### Title
Distributed Continuous-Time Optimization: Nonuniform Gradient Gains, Finite-Time Convergence, and Convex Constraint Set




### Journal
IEEE Transactions on Automatic Control, 62(5)

### ISSN
0018-9286 1558-2523



### Authors
Lin, Peng
Ren, Wei
Farrell, Jay A



### Publication Date
2017-05-01

### DOI
10.1109/TAC.2016.2604324

Peer reviewed

# Distributed Continuous-Time Optimization: Nonuniform Gradient Gains, Finite-Time Convergence, and Convex Constraint Set

Peng Lin, Wei Ren, *Fellow, IEEE*, and Jay A. Farrell, *Fellow, IEEE*

***Abstract*—In this paper, a distributed optimization problem with general differentiable convex objective functions is studied for continuous-time multi-agent systems with single-integrator dynamics. The objective is for multiple agents to cooperatively optimize a team objective function formed by a sum of local objective functions with only local interaction and information while explicitly taking into account nonuniform gradient gains, finite-time convergence, and a common convex constraint set. First, a distributed nonsmooth algorithm is introduced for a special class of convex objective functions that have a quadratic-like form. It is shown that all agents reach a consensus in finite time while minimizing the team objective function asymptotically. Second, a distributed algorithm is presented for general differentiable convex objective functions, in which the interaction gains of each agent can be self-adjusted based on local states. A corresponding condition is then given to guarantee that all agents reach a consensus in finite time while minimizing the team objective function asymptotically. Third, a distributed optimization algorithm with state-dependent gradient gains is given for general differentiable convex objective functions. It is shown that the distributed continuous-time optimization problem can be solved even though the gradient gains are not identical. Fourth, a distributed tracking algorithm combined with a distributed estimation algorithm is given for general differentiable convex objective functions. It is shown that all agents reach a consensus while minimizing the team objective function in finite time. Fifth, as an extension of the previous results, a distributed constrained optimization algorithm with nonuniform gradient gains and a distributed constrained finite-time optimization algorithm are given. It is shown that both algorithms can be used to solve a distributed continuous-time optimization problem with a common convex constraint set. Numerical examples are included to illustrate the obtained theoretical results.**

Manuscript received March 4, 2016; revised June 24, 2016; accepted August 14, 2016. Date of publication August 30, 2016; date of current version April 24, 2017. This work was supported by the National Science Foundation under Grant ECCS-1611423, the National Natural Science Foundation of China under Grants 61203080, 61573082, 61528301, and the State Key Laboratory of Intelligent Control and Decision of Complex Systems of Beijing Institute of Technology. Recommended by Associate Editor Prof. Z. Sun.

Peng Lin is with the School of Information Science and Engineering, Central South University, Changsha 410083, China (e-mail: lin_peng0103@sohu.com).

Wei Ren and Jay A. Farrell are with the Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521, USA (e-mail: ren@ee.ucr.edu; jay.farrell@ucr.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2016.2604324

***Index Terms*—Consensus, convex set constraint, distributed optimization, finite-time convergence, multi-agent systems, nonuniform gradient gains.**

## I. Introduction

DISTRIBUTED optimization problems for multi-agent systems have received significant attention in the control community [1]–[16]. The objective is to solve an optimization problem cooperatively in a distributed manner where the team objective function is composed of a sum of local objective functions, each of which is known to only one agent. Earlier work about distributed optimization problems mostly concentrated on discrete-time algorithms. For example, article [1] gave a discrete-time projection algorithm where each agent is required to lie in a closed convex set and showed that all agents reach an optimal point in the intersection of all the convex sets even when the communication topologies are dynamically changing as long as the edge weight matrix is doubly stochastic. Moreover, articles [4]–[8] addressed distributed optimization problems with inequality-equality constraints or using other discrete-time algorithms and derived conditions to ensure that all agents converge to the optimal point or its neighborhood. Recently, some researchers turned their attention to distributed optimization problems for multi-agent systems with continuous-time dynamics. For example, article [9] proposed a continuous-time zero-gradient-sum algorithm. Article [10] and its extension [16] studied a continuous-time version of the work in [1]. Article [11] studied the continuous-time distributed optimization problem for undirected graphs and derived explicit expressions for a lower bound on the algorithm's convergence rate. Article [12] proposed a novel distributed continuous-time algorithm for distributed convex optimization by introducing a dynamic integrator. Founded on the work of [12], articles [13], [14] studied the distributed continuous-time optimization problem for general strongly connected balanced directed graphs and gave the estimate of the convergence rate of the algorithm.

Although excellent work has been presented in [1]–[16] to solve the distributed optimization problem, there are still issues that need to be addressed, in particular when nonuniform gradient or subgradient gains and finite-time convergence are taken into account. For example, in [5], a distributed optimization problem with nonuniform subgradients was studied from a view point of stochastic theory, but by taking the mathematical expectation, it can be seen that the given algorithm uses uniform subgradient gains in nature. In reality, it is a difficult task to keep the gradient or subgradient gains uniform for different agents all the time, in particular when the number of agents is huge. It is important and necessary to design algorithms for distributed optimization with nonuniform gradient gains. However, due to coexistence of the nonuniformity of the gradient gains and the nonlinearity of the gradients of the local objective functions, the convergence rates of the local objective functions are no longer uniform for the distributed optimization problem with nonuniform gradient gains and hence the existing approaches cannot be applied directly. Besides this aspect, most of the existing works on the distributed optimization problem (e.g., [1]–[16]), studied only the asymptotical stability of the algorithm and rare results are concerned about the finite-time convergence of the algorithms. Due to the existence and the nonlinearity of the objective functions, the existing approaches for the distributed finite-time consensus problem (e.g., [17], [18]) cannot be extended directly to the distributed finite-time optimization problems. Though some results have been obtained in our previous works in [19], [20] for the distributed finite-time optimization problem, they are limited to a special class of convex objective functions that have a quadratic-like form and the approaches cannot be applied to more general convex objective functions. It is meaningful and challenging to study the distributed finite-time optimization problem for more general convex objective functions.

In this paper, we are interested in solving the distributed optimization problem with general differentiable convex objective functions for continuous-time multi-agent systems with the consideration of nonuniform gradient gains, finite-time convergence, and a common convex constraint set. First, a distributed nonsmooth algorithm is introduced for a special class of convex objective functions that have a quadratic-like form. It is shown that all agents reach a consensus in finite time while minimizing the team objective function asymptotically. Second, an adaptive distributed algorithm is presented where the interaction gains of each agent can be self-adjusted based on local states. It is shown that the distributed continuous-time optimization problem can be solved when general differentiable convex local objective functions are taken into account. Third, to relax the synchronization requirement of the system on the gains of the gradients, a distributed algorithm with state-dependent gradient gains is given. It is shown that the optimization problem can be solved even though the gradient gains are not identical. After that, a distributed tracking algorithm combined with a distributed estimation algorithm is given. It is shown that all agents reach a consensus while minimizing the team objective function in finite time. Finally, as an extension of the previous results, a distributed constrained optimization algorithms with nonuniform gradient gains and a distributed constrained finite-time optimization algorithm are given. It is shown that both algorithms can be used to solve a distributed continuous-time optimization problem with a common convex constraint set.

## II. Notation and Preliminaries

We adopt the following notation throughout this paper: $\mathbb{R}^m$ denotes the set of all $m$ dimensional real column vectors; $\mathbb{R}^{m\times n}$ denotes the set of all $m\times n$ real matrices; $\mathcal{I}$ denotes the index set $\{1,\ldots,n\}$; $s_i$ denotes the $i$th component of the vector $s$; $A_{ij}$ denotes the $ij$th entry of the matrix $A$; $s^T$ and $A^T$ denote, respectively, the transpose of the vector $s$ and the matrix $A$; $||s||$ denotes the Euclidean norm of the vector $s$; $\frac{\mathrm{d}}{\mathrm{d}s}$ and $\frac{\partial}{\partial s}$ denote, respectively, the differential operator and partial differential operator with respect to $s$; $\nabla f(s)$ denotes the gradient of the function $f(s)$ at $s$ with $[\nabla f(s)]_i = \frac{\partial f(s)}{\partial s_i}$; the matrix $\nabla^2 f(s)$ denotes the Hessian or second-order partial derivative matrix of the function $f(s)$ at $s$ with $[\nabla^2 f(s)]_{ij} = \frac{\partial^2 f(s)}{\partial s_i \partial s_j}$; $\mathrm{sgn}(s)$ denotes a component-wise sign function of $s$; the symbol / denotes the division sign; $Y - X$ denotes the relative complement set of $X$ in $Y$ for any two sets $X$ and $Y$; and $P_X(s)$ denotes the projection of the vector $s$ onto the closed convex set $X$, i.e., $P_X(s) = \arg\min_{\bar{s}\in X}\|s-\bar{s}\|$.

Let $\mathcal{G}(\mathcal{V},\mathcal{E},\mathcal{A})$ be a graph of order $n$, where $\mathcal{V}=\{1,\ldots,n\}$ is the set of nodes, $\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ is the set of edges, and $\mathcal{A}=[a_{ij}]\in\mathbb{R}^{n\times n}$ is the weighted adjacency matrix. An edge of $(i,j)\in\mathcal{E}$ denotes that agent $i$ can obtain information from agent $j$. The weighted adjacency matrix $\mathcal{A}$ is defined as $a_{ii}=0$ and $a_{ij}=1$ if $(i,j)\in\mathcal{E}$ and $a_{ij}=0$ otherwise. The graph $\mathcal{G}$ is undirected if $a_{ij}=a_{ji}$ for all $i,j$. The set of neighbors of node $i$ is denoted by $N_i=\{j\in\mathcal{V}:(i,j)\in\mathcal{E}\}$. A path is a sequence of edges of the form $(i_1,i_2),(i_2,i_3),\ldots$, where $i_j\in\mathcal{V}$. The graph $\mathcal{G}$ is connected, if there is a path from every node to every other node.

