(-x_1 - 2, x_1, -x_2 - 2, x_2)$, $g_2(\mathbf{x}) = col(-x_1 - 2, x_1, -x_2 - 0.01, x_2 - 0.19, 0.1x_1 - x_2 + 0.19)$.

We make a simulation comparison between our algorithm and the one in [11] as shown in Figures 1 and 2. Our algorithm makes the agents reach a consensus and converge to the optimal solution, whereas the algorithm in [11] fails to find the correct solution.

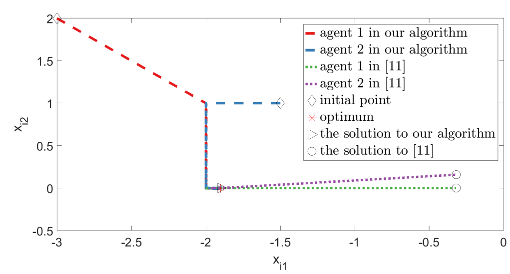

Fig. 1. The comparison of our algorithm and the algorithm in [11].

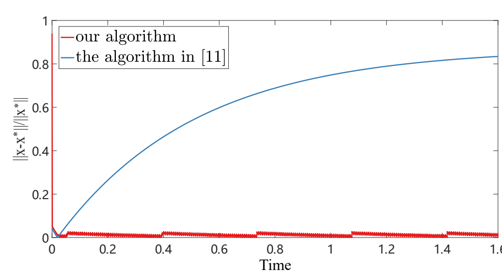

Fig. 2. The error comparison of our algorithm and the algorithm in [11].

Example 2: We discuss the case of strongly convex cost functions to verify the exponential convergence. Consider the optimization problem (2) with a multi-agent network consisting of 30 agents connected by a cyclic graph. For $i = 1, 2, \cdots, 30$, $f_i(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T P_i\mathbf{x} + q_i^T\mathbf{x} + r_i|\mathbf{x}|$, where $\mathbf{x} \in \mathbb{R}^{10}$

is the decision variable, $P_i \in \mathbb{R}^{10\times10}$ is a positive definite matrix, and $q_i \in \mathbb{R}^{10}$, $r_i \in \mathbb{R}^1_{>0}$ are randomly generated known coefficients. The local constraints are $g_i(\mathbf{x}) \leq \mathbf{0}$, $i = 1, 2, \cdots, 30$, where $g_i(\mathbf{x}) : \mathbb{R}^{10} \to \mathbb{R}^2$ and $g_i(\mathbf{x}) =$

$$\begin{cases} col(x_1 + \cdots + x_i^2 + \cdots + x_{10} - 9 - i, 0), \ i = 1, \cdots, 10, \\ col(x_1 + x_2 + \cdots + x_{10} + 6 - i, 0), \ i = 11, \cdots, 20, \\ col(x_{i-20} + 16 - i, -x_{i-20} + 16 - i), \ i = 21, \cdots, 30. \end{cases}$$

The simulation results are shown in Figures 3 and 4, where V is given in (11). As can be seen from Figure 3, our algorithm achieves exponential convergence. Also, Figure 4 shows that both our algorithm and the adaptive one in [7] converge to the optimal value, while our algorithm converges fast.

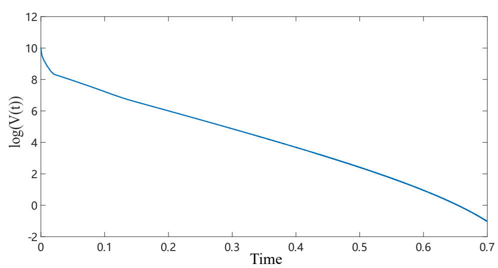

Fig. 3. Exponential convergence of the algorithm.

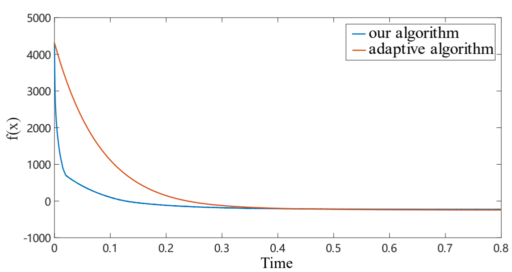

Fig. 4. Comparison of our algorithm and the adaptive algorithm.

VI. Conclusion

In this paper, a distributed continuous-time differential inclusion algorithm has been designed to solve a distributed optimization problem with different local constraints. A novel method by removing both the consensus constraint and different local constraints based on the exact penalty have been proposed. Moreover, the feasibility and convergence of the algorithm have been proved theoretically. The convergence rate of $\mathcal{O}(\frac{1}{t})$ has been achieved when the cost functions are convex, and the exponential convergence rate has been achieved when the cost functions are strongly convex. Finally, the convergence and effectiveness of the algorithm have been verified by two numerical examples.