*Lemma 1: [22]* Let $f_0(s):\mathbb{R}^m\to\mathbb{R}$ be a differentiable convex function. $f_0(s)$ is minimized if and only if $\nabla f_0(s)=0$.

*Lemma 2: [23]* Suppose that $Y\neq\emptyset$ is a closed convex set in $\mathbb{R}^m$. The following two statements hold:

(a) For any $y\in\mathbb{R}^m$, $\|y-P_Y(y)\|$ is continuous with respect to $y$ and $\nabla\frac{1}{2}\|y-P_Y(y)\|^2 = y-P_Y(y)$.

(b) For any $y,z\in\mathbb{R}^m$ and all $\xi\in Y$, $[y-P_Y(y)]^T(y-\xi)\geq 0$, $\|P_Y(y)-\xi\|^2\leq\|y-\xi\|^2-\|P_Y(y)-y\|^2$ and $\|P_Y(y)-P_Y(z)\|\leq\|y-z\|$.

## III. Distributed Continuous-Time Optimization Without Constraints

Consider a multi-agent system consisting of $n$ agents. Each agent is regarded as a node in an undirected graph $\mathcal{G}(t)$, and each agent can interact with only its neighbors. Suppose that the agents satisfy the continuous-time dynamics

$$\dot{x}_i(t) = u_i(t),\quad i\in\mathcal{I}, \tag{1}$$

where $x_i(t)\in\mathbb{R}^m$ is the state of agent $i$, and $u_i(t)\in\mathbb{R}^m$ is the control input of agent $i$. Our objective is to design $u_i(t)$ using only local interaction and information, such that all agents cooperatively find the optimal state $s^*$ that solves the optimization problem

$$\text{minimize } \textstyle\sum_{i=1}^n f_i(s) \text{ subject to } s\in\mathbb{R}^m,$$

where $f_i(s):\mathbb{R}^m\to\mathbb{R}$ denotes the local objective function of agent $i$, which is convex, differentiable, and known only to agent $i$. Clearly, $\sum_{i=1}^n f_i(s)$ is also a differentiable function. The problem described above is equivalent to the problem that all agents reach a consensus while minimizing the team objective function $\sum_{i=1}^n f_i(x_i)$, i.e.,

$$\text{minimize } \textstyle\sum_{i=1}^n f_i(x_i) \text{ subject to } x_i = x_j\in\mathbb{R}^m. \tag{2}$$

*Assumption 1:* Each set $X_i \triangleq \{s | \nabla f_i(s) = 0\}$ is nonempty and bounded.

To illustrate Assumption 1, we consider the convex function $f_i(s): \mathbb{R}^2 \to \mathbb{R}$:

$$f_i(s) = \begin{cases} 0, & \text{if } \|s\| \le 1, \\ 0.5(\|s\|-1)^2, & \text{otherwise.} \end{cases}$$

By simple calculations, we have

$$\nabla f(s) = \begin{cases} [0,0]^T, \text{if } \|s\| \le 1, \\ (\|s\|-1)\frac{s}{\|s\|} \neq [0,0]^T, \text{otherwise.} \end{cases}$$

Clearly, $X_i = \{s | \|s\| \le 1\}$ and hence $f_i(s)$ satisfies Assumption 1.

In Assumption 1, we only make assumptions on each $f_i(s)$ rather than the team objective function $\sum_{i=1}^n f_i(s)$ because the team objective function $\sum_{i=1}^n f_i(s)$ is a global information for all agents and cannot be used by each agent in a distributed way.

Define $X \triangleq \{s | \sum_{i=1}^n \nabla f_i(s) = 0\}$. From Lemma 1, all $X_i$ and $X$ are the minimum sets of $f_i(s)$ and $\sum_{i=1}^n f_i(s)$ for all $i$.

*Lemma 3:* Let $f(s): \Xi \mapsto \mathbb{R}$ be a differentiable convex function and $Y$ be its minimum set in $\Xi$, where $\Xi \subseteq \mathbb{R}^n$ is a closed convex set. Suppose that $Y$ is bounded and closed. For any $z = \lambda P_Y(y) + (1-\lambda)y$ with $\lambda \in (0,1)$,

$$0 < \nabla f(z)^T \frac{y - P_Y(y)}{\|y - P_Y(y)\|} \le \nabla f(y)^T \frac{y - P_Y(y)}{\|y - P_Y(y)\|}$$

for any $y \in \Xi - Y$.

*Proof:* Let

$$z = \lambda P_Y(y) + (1-\lambda)y$$

with $\lambda \in (0,1)$ for any $y \in \Xi - Y$. Clearly, $z \in \Xi - Y$. From the convexity of the function $f(s)$, we have $f(y) - f(z) \ge \nabla f(z)^T(y-z)$ and $f(z) - f(y) \ge \nabla f(y)^T(z-y)$. Thus, $0 = f(y) - f(z) + f(z) - f(y) \ge [\nabla f(z) - \nabla f(y)]^T (y - z)$. Note that $y \neq z$, $y \neq P_Y(y)$ and $z \neq P_Y(y)$. Hence $\frac{y-z}{\|y-z\|} = \frac{y-P_Y(y)}{\|y-P_Y(y)\|} = \frac{z-P_Y(y)}{\|z-P_Y(y)\|}$, and $\nabla f(y)^T \frac{y-P_Y(y)}{\|y-P_Y(y)\|} = \nabla f(y)^T \frac{y-z}{\|y-z\|} \ge \nabla f(z)^T \frac{y-z}{\|y-z\|} = \nabla f(z)^T \frac{y-P_Y(y)}{\|y-P_Y(y)\|}$, i.e., $\nabla f(y)^T \frac{y-P_Y(y)}{\|y-P_Y(y)\|} \ge \nabla f(z)^T \frac{y-P_Y(y)}{\|y-P_Y(y)\|}$. Furthermore, since $P_Y(y) \in Y$ and $Y$ is the minimum set, from the convexity of the function $f(s)$, we have $0 > f(P_Y(y)) - f(z) \ge \nabla f(z)^T (P_Y(y) - z)$. That is, $\nabla f(z)^T \frac{z-P_Y(y)}{\|z-P_Y(y)\|} > 0$. Recalling that $\frac{y-P_Y(y)}{\|y-P_Y(y)\|} = \frac{z-P_Y(y)}{\|z-P_Y(y)\|}$, we have $\nabla f(z)^T \frac{y-P_Y(y)}{\|y-P_Y(y)\|} > 0$. ■

*Lemma 4:* Under Assumption 1, the following two statements hold:

(1) $\lim_{\|y\|\to+\infty} f_i(y) = +\infty$ for all $i$ and accordingly $\lim_{\|y\|\to+\infty} \sum_{i=1}^n f_i(y) = +\infty$.

(2) All $X_i$ and $X$ are nonempty closed bounded convex sets for all $i$.

*Proof:* From the convexity of the functions $f_i(s)$, $\sum_{i=1}^n f_i(s)$ is convex and hence all $X_i$ and $X$ are also convex. Under Assumption 1, each $X_i$ is nonempty and bounded. Now, we prove that all $X_i$ are closed sets. If this is not true, there exists an integer $i_e$ such that $X_{i_e}$ is an open set. Then there must exist a vector $s_e \notin X_{i_e}$ and a vector sequence $\{\tilde{s}_k \in X_{i_e}\}$ such that $s_e = \lim_{k\to+\infty} \tilde{s}_k$. Clearly, $f_{i_e}(\tilde{s}_k) = \rho_{i_e}$ for all $k$, where $\rho_{i_e}$ denotes the minimum value of the function $f_{i_e}(s)$. From the continuity of the function $f_{i_e}(s)$, we have $f_{i_e}(s_e) = \lim_{k\to+\infty} f_{i_e}(\tilde{s}_k) = \rho_{i_e}$. This implies that $s_e \in X_{i_e}$, which yields a contradiction. Therefore, all $X_i$ are closed sets.

Let $Y_i$ be a closed bounded convex set such that $X_i \subset Y_i$ and $\min_{y \notin Y_i} \|y - P_{X_i}(y)\| > 0$. Clearly, $\max_{y\in Y_i} \|y - P_{X_i}(y)\| < \varpi_0$ for some positive constant $\varpi_0 > 0$. From the property of a continuous function on a closed bounded set and Lemma 3, we have $\varpi_1 \triangleq \min_{y\in\bar{\partial}Y_i} \sum_{i=1}^n \nabla f_i(y)^T \frac{y-P_{X_i}(y)}{\|y-P_{X_i}(y)\|} > 0$ and $\varpi_2 \triangleq \max_{y\in\bar{\partial}Y_i} \sum_{i=1}^n \nabla f_i(y)^T \frac{y-P_{X_i}(y)}{\|y-P_{X_i}(y)\|} > 0$, where $\bar{\partial}Y_i$ denotes the boundary of the set $Y_i$. Integrating $\nabla f_i(s)$ along the line from $P_{Y_i}(y)$ to $y$, from Lemma 3, we have $f_i(y) - f_i(P_{X_i}(y)) = \int_{P_{X_i}(y)}^{y} \nabla f_i(s)^T \mathrm{d}s = \int_0^{\|y-P_{X_i}(y)\|} \nabla f_i(P_{X_i}(y) + \frac{y-P_{X_i}(y)}{\|y-P_{X_i}(y)\|} s)^T \frac{y-P_{X_i}(y)}{\|y-P_{X_i}(y)\|} \mathrm{d}s \ge \varpi_1 \|y - P_{Y_i}(y)\| - \varpi_2\varpi_0$. It follows that $\lim_{\|y\|\to+\infty} f_i(y) = +\infty$. Thus,

$$\lim_{\|y\|\to+\infty} \sum_{i=1}^n f_i(y) = +\infty. \tag{3}$$

On the other hand, since each $f_i(y)$ is lower bounded, $\sum_{i=1}^n f_i(y)$ is lower bounded and hence its infimum exists, denoted by $\omega_2$. From (3), for any sufficiently large constant $\omega_3 > \omega_2$, there exists a constant $h_l > 0$ such that $\sum_{i=1}^n f_i(y) > \omega_3$ for any $\|y\| > h_l$. Let $\tilde{Y} = \{\|y\| \le h_l\}$. It is clear that if $X$ is nonempty, $X \subset \tilde{Y}$. Note that $\tilde{Y}$ is a closed bounded set. Since $\sum_{i=1}^n f_i(y)$ is continuous with respect to $y$, from the property of a continuous function on a closed bounded set, we have the minimum set of $\sum_{i=1}^n f_i(y)$ in $\tilde{Y}$ is nonempty. That is, $X$ is nonempty. Then by using the same analysis approach as for $X_i$, it can be proved that $X$ is bounded and closed. ■

*Assumption 2:* The length of the time interval between any two contiguous switching times is no smaller than a given constant, denoted by $d_w$.

Arbitrary switching of the graph $\mathcal{G}(t)$ might lead to the Zeno behavior. Hence Assumption 2 is imposed to prevent the system from exhibiting the Zeno behavior. Throughout this paper, our analysis is founded on Assumption 2. For simplicity, this will not be repeatedly mentioned except when it is necessary.

### *A. Distributed Gradient Optimization*

In this subsection, we design $u_i(t)$ for (1) to solve the convex optimization problem (2). In particular, all agents are driven to reach a consensus in finite time while minimizing the team objective function as $t \to +\infty$. We propose the following algorithm

$$u_i(t) = \alpha \sum_{j\in N_i(t)} \mathrm{sgn}\big(x_j(t) - x_i(t)\big) - \gamma \nabla f_i(x_i(t)), \tag{4}$$

where $\alpha > 0$ and $\gamma > 0$ are two constants. In (4), the role of the first term, $\alpha\sum_{j\in N_i(t)} \mathrm{sgn}\big(x_j(t) - x_i(t)\big)$, is to drive all agents to reach a consensus, while the second term, $-\gamma\nabla f_i(x_i(t))$, is a weighted negative gradient of $f_i(x_i(t))$ playing a role in minimizing $f_i(x_i(t))$.

*Remark 1:* As our algorithms discussed in this paper contain sign functions that is piecewise differentiable, the solution of the system (1) would be considered in the sense of Filippov [24].

*Assumption 3:* Let $\nabla f_i(s) = \sigma s + \phi_i(s)$, where $\sigma \geq 0$ and $\|\phi_i(s)\| < g$ for a certain positive number $g$ and all $s \in \mathbb{R}^m$.

In [2], the subgradients of the local objective functions were assumed to be bounded and the most common quadratic convex functions were not considered. Under Assumption 3, when $\sigma = 0$, the gradient of each local objective function is bounded, and when $\sigma > 0$, the gradient of each local objective function contains a linear term and a bounded term, which includes the scenarios of [2] and the quadratic convex functions as special cases.

*Proposition 1:* Suppose that the graph $\mathcal{G}(t)$ is undirected and connected for all $t$ and Assumptions 2 and 3 hold. For system (1) with algorithm (4), if $\alpha/\gamma > 2ng$, all agents reach a consensus in finite time. That is, there exists a positive number $T$ such that $x_j(t) = x_i(t)$ for all $t > T$ and all $i, j \in \mathcal{I}$.

*Proof:* Consider the Lyapunov function candidate

$$V(t) = \frac{1}{2}\sum_{i=1}^{n} \left\|x_i(t) - \frac{1}{n}\sum_{j=1}^{n} x_j(t)\right\|^2. \tag{5}$$

It is clear that when $V(t) = 0$, $x_i(t) = x_j(t)$ for all $i, j$. Calculating $\dot{V}(t)$ along the solutions of system (1) with (4), we have

$$\begin{aligned}
\dot{V}(t) &= \sum\nolimits_{i=1}^{n}[x_i(t) - \frac{1}{n}\sum\nolimits_{k=1}^{n} x_k(t)]^T \\
&\quad \times \left[\alpha\sum\nolimits_{j\in N_i(t)}\operatorname{sgn}(x_j(t) - x_i(t)) \right.\\
&\quad \left. - \gamma\nabla f_i(x_i(t)) - \frac{1}{n}\sum\nolimits_{j=1}^{n}\dot{x}_j(t)\right] \\
&= \sum\nolimits_{i=1}^{n}\left[x_i(t) - \frac{1}{n}\sum\nolimits_{k=1}^{n} x_k(t)\right]^T \\
&\quad \times \left[\alpha\sum\nolimits_{j\in N_i(t)}\operatorname{sgn}(x_j(t) - x_i(t)) \right.\\
&\quad - [\gamma\nabla f_i(x_i(t)) - \gamma\sigma x_i(t) + \gamma\sigma x_i(t)] \\
&\quad \left. + \frac{\gamma\sigma}{n}\sum\nolimits_{j=1}^{n} x_j(t)\right]
\end{aligned} \tag{6}$$

where the second equality holds because $\sum_{i=1}^{n}[x_i(t) - \frac{1}{n}\sum_{k=1}^{n}x_k(t)]^T \frac{1}{n}\sum_{j=1}^{n}\dot{x}_j(t) = 0 \times \frac{1}{n}\sum_{j=1}^{n}\dot{x}_j(t) = 0$ and $\sum_{i=1}^{n}[x_i(t) - \frac{1}{n}\sum_{k=1}^{n}x_k(t)]^T \frac{1}{n}\sum_{j=1}^{n}x_j(t) = 0 \times \frac{1}{n}\sum_{j=1}^{n}x_j(t) = 0$. Since the graph $\mathcal{G}(t)$ is undirected, then $(i, j) \in \mathcal{E}$ if and only if $(j, i) \in \mathcal{E}$. Thus,

$$\begin{aligned}
&\sum\nolimits_{i=1}^{n}\left[x_i(t) - \frac{1}{n}\sum\nolimits_{k=1}^{n}x_k(t)\right]^T \\
&\quad \times \alpha\sum\nolimits_{j\in N_i(t)}\operatorname{sgn}(x_j(t) - x_i(t)) \\
&= \sum\nolimits_{i=1}^{n}\sum\nolimits_{j\in N_i(t)}\alpha\left[x_i(t) - \frac{1}{n}\sum\nolimits_{k=1}^{n}x_k(t)\right]^T \\
&\quad \times \operatorname{sgn}(x_j(t) - x_i(t)) \\
&= \sum\nolimits_{i=1}^{n}\sum\nolimits_{j\in N_i(t)}\left\{\frac{\alpha}{2}\right. \\
&\quad \times \left[x_i(t) - \frac{1}{n}\sum\nolimits_{k=1}^{n}x_k(t)\right]^T \operatorname{sgn}(x_j(t) - x_i(t)) \\
&\quad \left. + \frac{\alpha}{2}\left[x_j(t) - \frac{1}{n}\sum\nolimits_{k=1}^{n}x_k(t)\right]^T \operatorname{sgn}(x_i(t) - x_j(t))\right\} \\
&= \sum\nolimits_{i=1}^{n}\sum\nolimits_{j\in N_i(t)}\left\{\frac{\alpha}{2}\left[x_i(t) - \frac{1}{n}\sum\nolimits_{k=1}^{n}x_k(t)\right.\right. \\
&\quad \left.\left. - x_j(t) + \frac{1}{n}\sum\nolimits_{k=1}^{n}x_k(t)\right]^T \operatorname{sgn}(x_j(t) - x_i(t))\right. \\
&= \sum\nolimits_{i=1}^{n}\sum\nolimits_{j\in N_i(t)}\frac{\alpha}{2}[x_i(t) - x_j(t)]^T \\
&\quad \times \operatorname{sgn}(x_j(t) - x_i(t)).
\end{aligned} \tag{7}$$

Under Assumption 3, $\nabla f_i(x_i(t)) = \sigma x_i(t) + \phi_i(x_i(t)), \sigma > 0$ and $\|\phi_i(x_i(t))\| < g$. Since $\gamma > 0$, it follows from (6) and (7) that

$$\begin{aligned}
&\dot{V}(t) \\
&\quad = \sum\nolimits_{i=1}^{n}\sum\nolimits_{j\in N_i(t)}\frac{\alpha}{2}[x_i(t) - x_j(t)]^T \\
&\qquad \times \operatorname{sgn}(x_j(t) - x_i(t)) \\
&\qquad - \sum\nolimits_{i=1}^{n}\left[x_i(t) - \frac{1}{n}\sum\nolimits_{j=1}^{n}x_j(t)\right]^T \gamma\phi_i(x_i(t)) \\
&\qquad - \gamma\sigma\sum\nolimits_{i=1}^{n}\left\|x_i(t) - \frac{1}{n}\sum\nolimits_{j=1}^{n}x_j(t)\right\|^2 \\
&\quad \leq \sum\nolimits_{i=1}^{n}\sum\nolimits_{j\in N_i(t)}\frac{\alpha}{2}[x_i(t) - x_j(t)]^T \\
&\qquad \times \operatorname{sgn}(x_j(t) - x_i(t)) \\
&\qquad + \sum\nolimits_{i=1}^{n}\|x_i(t) - \frac{1}{n}\sum\nolimits_{j=1}^{n}x_j(t)\|\gamma g.
\end{aligned} \tag{8}$$