References

[1] M. Blot, D. Picard, N. Thome, and M. Cord, "Distributed optimization for deep learning with gossip exchange," *Neurocomputing*, vol. 330, pp. 287–296, 2019.

[2] X. Jiang, S. Qin, and X. Xue, "A penalty-like neurodynamic approach to constrained nonsmooth distributed convex optimization," *Neurocomputing*, vol. 377, pp. 225–233, 2020.

[3] D. Wang, M. Chen, and W. Wang, "Distributed extremum seeking for optimal resource allocation and its application to economic dispatch in smart grids," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 10, pp. 3161–3171, 2019.

[4] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson, "A survey of distributed optimization," *Annual Reviews in Control*, vol. 47, pp. 278–305, 2019.

[5] S. Liang, X. Zeng, and Y. Hong, "Distributed nonsmooth optimization with coupled inequality constraints via modified lagrangian function," *IEEE Transactions on Automatic Control*, vol. 63, no. 6, pp. 1753–1759, 2017.

[6] K. Lu, G. Jing, and L. Wang, "Distributed algorithm for solving convex inequalities," *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 2670–2677, 2017.

[7] W. Jia and S. Qin, "An adaptive penalty-like continuous-time algorithm to constrained distributed convex optimization," *Journal of the Franklin Institute*, vol. 359, no. 8, pp. 3692–3716, 2022.

[8] C. Xu, Q. Liu, and T. Huang, "Resilient penalty function method for distributed constrained optimization under byzantine attack," *Information Sciences*, vol. 596, pp. 362–379, jun 2022.

[9] D. Yuan, D. W. Ho, and G.-P. Jiang, "An adaptive primal-dual subgradient algorithm for online distributed constrained optimization," *IEEE Transactions on Cybernetics*, vol. 48, no. 11, pp. 3045–3055, 2017.

[10] Y. Zhu, W. Yu, G. Wen, and G. Chen, "Projected primal–dual dynamics for distributed constrained nonsmooth convex optimization," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1776–1782, 2018.

[11] W. Li, X. Zeng, S. Liang, and Y. Hong, "Exponentially convergent algorithm design for constrained distributed optimization via nonsmooth approach," *IEEE Transactions on Automatic Control*, vol. 67, no. 2, pp. 934–940, 2021.

[12] S. A. Alghunaim and A. H. Sayed, "Linear convergence of primal–dual gradient methods and their performance in distributed optimization," *Automatica*, vol. 117, p. 109003, jul 2020.

[13] L. Romao, K. Margellos, G. Notarstefano, and A. Papachristodoulou, "Subgradient averaging for multi-agent optimisation with different constraint sets," *Automatica*, vol. 131, p. 109738, 2021.

[14] S. A. Alghunaim, E. K. Ryu, K. Yuan, and A. H. Sayed, "Decentralized proximal gradient algorithms with linear convergence rates," *IEEE Transactions on Automatic Control*, vol. 66, no. 6, pp. 2787–2794, 2020.

[15] A. Makhdoumi and A. Ozdaglar, "Convergence rate of distributed ADMM over networks," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 5082–5095, 2017.

[16] J. Zhang, K. You, and L. Xie, "Innovation compression for communication-efficient distributed optimization with linear convergence," *IEEE Transactions on Automatic Control*, pp. 1–8, 2023.

[17] J. Cortés and S. K. Niederländer, "Distributed coordination for nonsmooth convex optimization via saddle-point dynamics," *Journal of Nonlinear Science*, vol. 29, no. 4, pp. 1247–1272, nov 2018.

[18] I. Masubuchi, T. Wada, T. Asai, N. T. H. Linh, Y. Ohta, and Y. Fujisaki, "Distributed constrained optimization protocol via an exact penalty method," in *2015 European Control Conference (ECC)*. IEEE, jul 2015.

[19] J.-P. Aubin and A. Cellina, *Differential Inclusions*. Berlin, German: Springer, 1984.

[20] W. M. Haddad and V. Chellaboina, *Nonlinear dynamical systems and control*. Princeton, NJ, USA: Princeton University Press, 2011.

[21] H. Zhou, X. Zeng, and Y. Hong, "Adaptive exact penalty design for constrained distributed optimization," *IEEE Transactions on Automatic Control*, vol. 64, no. 11, pp. 4661–4667, 2019.

[22] 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.

[23] D. Bertsekas, *Convex optimization algorithms*. Nashua, NH, USA: Athena Scientific, 2015.

[24] 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, 2015.