Consider the quantity $x_i(t) - \frac{1}{n}\sum_{j=1}^{n}x_j(t)$ for all $i \in \mathcal{I}$. Let $i_0, j_0 \in \mathcal{I}$ be the integers such that $\|x_{i_0}(t) - x_{j_0}(t)\| = \max_{i,j\in\mathcal{I}}\|x_i(t) - x_j(t)\|$ at time $t$. It is clear that $\|x_i(t) - \frac{1}{n}\sum_{j=1}^{n}x_j(t)\| \leq \frac{1}{n}\sum_{j=1}^{n}\|x_i(t) - x_j(t)\| \leq \|x_{i_0}(t) - x_{j_0}(t)\|$. Since $\mathcal{G}(t)$ is connected, there must exist a path $(i_0, i^0), (i^0, i^1), \ldots, (i^{h-1}, i^h), (i^h, j_0)$ that connects nodes $i_0$ and $j_0$. Note that $\|x_{i_0}(t) - x_{j_0}(t)\| \leq \|x_{i_0}(t) - x_{i^0}(t)\| + \sum_{k=1}^{h}\|x_{i^{k-1}}(t) - x_{i^k}(t)\| + \|x_{i^h}(t) - x_{j_0}(t)\| \leq \sum_{i=1}^{n}\sum_{j\in N_i(t)}\|x_i(t) - x_j(t)\|$. Therefore, $\|x_{i_0}(t) - x_{j_0}(t)\| \leq \sum_{i=1}^{n}\sum_{j\in N_i(t)}\|x_i(t) - x_j(t)\|$. Also, note that $\|x_i(t) - x_j(t)\| \leq -[x_i(t) - x_j(t)]^T\operatorname{sgn}(x_j(t) - x_i(t))$ for all $i, j \in \mathcal{I}$ from the relations of operator norms. It follows that

$$\begin{aligned}
&\left\|x_i(t) - \frac{1}{n}\sum\nolimits_{j=1}^{n}x_j(t)\right\| \\
&\quad \leq \|x_{i_0}(t) - x_{j_0}(t)\| \\
&\quad \leq -\sum\nolimits_{i=1}^{n}\sum\nolimits_{j\in N_i(t)}[x_i(t) - x_j(t)]^T \\
&\qquad \times \operatorname{sgn}(x_j(t) - x_i(t)).
\end{aligned} \tag{9}$$

If $\alpha/\gamma > 2ng$, it follows from (8) and (9) that

$$\dot{V}(t) \leq \sum_{i=1}^{n}\sum_{j\in N_i(t)}[x_i(t)-x_j(t)]^T \times \mathrm{sgn}(x_j(t)-x_i(t))\left(\frac{\alpha}{2}-\gamma ng\right) \leq 0.$$

This implies that $V(t)$ and hence $\|x_{i_0}(t)-x_{j_0}(t)\|$ are bounded. When $V(t)\neq 0$, i.e., $\|x_{i_0}(t)-x_{j_0}(t)\|\neq 0$, noting from (9) that $\sqrt{V(t)} \leq \sqrt{\frac{n}{2}}\|x_{i_0}(t)-x_{j_0}(t)\|$, it follows that

$$\frac{\dot{V}(t)}{\sqrt{V(t)}} \leq \frac{\sum_{i=1}^{n}\sum_{j\in N_i(t)}[x_i(t)-x_j(t)]^T\mathrm{sgn}(x_j(t)-x_i(t))}{\frac{1}{\sqrt{2}}\sqrt{n}\|x_{i_0}(t)-x_{j_0}(t)\|} \times\left(\frac{\alpha}{2}-\gamma ng\right) \leq -(\alpha/\sqrt{2n}-\sqrt{2n}\gamma g) < 0.$$

Integrating both sides of this inequality, we have

$$2\sqrt{V(t)}-2\sqrt{V(0)} < -(\alpha/\sqrt{2n}-\sqrt{2n}\gamma g)t. \qquad (10)$$

It is clear that $V(t)$ converges to zero in finite time. Namely, all agents reach a consensus in finite time. ■

*Theorem 1:* Suppose that the graph $\mathcal{G}(t)$ is undirected and connected for all $t$ and Assumptions 1, 2 and 3 hold. For system (1) with algorithm (4), if $\alpha/\gamma > 2ng$, all agents reach a consensus in finite time and minimize the team objective function (2) as $t\to+\infty$.

*Proof:* Define

$$x^*(t) \triangleq \frac{1}{n}\sum_{j=1}^{n}x_j(t). \qquad (11)$$

Under Assumption 3, Proposition 1 holds. From Proposition 1, there exists a positive number $T$ such that $x_i(t)=x^*(t)$ for all $t>T$ and all $i\in\mathcal{I}$. Since the graph $\mathcal{G}(t)$ is undirected, it follows that for all $t>T$,

$$\begin{aligned}\dot{x}^*(t) &= \frac{1}{n}\sum_{i=1}^{n}\dot{x}_i(t)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\alpha\sum_{j\in N_i(t)}\mathrm{sgn}(x_j(t)-x_i(t)) - \gamma\nabla f_i(x_i(t))\right] = -\frac{\gamma}{n}\sum_{i=1}^{n}\nabla f_i(x^*(t)). \qquad (12)\end{aligned}$$

Consider the Lyapunov function candidate $V(t)=\frac{1}{2}\|x^*(t)-P_X(x^*(t))\|^2$ for $t>T$. Calculating $\dot{V}(t)$ along the solutions of (12), it follows from Lemma 2 and the convexity of $\sum_{i=1}^{n}f_i(s)$ that

$$\begin{aligned}\dot{V}(t) &= [x^*(t)-P_X(x^*(t))]^T\dot{x}^*(t)\\ &= -\frac{\gamma}{n}[x^*(t)-P_X(x^*(t))]^T\sum_{i=1}^{n}\nabla f_i(x^*(t))\\ &\leq -\frac{\gamma}{n}\left[\sum_{i=1}^{n}f_i(x^*(t))-\sum_{i=1}^{n}f_i(P_X(x^*(t)))\right] \qquad (13)\end{aligned}$$

for $t>T$. Let $Y=\{s\in\mathbb{R}^m \mid \|s-P_X(s)\|\leq l_1\}$ for some constant $l_1>0$ and $\rho=\min_{s\in\bar{\partial}Y}\sum_{i=1}^{n}[f_i(s)-f_i(P_X(s))]$, where $\bar{\partial}Y$ denotes the boundary of $Y$. Since $P_X(s)\in X$, from the definition of $X$, we have $\rho>0$. Moreover, from Lemma 3, $\sum_{i=1}^{n}[f_i(s)-f_i(P_X(s))]>\rho$ for any $s\notin Y$. Thus, $\dot{V}(t)\leq -\frac{\gamma}{n}\rho$ for any $x^*(t)\notin Y$ and $t>T$. This implies that there exists a constant $T_0>T$ for any $l_1>0$ such that $\|x^*(t)-P_X(x^*(t))\|\leq l_1$ for all $t>T_0$. In view of the arbitrariness of $l_1$, letting $l_1\to 0$, we have $\lim_{t\to+\infty}\|x^*(t)-P_X(x^*(t))\|=0$. It follows from the definition of $X$ that the team objective function (2) is minimized as $t\to+\infty$. ■

## B. Distributed Adaptive Gradient Optimization Algorithm

In algorithm (4), it is required that the gains $\alpha$ and $\gamma$ should be known to all agents and it can only be used to deal with quadratic-like convex objective functions. In this subsection, we design a distributed adaptive algorithm for (1) to solve the optimization problem (2) for general convex local objective functions. The algorithm is given by

$$\begin{aligned}u_i(t) &= \sum_{j\in N_i(t)}q_{ij}(t)\mathrm{sgn}\big(x_j(t)-x_i(t)\big) - \nabla f_i(x_i(t)),\\ \dot{q}_{ij}(t) &= \begin{cases}\mathrm{sgn}(\max_{s\in[t-c_0,t]}\|x_j(s)-x_i(s)\|), & \text{if } (i,j)\in\mathcal{G}(t),\\ 0, & \text{otherwise,}\end{cases}\\ q_{ij}(0) &= q_{ji}(0)=0, \qquad (14)\end{aligned}$$

where $c_0>0$ is an arbitrary constant. In (14), the role of the first term, $\sum_{j\in N_i(t)}q_{ij}(t)\mathrm{sgn}\big(x_j(t)-x_i(t)\big)$, is to drive all agents to reach a consensus, while the second term, $-\nabla f_i(x_i(t))$, is the negative gradient of $f_i(x_i(t))$ playing a role in minimizing $f_i(x_i(t))$.

*Theorem 2:* Suppose that the graph $\mathcal{G}(t)$ is undirected and connected for all $t$ and Assumptions 1 and 2 hold. For system (1) with algorithm (14), all agents reach a consensus in finite time and minimize the team objective function (2) as $t\to+\infty$.

*Proof:* We first show that all $x_i(t)$ remain in a bounded region. Under Assumption 1, it follows from Lemma 4 that all $X_i$ and $X$ are nonempty closed bounded convex sets for all $i$. Therefore, there is a closed bounded convex set $Y$ such that $x_i(0)\in Y$, $X\subset Y$ and $X_i\subset Y$ for all $i$. Consider the Lyapunov function candidate $V(t)=\sum_{i=1}^{n}\|x_i(t)-z\|^2$ for some $z\in X$. Let $Y$ be sufficiently large for all $z_j\in X_j$ such that $f_i(x_i(t))-f_i(z)\geq\sum_{j=1,j\neq i}^{n}[f_j(z)-f_j(z_j)]$ for all $i$ and all $x_i(t)\notin Y$. Calculating $\dot{V}(t)$, we have

$$\begin{aligned}\dot{V}(t) &= \sum_{i=1}^{n}\big(x_i(t)-z\big)^T \times\sum_{j\in N_i(t)}q_{ij}(t)\mathrm{sgn}\big(x_j(t)-x_i(t)\big)\\ &\quad -\sum_{i=1}^{n}\big(x_i(t)-z\big)^T\nabla f_i(x_i(t)). \qquad (15)\end{aligned}$$

Since $z\in X$, from the convexity of the function $f_i(x_i(t))$, we have $\nabla f_i(x_i(t))^T(z-x_i(t))\leq f_i(z)-f_i(x_i(t))$. Moreover, since the graph $\mathcal{G}(t)$ is undirected, similar to the proof of Proposition 1, we have

$$\begin{aligned}&\sum_{i=1}^{n}x_i(t)^T\sum_{j\in N_i(t)}q_{ij}(t)\mathrm{sgn}\big(x_j(t)-x_i(t)\big)\\ &= \sum_{i=1}^{n}\sum_{j\in N_i(t)}\frac{q_{ij}(t)}{2}\big(x_i(t)-x_j(t)\big)^T \times\mathrm{sgn}\big(x_j(t)-x_i(t)\big)\leq 0 \qquad (16)\end{aligned}$$

and

$$\sum_{i=1}^{n} z^T \sum_{j\in N_i(t)} q_{ij}(t)\mathrm{sgn}\big(x_j(t)-x_i(t)\big)=0. \quad (17)$$

From (15), (16) and (17), we have $\dot V(t) \le -\sum_{i=1}^n [f_i(x_i(t)) - f_i(z)]$. If $x_{i_0}(t)\notin Y$ for some $i_0$, we have $f_{i_0}(x_{i_0}(t)) - f_{i_0}(z) \ge \sum_{j=1,j\ne i_0}^n [f_j(z)-f_j(z_j)]$ for all $z_j\in X_j$ and hence $\dot V(t) \le -[f_{i_0}(x_{i_0}(t)) - f_{i_0}(z)] + \sum_{j=1,j\ne i_0}^n [f_j(z) - f_j(z_j)] \le 0$. This implies that all $x_i(t)$ remain in $Y$. Note that each function $f_i(s)$ is differentiable and $Y$ is bounded. Thus each $\nabla f_i(s)$ is bounded. That is, $\|\nabla f_i(s)\| < \rho_f$ for some constant $\rho_f > 0$.

Now, we show that all agents reach a consensus in finite time. Let $0 < t_{k1} \le t_{k2} < t_{k+1,1} \le t_{k+1,2}$ denote the contiguous switching times for all $k\in\{1,2,\ldots\}$ such that $x_i(t)\ne x_j(t)$ for some two integers $i,j\in\mathcal{I}$ and all $t\in[t_{k1},t_{k2})$ and $x_i(t) = x_j(t)$ for all $i,j\in\mathcal{I}$ and all $t\in[t_{k2},t_{k+1,1})$. Suppose that consensus is not reached in finite time and $\sum_{k=1}^{+\infty}(t_{k2}-t_{k1}) < +\infty$. It is clear that $t_{k2}-t_{k1}>0$ when $k\to+\infty$. From the dynamics of $q_{ij}(t)$, we have that $\lim_{t\to+\infty}\sum_{i=1}^n\sum_{j=1}^n q_{ij}(t) = +\infty$. Then by a similar approach to the following case of $\sum_{k=1}^{+\infty}(t_{k2}-t_{k1}) = +\infty$, it can be shown that consensus can be reached in finite time, which is a contradiction.

Suppose that $\sum_{k=1}^{+\infty}(t_{k2}-t_{k1}) = +\infty$. Then from the dynamics of $q_{ij}(t)$, there must exist a pair of agents, denoted by $i_0\ne j_0$, such that $\lim_{t\to+\infty} q_{i_0j_0}(t) = +\infty$. In the following, we prove that there exist a pair of agents, denoted by $i_1\ne j_1$, such that $(i_1,j_1)\notin\{(i_0,j_0),(j_0,i_0)\}$, $i_1\in\{i_0,j_0\}$ and $\lim_{t\to+\infty} q_{i_1j_1} = +\infty$. If this is not true, we have $q_{ii_0}(t) < \gamma_q$ and $q_{ij_0}(t) < \gamma_q$ for some constant $\gamma_q > \rho_f$, all $t$ and all $i\in\cup_{s\in[0,+\infty)}[N_{i_0}(s)\cup N_{j_0}(s)]$ with $i\ne i_0$ and $i\ne j_0$. Since $\lim_{t\to+\infty} q_{i_0j_0}(t) = +\infty$, there exists a sufficiently large constant $T_0 > 0$ for any $\gamma_0 > 6nm\gamma_q$ such that $q_{i_0j_0}(t) > \gamma_0$ for all $t > T_0$. By simple calculations based on (14), when $(i_0,j_0)\in\mathcal{G}(t)$ and $\|x_{i_0}(t)-x_{j_0}(t)\|\ne 0$ for $t > T_0$, we have $\frac{\mathrm{d}}{\mathrm{d}t}\|x_{i_0}(t)-x_{j_0}(t)\| \le \frac{x_{i_0}(t)-x_{j_0}(t)}{\|x_{i_0}(t)-x_{j_0}(t)\|} 2q_{i_0j_0}(t)\mathrm{sgn}\big(x_{j_0}(t) - x_{i_0}(t)\big) + 2nm\gamma_q \le -2nm\gamma_q$. When there exist at least an agent $i$ such that $i\in N_{\tilde i}(t)$ and $\|x_{\tilde i}(t) - x_i(t)\|\ne 0$ for $\tilde i\in\{i_0,j_0\}$ and either $(i_0,j_0)\notin\mathcal{G}(t)$ or $\|x_{i_0}(t)-x_{j_0}(t)\| = 0$ holds, we have $\frac{\mathrm{d}}{\mathrm{d}t}\|x_{i_0}(t)-x_{j_0}(t)\| \le 2nm\gamma_q$ for $t > T_0$. Since all $x_i(t)$ remain in a bounded region, $\|x_{i_0}(t)-x_{j_0}(t)\| < \rho_v$ for some positive constant $\rho_v$. Let $\tau_a(T_1)$ and $\tau_b(T_1)$, respectively, denote the total time in the interval $(T_0,T_1)$ for any $T_1 > T_0$ for the case when $(i_0,j_0)\in\mathcal{G}(t)$ and $\|x_{i_0}(t)-x_{j_0}(t)\|\ne 0$ and the case when there exist at least an agent $i$ such that $i\in N_{\tilde i}(t)$ and $\|x_{\tilde i}(t)-x_i(t)\|\ne 0$ for $\tilde i\in\{i_0,j_0\}$ and either $(i_0,j_0)\notin\mathcal{G}(t)$ or $\|x_{i_0}(t)-x_{j_0}(t)\| = 0$ holds. Thus,

$$\begin{aligned} 0 \le\ & \|x_{i_0}(T_1) - x_{j_0}(T_1)\| \\ \le\ & \|x_{i_0}(T_0) - x_{j_0}(T_0)\| + 2nm\gamma_q\tau_b(T_1) \\ & - 2nm\gamma_q\tau_a(T_1) \\ \le\ & \rho_v + 2nm\gamma_q\tau_b(T_1) - 2nm\gamma_q\tau_a(T_1). \end{aligned} \quad (18)$$

Since $\lim_{t\to+\infty} q_{i_0j_0}(t) = +\infty$, from the dynamics of $q_{ij}(t)$, we have $\lim_{T_1\to+\infty}\tau_a(T_1) = +\infty$ and hence from (18) we have $\lim_{T_1\to+\infty}\tau_b(T_1) = +\infty$. That is, there exist a pair of agents $i_1\ne j_1$ such that $(i_1,j_1)\notin\{(i_0,j_0),(j_0,i_0)\}$, $i_1\in\{i_0,j_0\}$ and $\lim_{t\to+\infty} q_{i_1j_1}(t) = +\infty$. Similarly, it can be proved that there exist a pair of agents $i_2\ne j_2$ such that $(i_2,j_2)\notin\{(i_0,j_0),(j_0,i_0),(i_1,j_1),(j_1,i_1)\}$, $i_2\in\{i_0,j_0,i_1\}$ and $\lim_{t\to+\infty} q_{i_2j_2}(t) = +\infty$. By analogy, it can be proved that $\lim_{t\to+\infty} q_{ij}(t) = +\infty$ for all $i,j$. Since each $\|\nabla f_i(x_i(t))\|$ is bounded for all $t$, there is a constant $T_2 > 0$ such that $q_{ij}(t) > 2n\max_k \|\nabla f_k(x_k(t))\|$ for all $i,j$ and all $t > T_2$. Similar to the proof of Proposition 1, we have all agents reach a consensus in finite time. This contradicts with the precondition that $\sum_{k=1}^{+\infty}(t_{k2}-t_{k1}) = +\infty$.

Summarizing the above analysis, all agents reach a consensus in finite time. Then there exists a number $T > 0$ such that $x_i(t) = x^*(t)$, where $x^*(t)$ is defined in (11), for all $t > T$ and all $i\in\mathcal{I}$. Similar to the proof of Theorem 1, it can be proved that the team objective function (2) is minimized as $t\to+\infty$. ■

### C. Distributed Optimization Algorithm With Nonuniform Gradient Gains

In the existing works, the gradient gains are usually assumed to be uniform and need to be known in advance, e.g., [1]. In this subsection, we extend to consider the nonuniform gradient gains based on the agents' states for general convex local objective functions. The algorithm is given by

$$\begin{aligned} \dot q_i(t) =\ & \arctan(e^{\|x_i(t)\|}),\ q_i(0) > 0, \\ u_i(t) =\ & \sum_{j\in N_i(t)} \mathrm{sgn}\big(x_j(t)-x_i(t)\big) - \frac{\nabla f_i(x_i(t))}{\sqrt{q_i(t)}} \end{aligned} \quad (19)$$

for all $i$. Here, the gain $1/\sqrt{q_i(t)}$ is used to ensure the weighted gradient $\frac{\mathrm{d}f_i(x_i(t))}{\sqrt{q_i(t)}\mathrm{d}x_i(t)}$ to tend to zero as time evolves. In practical applications, it is hard for all agents to have a uniform system clock and know its value accurately at any time. So, we do not use the information of the system clock directly in the design of the gradient gains.

*Remark 2:* Here, we use the inverse tangent functions and the exponential functions to guarantee $\dot q_i(t)$ to be upper and lower bounded by two positive constants (here the two constants are $\frac{\pi}{2}$ and $\frac{\pi}{4}$). As a matter of fact, there are some other functions, e.g., saturation function, that can be used to play the same role. For easy readability, we do not give the general form of such functions.

*Theorem 3:* Suppose that the graph $\mathcal{G}(t)$ is undirected and connected for all $t$ and Assumptions 1 and 2 hold. For system (1) with algorithm (19), all agents reach a consensus in finite time and minimize the team objective function (2) as $t\to+\infty$.

*Proof:* Note that $\pi/4 \le \arctan(e^{\|x_i(s)\|}) \le \pi/2$ for all $s$ and all $i$. There exists a constant $T > 0$ such that $2\sqrt{t} > \sqrt{q_i(t)} > \frac{\sqrt{t}}{2}$ for all $i$ and all $t > T$. Consider the Lyapunov function candidate $V(t) = \sum_{i=1}^n \|x_i(t) - z\|^2$ for $z\in X$ and $t > T$. Under Assumption 1, it follows from Lemma 4 that all $X_i$ and $X$ are nonempty closed bounded convex sets for all $i$. Let $Y$ be a closed bounded convex set such that $x_i(T)\in Y$, $X\subset Y$, $X_i\subset Y$ and $f_i(x_i(t)) - f_i(z) \ge 4\sum_{j=1,j\ne i}^n [f_j(z) - f_j(z_j)]$ for all $i$, all $z_j \in X_j$ and all $x_i(t) \notin Y$. It follows that $\frac{1}{\sqrt{q_i(t)}}[f_i(x_i(t)) - f_i(z)] \geq \sum_{j=1,j\neq i}^{n} \frac{1}{\sqrt{q_j(t)}}[f_j(z) - f_j(z_j)]$ for all $t > T$, all $i$, all $z_j \in X_j$ and all $x_i(t) \notin Y$. Then similar to the proof of Theorem 2, it can be proved that all $\|x_i(t)\|$ and all $\nabla f_i(x_i(t))$ are bounded for all $t > T$. That is, $|f_i(x_i(t))| < \mu_c$ and $\|\nabla f_i(x_i(t))\| < \mu_c$ for some constant $\mu_c > 0$, all $i$ and all $t > T$. Moreover, note that $X$ is bounded and each $f_i(s)$ is differentiable for all $i$ and all $s$. Let $\mu_c$ be sufficiently large such that $\mu_c > 2n\|\nabla f_i(x_i(t))\|$ and $\mu_c > |f_i(s)|$ for all $i$, all $t > T$ and all $s \in X$. Let $T_0 > T$ be a constant such that $\frac{\sqrt{T_0}}{2} > \mu_c$. Similar to the proof of Proposition 1, it can be proved that all agents reach a consensus in finite time. That is, there exists a constant $T_1 > T_0$ such that $x_i(t) = x^*(t)$ for all $i$ and all $t > T_1$, where $x^*(t)$ is defined in (11).

Now, we prove that the team objective function (2) is minimized as $t \to +\infty$. Let $E = \{s \in \mathbb{R}^m \mid \|s - P_X(s)\| \leq l_1\}$ for some constant $l_1 > 0$ and $\rho = \min_{s\in\bar{\partial}E} \sum_{i=1}^{n}[f_i(s) - f_i(P_X(s))]$, where $\bar{\partial}E$ denotes the boundary of $E$. Since $P_X(s) \in X$, from the definition of $X$, we have $\rho > 0$. From Lemma 3, $\sum_{i=1}^{n}[f_i(s) - f_i(P_X(s))] > \rho$ for any $s \notin E$. Note that $q_i(t) - q_i(T_1) = q_j(t) - q_j(T_1) = \int_{T_1}^{t} \arctan(e^{\|x^*(s)\|})\mathrm{d}s \triangleq q^*(t)$ and $q_i(t) = q^*(t)/\Delta_i(t)$ for all $i, j$ and $t > T_1$, where $\Delta_i(t) = 1/(1 + \frac{q_i(T_1)}{q^*(t)})$. Since the graph $\mathcal{G}(t)$ is undirected and connected, it follows that for all $t > T_1$,

$$\begin{aligned}
&\dot{x}^*(t)\\
&= \frac{1}{n}\sum_{i=1}^{n} \dot{x}_i(t)\\
&= \frac{1}{n}\sum_{i=1}^{n} \left[\sum_{j\in N_i(t)} \mathrm{sgn}(x_j(t) - x_i(t)) - \frac{\nabla f_i(x_i(t))}{\sqrt{q_i(t)}}\right]\\
&= -\frac{1}{n}\sum_{i=1}^{n} \frac{\sqrt{\Delta_i(t)}\nabla f_i(x^*(t))}{\sqrt{q^*(t)}}.
\end{aligned} \tag{20}$$

On the other hand, recall that $\pi/4 \leq \arctan(e^{\|x_i(s)\|}) \leq \pi/2$ for all $s$ and all $i$. From the definition of $q^*(t)$, there exists a constant $T_2 > T_1$ for any $0 < \epsilon < \frac{\rho}{4n}$ such that $2\sqrt{t} > \sqrt{q^*(t)} > \frac{\sqrt{t}}{2}$ and $1 - \sqrt{\Delta_i(t)} < \frac{\epsilon}{2\mu_c}$ for all $i$ and all $t > T_2$. Let $\phi_i(t) = f_i(x^*(t)) - f_i(P_X(x^*(t)))$ for all $i$. Since $|f_i(x^*(t))| < \mu_c$ and $|f_i(P_X(x^*(t)))| < \mu_c$, it follows that $|\phi_i(t)[1 - \sqrt{\Delta_i(t)}]| < \epsilon$ for all $i$.

Consider the Lyapunov function candidate $\bar{V}(t) = \frac{1}{2}\|x^*(t) - P_X(x^*(t))\|^2$ for $t > T_2$. Calculating $\dot{\bar{V}}(t)$ along the solutions of (20), it follows from Lemma 2 and the convexity of $\sum_{i=1}^{n} f_i(s)$ that

$$\begin{aligned}
\dot{\bar{V}}(t) &= [x^*(t) - P_X(x^*(t))]^T \dot{x}^*(t)\\
&= -\frac{1}{n\sqrt{q^*(t)}}[x^*(t) - P_X(x^*(t))]^T\\
&\quad \times \sum_{i=1}^{n} \nabla f_i(x^*(t))\big(1 - 1 + \sqrt{\Delta_i(t)}\big)\\
&\leq -\frac{1}{n\sqrt{q^*(t)}}\sum_{i=1}^{n} \phi_i(t) - \sum_{i=1}^{n} \phi_i(t)\frac{1 - \sqrt{\Delta_i(t)}}{n\sqrt{q^*(t)}}\\
&\leq -\frac{1}{n\sqrt{q^*(t)}}\sum_{i=1}^{n} \phi_i(t) + \frac{\epsilon}{\sqrt{q^*(t)}}
\end{aligned} \tag{21}$$

for $t > T_2$. Since $2\sqrt{t} > \sqrt{q^*(t)} > \frac{\sqrt{t}}{2}$ for all $t > T_2$ and $\epsilon < \frac{\rho}{4n}$, then $\dot{\bar{V}}(t) \leq -\frac{1}{\sqrt{q^*(t)}}[\frac{\rho}{n} - \epsilon] \leq -\frac{1}{\sqrt{t}}[\frac{\rho}{2n} - 2\epsilon] < 0$ for any $x^*(t) \notin E$ and all $t > T_2$. Integrating both sides of this inequality from $T_2$ to $t$, we have $\dot{\bar{V}}(t) \leq -(\sqrt{t} - \sqrt{T_2})[\frac{\rho}{n} - 4\epsilon]$. This implies that there exists a constant $T_3 > T_2$ for any $l_1 > 0$ such that $\|x^*(t) - P_X(x^*(t))\| \leq l_1$ for all $t > T_3$. In view of the arbitrariness of $l_1$, letting $l_1 \to 0$, we have $\lim_{t\to+\infty} \|x^*(t) - P_X(x^*(t))\| = 0$. That is, the team objective function (2) is minimized as $t \to +\infty$. ■

*Remark 3:* In the existing works, the gradient or subgradient gains are usually assumed to be uniform for all agents at any time instant, and moreover their values for all time instants need be known in advance. For example, in [2], for discrete-time multi-agent systems, the gains should satisfy that $\sum_{i=1}^{n}\alpha_k = +\infty$ and $\sum_{i=1}^{n}\alpha_k^2 = +\infty$, where $\alpha_k$ denotes the uniform subgradient gain for all agents at time instant $k$. One example of the selection is $\alpha_k = \frac{1}{1+k}$, $k = 0, 1, \ldots$. To determine the gains, the exact time clock (i.e., $k$ in the discrete-time case) should be known by all agents and the values of the gains for all agents need be kept identical at any time instant. In contrast, in this paper, the gradient gains in algorithm (19) are state-dependent and can be self-adjusted based on the agents' current states. At each time instant, the agents only need to know their current states $x_i(t)$ to determine their own gains and there is no need to know the current time clock (i.e., $t$ in the continuous-time case). Note that while $x_i(t)$ is a function of time, the agents do not use the current time to calculate $x_i(t)$ but instead the states are obtained by measurements without the need for explicitly knowing the exact clock. The gradient gains can be different for different agents, which might distinctly relax the synchronization requirement on the system. In [5], a distributed algorithm with nonuniform subgradient gains was also given to solve the distributed optimization problem, but the algorithm can only be used in the stochastic sense. By taking the mathematical expectation, it uses uniform subgradient gains for all agents in nature.

## D. Distributed Finite-Time Optimization Algorithm

Most of the existing works on the distributed optimization problem (e.g., [1]–[16]) as well as the algorithms introduced in Section III.A–III.C, studied only the asymptotic stability of the algorithm, and are rarely concerned with the finite-time convergence of the algorithms. To this end, in this subsection, we design one algorithm for (1) such that distributed optimization can not only be achieved, but achieved in finite time.

The finite-time algorithm for system (1) is given by

$$\begin{aligned}
\dot{\psi}_i(t) &= \sum_{j\in N_i(t)} p_{ij}(t)\mathrm{sgn}(\theta_j(t) - \theta_i(t)),\\
\theta_i(t) &= \psi_i(t) + \nabla f_i(x_i(t)), \psi_i(0) = 0,
\end{aligned}$$

$$\begin{aligned}
\dot{p}_{ij}(t) =& \begin{cases}\text{sgn}(\max_{s\in[t-c_0,t]}\|\theta_j(s)-\theta_i(s)\|), \text{ if } (i,j)\in\mathcal{G}(t),\\ 0, \text{ otherwise},\end{cases}\\
p_{ij}(0) =& \; p_{ji}(0) = 0,\\
u_i(t) =& \sum\nolimits_{j\in N_i(t)} q_{ij}(t)\text{sgn}(x_j(t)-x_i(t))\\
& -\frac{\theta_i(t)}{\|\theta_i(t)\|} - gc_i(t),\\
\dot{q}_{ij}(t) =& \begin{cases}\text{sgn}(\max_{s\in[t-c_0,t]}\|x_j(s)-x_i(s)\|), \text{ if } (i,j)\in\mathcal{G}(t),\\ 0, \text{ otherwise},\end{cases}\\
q_{ij}(0) =& \; q_{ji}(0) = 0,\\
gc_i(t) =& \begin{cases} 0, \text{ if } \theta_j(t)=\theta_i(t)\\ \text{and } x_i(t)=x_j(t) \text{ for all } j\in N_i(t),\\ x_i(t)-P_{X_i}(x_i(t)), \text{ otherwise},\end{cases}
\end{aligned} \tag{22}$$

where $c_0 > 0$ is an arbitrary constant, and $\theta_i(t)$ and $\psi_i(t)$ are the internal states of the dynamic averaging estimator for all $i$. Here the dynamic averaging estimator is motivated by [21]. Here, to eliminate the singular point of the function $\frac{x}{\|x\|}$, we define $\frac{x}{\|x\|} = 0$ when $x = 0$.

In algorithm (22), the role of $\theta_i(t)$ is to estimate the average derivative of all local objective functions $f_i(x_i(t))$ with respect to $x_i(t)$, the role of the time-varying gains $p_{ij}(t)$ is to ensure the influence of $\nabla^2 f_i(x^*(t))\dot{x}^*(t)$ on the tracking of the average derivative of all $f_i(x_i(t))$ to vanish to zero as time evolves, and the role of the time-varying gains $q_{ij}(t)$ is to force all agents to reach a consensus and move along the negative direction of the average derivative of all local objective functions $f_i(x_i(t))$.

*Remark 4:* There are three difficulties in the analysis of system (1) with algorithm (22): (a) this system is a time-varying system with a strong nonlinearity since the interaction gains $p_{ij}(t)$ and $q_{ij}(t)$ are time-varying and this system contains a strongly nonlinear term $\nabla^2 f_i(x^*(t))\dot{x}^*(t)$ as shown later in (31); (b) there exist four strong couplings: the first one is between the variables $\theta_i(t)$ and $x_i(t)$ in each agent; the second one is among the variables $\theta_i(t)$ for neighbor agents; the third one is among the variables $x_i(t)$ for neighbor agents; and the last one is between the variables $\theta_i(t)$ and $x_j(t)$ for neighbor agents; (c) each $\nabla f_i(x_i(t))$ and each $\nabla^2 f_i(x^*(t))$ are not bounded and hence $\theta_i(t)$ might tend to infinity as time evolves, which might destroy the system stability.

*Assumption 4:* Suppose that each $[\nabla^2 f_i(s)]_{jk} = \frac{\partial^2 f(s)}{\partial s_j \partial s_k}$ is continuous with respect to $s$, and either one of the following conditions holds:

(a) There exists a scalar $\delta > 0$ and a vector $\bar{s} \in X$ such that $\{\xi \,|\, \|\xi - \bar{s}\| \le \delta\} \subset X$ [1].

(b) There is a neighborhood of $X$, denoted by $S$, and a uniform constant $0 < c_s \le 1$ such that $(s - P_X(s))^T \frac{1}{n}\sum_{i=1}^n \nabla f_i(s) \ge c_s \left\|\frac{1}{n}\sum_{i=1}^n \nabla f_i(s)\right\| \|s - P_X(s)\|$ for all $s \in S$.

Below are some lemmas that will be used in the proof of the main theorem.

*Lemma 5:* Let $Z$ be a closed bounded convex set containing $X$, and $\bar{s}$ be defined in Assumption 4(a). Under Assumption 4(a), there exists a uniform constant $0 < c_x \le 1$ such that $(s - \bar{s})^T \frac{1}{n}\sum_{i=1}^n \nabla f_i(s) \ge c_x \left\|\frac{1}{n}\sum_{i=1}^n \nabla f_i(s)\right\| \|s - \bar{s}\|$ for any $s \in Z - X$.

*Proof:* As all $f_i(s)$ are twice differentiable convex functions, $\frac{1}{n}\sum_{i=1}^n f_i(s)$ is a twice differentiable convex function as well. It follows that

$$\begin{aligned}
&\frac{1}{n}\sum\nolimits_{i=1}^n f_i(s^0)\\
\ge\; & \frac{1}{n}\sum\nolimits_{i=1}^n f_i(s) + \frac{1}{n}\sum\nolimits_{i=1}^n \nabla f_i(s)^T (s^0 - s),
\end{aligned}$$

i.e.,

$$\begin{aligned}
&\left(\frac{1}{n}\sum\nolimits_{i=1}^n \nabla f_i(s)\right)^T (s - s^0)\\
\ge\; & \frac{1}{n}\sum\nolimits_{i=1}^n f_i(s) - \frac{1}{n}\sum\nolimits_{i=1}^n f_i(s^0) > 0,
\end{aligned} \tag{23}$$

for all $s^0 \in X$ and all $s \in Z - X$. If this lemma does not hold, there exists a sequence of vectors $\{y_k \in Z - X\}$ such that $\lim_{k\to+\infty} \frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)^T (y_k - \bar{s})/\left\|\frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)\right\|/\|y_k - \bar{s}\| = 0$. Since each $[\nabla^2 f_i(s)]_{jk} = \frac{\partial^2 f(s)}{\partial s_j \partial s_k}$ is continuous, $\frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)$ is continuous. Since $Z$ is bounded, $\left\|\frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)\right\|$ and $\|y_k - \bar{s}\|$ are both upper bounded. Thus, $\lim_{k\to+\infty} \frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)^T (y_k - \bar{s}) = 0$. Let $d_k \in \mathbb{R}^m$ be an arbitrary unit vector for any $k$. Under Assumption 4(a), $\bar{s} + \frac{1}{2}\delta d_k \in X$ for all $k$, where $\delta$ is defined in Assumption 4(a). In view of the arbitrariness of the direction $d_k$, we can adopt a proper $d_k$ such that $\frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)^T d_k = \left\|\frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)\right\|$. As $k \to +\infty$, $\frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)^T (y_k - \bar{s} - \frac{1}{2}\delta d_k)/\left\|\frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)\right\|/\|y_k - \bar{s} - \frac{1}{2}\delta d_k\| = -\frac{1}{2}\delta/\|y_k - \bar{s} - \frac{1}{2}\delta d_k\|$. Since $y_k \in Z - X$ and $\bar{s} + \frac{1}{2}\delta d_k \in X$, we have that $\|y_k - \bar{s} - \frac{1}{2}\delta d_k\|$ is upper bounded from the boundedness of $X$ and $Z$ and $\|y_k - \bar{s} - \frac{1}{2}\delta d_k\| \ge \frac{1}{2}\delta$ from the definition of $\bar{s}$. Therefore, $\frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)^T (y_k - \bar{s} - \frac{1}{2}\delta d_k)/\left\|\frac{1}{n}\sum_{i=1}^n \nabla f_i(y_k)\right\|/\|y_k - \bar{s} - \frac{1}{2}\delta d_k\|$ is upper bounded by a negative constant as $k \to +\infty$, which contradicts with (23). ■

*Lemma 6:* Consider the system given by $\dot{y}(t) = -\sum_{i=1}^n \nabla f_i(y(t))/\left\|\sum_{i=1}^n \nabla f_i(y(t))\right\|$. Let $Z$ be a closed bounded convex set containing $X$. If $y(t) \in Z$ for all $t$ and Assumption 4 holds, there exists a constant $T > 0$ such that $y(t) \in X$ for all $t > T$.

*Proof:* (a) Under Assumption 4(a), Lemma 5 holds and consider the Lyapunov function candidate $V_a(t) = \|y(t) - \bar{s}\|$ for all $t$. Calculating $\dot{V}_a(t)$, we have

$$\dot{V}_a(t) = -\frac{(y(t)-\bar{s})^T \sum_{i=1}^n \nabla f_i(y(t))}{\left\|y(t)-\bar{s}\right\| \left\|\sum_{i=1}^n \nabla f_i(y(t))\right\|} \le -c_x$$

for a constant $c_x > 0$ and all $y(t) \in Z - X$. Integrating both sides of this inequality, we have $V_a(t) - V_a(0) \le -c_x t$ for all $y(t) \in Z - X$. It is clear that there exists a constant $T > 0$ such that $y(t) \in X$ for all $t > T$.

(b) Under Assumption 4(b), there is a neighborhood of $X$, denoted by $S$, and a uniform constant $0 < c_s \le 1$ such that

$$(y(t) - P_X(y(t)))^T \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(y(t)) \ge c_s \left\| \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(y(t)) \right\| \|y(t) - P_X(y(t))\|$$

for all $y(t) \in S$. Similar to the proof of Lemma 4, $\min_{y \in \bar{\partial} S} \sum_{i=1}^n \nabla f_i(y)^T \frac{y - P_X(y)}{\|y - P_X(y)\|} > 0$, where $\bar{\partial} S$ denotes the boundary of the set $S$. From Lemma 3, it follows that $\varpi \le \frac{1}{n}\sum_{i=1}^n \nabla f_i(y(t))^T \frac{y(t) - P_X(y(t))}{\|y(t) - P_X(y(t))\|}$ for some constant $\varpi > 0$ and any $y(t) \in Z - S$. Since $Z$ is bounded, $\left\|\frac{1}{n}\sum_{i=1}^n \nabla f_i(y(t))\right\| < c_g$ for some constant $c_g > 0$ and all $y(t) \in Z$. Hence

$$(y(t) - P_X(y(t)))^T \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(y(t)) \ge \frac{\varpi}{c_g} \left\| \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(y(t)) \right\| \|y(t) - P_X(y(t))\|$$

for some constant $\varpi > 0$ and any $y(t) \in Z - S$. Consider the Lyapunov function candidate $V_b(t) = \|y(t) - P_X(y(t))\|$ for all $t$. In the same way as the proof of (a), it can be proved that there exists a constant $T > 0$ such that $y(t) \in X$ for all $t > T$. ■

*Remark 5:* Under Assumption 4(a), $X$ is a nonempty closed convex set and contains at least one interior point while Assumption 4(b) considers the case that $X$ has no interior points and excludes the singular situation where

$$\lim_{\|s - P_X(s)\| \to 0} \frac{(s - P_X(s))^T}{\|s - P_X(s)\|} \times \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(s) / \left\| \frac{1}{n}\sum_{i=1}^{n} \nabla f_i(s) \right\| = 0,$$

i.e., $\frac{1}{n}\sum_{i=1}^n \nabla f_i(s)$ tends to be orthogonal to $s - P_X(s)$ as $s$ converges to $X$. In [25], some finite-time results are given for nonconvex functions, but when convex functions are considered, the results can only be used to a special case of Assumption 4(b) because the convexity of the functions was not fully exploited.

*Lemma 7:* Suppose that the graph $\mathcal{G}(t)$ is undirected and connected for all $t$ and Assumptions 1, 2 and 4 hold. For system (1) with algorithm (22), the following statements hold:

(a) $x_i(t) \in Z$ for all $t$ and a closed bounded region $Z$ and there exists a constant $T_0 > 0$ such that $x_i(t) = x_j(t)$ for all $t > T_0$;

(b) Each $\|\theta_i(t)\|$ is bounded for all $i$ and all $t$.

*Proof:* Under Assumption 1, it follows from Lemma 4 that all $X_i$ and $X$ are nonempty closed bounded convex sets for all $i$. Consider the Lyapunov function candidate $V(t) = \sum_{i=1}^n \|x_i(t) - z_0\|^2$ for some $z_0 \in X$. Let $d_D = \max\{\|y_1 - y_2\| \mid y_1, y_2 \in (X \bigcup_{i=1}^n X_i)\}$ and $Y = \{y \mid \|y - P_X(y)\| \le L_0\} \subset \mathbb{R}^m$ for some constant $L_0 > 0$ be a closed bounded convex set such that

$$\min_i\{\|y - P_X(y)\|, \|y - P_{X_i}(y)\|\} \ge d_D \quad (24)$$

for any $y \notin Y$ and

$$\max_i\{\|y - P_X(y)\|, \|y - P_{X_i}(y)\|\} \le 3d_D \quad (25)$$

for all $y \in Y$. Then, from the triangle relationship, for any $y \notin Y$, the angle between $y - P_{X_i}(y)$ and $y - z_0$ is no larger than $\frac{\pi}{3}$ for all $i$. That is,

$$(y - P_{X_i}(y))^T (y - z_0) \ge \tfrac{1}{2}\|y - P_{X_i}(y)\| \|y - z_0\| \quad (26)$$

for any $y \notin Y$. Let $Z = \{y \mid \|y - P_X(y)\| \le L_1\} \subset \mathbb{R}^m$ for some constant $L_1 > L_0$ be a closed bounded convex set such that $Y \subset Z$, $x_i(0) \in Z$ and

$$\min_i\{\|y - P_{X_i}(y)\|, \|y - z_0\|\} > \max\{8n + 6nd_D, 0.5 \max_{z_1 \in Z} \|z_1 - z_0\|\} \quad (27)$$

for all $i$ and any $y \notin Z$.

We first consider the case where $\theta_i(t) \neq \theta_j(t)$ or $x_i(t) \neq x_j(t)$ for some $i \neq j$. Suppose that there exists an agent $i_0$ such that $x_{i_0}(t) \notin Z$. Then there must exist an agent $i_1$ such that $x_{i_0}(t) = x_{i_1}(t) \notin Z$ and $gc_{i_1}(t) \neq 0$. If this is not true, from (22), $x_j(t) = x_{i_0}(t) \notin Z$ and $\theta_j(t) = \theta_{i_0}(t)$ for all $j \in N_{i_0}(t)$. Since the graph $\mathcal{G}(t)$ is undirected and connected, it follows that $x_i(t) = x_j(t) \notin Z$ and $\theta_i(t) = \theta_j(t)$ for all $i, j$. This yields a contradiction. Without loss of generality, suppose that $\|x_{i_1}(t) - z_0\| = \max\{\|x_i(t) - z_0\| \mid x_i(t) \notin Z, gc_i(t) \neq 0\}$. Clearly, $\|x_{i_1}(t) - z_0\| = \max_i\{\|x_i(t) - z_0\| \mid x_i(t) \notin Z\}$.

Calculating $\dot{V}(t)$, from (16) and (27), we have

$$\begin{aligned}\dot{V}(t) = & -\sum_{i=1}^{n} (x_i(t) - z_0)^T \left[\frac{\theta_i(t)}{\|\theta_i(t)\|} + gc_i(t) \right. \\ & \left. - \sum_{j \in N_i(t)} q_{ij}(t) \mathrm{sgn}(x_j(t) - x_i(t))\right] \\ \le & \sum_{i=1}^{n} \sum_{j \in N_i(t)} \frac{q_{ij}(t)}{2} \big(x_i(t) - x_j(t)\big)^T \\ & \times \mathrm{sgn}\big(x_j(t) - x_i(t)\big) + \sum_{i=1}^{n} \|x_i(t) - z_0\| \\ & - \sum_{i=1}^{n} (x_i(t) - z_0)^T gc_i(t) \\ \le & \; 2n\|x_{i_1}(t) - z_0\| - \sum_{i=1}^{n} (x_i(t) - z_0)^T gc_i(t).\end{aligned}$$

From (22), (25) and (26), we have $(x_i(t) - z_0)^T gc_i(t) \le 3\|x_{i_1}(t) - z_0\| d_D$ for any $x_i(t) \in Y$ and $(x_i(t) - z_0)^T (x_i(t) - P_{X_i}(x_i(t))) \ge \frac{1}{2}\|x_i(t) - z_0\| \|x_i(t) - P_{X_i}(x_i(t))\|$ for any $x_i(t) \notin Y$. It follows from (27) that $-\sum_{i=1}^n (x_i(t) - z_0)^T gc_i(t) \le -\|x_{i_1}(t) - z_0\|(4n + 3nd_D - 3nd_D)$. Thus, $\dot{V}(t) \le -2n\|x_{i_1}(t) - z_0\|$.

Now, we consider the case where $\theta_i(t) = \theta_j(t)$ and $x_i(t) = x_j(t)$ for all $i, j$. Calculating $\dot{V}(t)$, we have

$$\begin{aligned}\dot{V}(t) = & -\sum_{i=1}^{n} (x_i(t) - z_0)^T \left[\frac{\theta_i(t)}{\|\theta_i(t)\|} \right. \\ & \left. - \sum_{j \in N_i(t)} q_{ij}(t) \mathrm{sgn}(x_j(t) - x_i(t))\right] \\ \le & -\sum_{i=1}^{n} (x_i(t) - z_0)^T \frac{\theta_i(t)}{\|\theta_i(t)\|}. \quad (28)\end{aligned}$$

Since $\mathcal{G}(t)$ is connected and $\psi_i(0) = 0$ for all $i$, we have $\sum_{i=1}^n \dot{\psi}_i(t) = 0$ and thus $\sum_{i=1}^n \psi_i(t) = 0$, which implies that

$$\frac{1}{n}\sum_{i=1}^{n}\theta_i(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\nabla f_i(x_i(t))\right]$$

for all $t$. Since $\theta_i(t) = \theta_j(t)$ and $x_i(t) = x_j(t)$ for all $i, j$, we have

$$\begin{aligned}\theta_i(t) =& \frac{1}{n}\sum_{i=1}^{n}\theta_i(t) = \frac{1}{n}\sum_{i=1}^{n}\left[\nabla f_i(x_i(t))\right]\\ =& \frac{1}{n}\sum_{i=1}^{n}\left[\nabla f_i\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} x_i(t)\right)\right] \end{aligned} \tag{29}$$

for all $i$. Moreover, since $z_0 \in X$, from the convexity of the functions $f_i(s)$, we have

$$\begin{aligned} 0 \le& \frac{1}{n}\sum\nolimits_{i=1}^{n} f_i\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} x_i(t)\right) - \frac{1}{n}\sum\nolimits_{i=1}^{n} f_i(z_0)\\ \le& \frac{1}{n}\sum\nolimits_{i=1}^{n}\nabla f_i\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} x_i(t)\right)\left(\frac{1}{n}\sum\nolimits_{i=1}^{n} x_i(t) - z_0\right) \end{aligned} \tag{30}$$

It follows from (28), (29) and (30) that $\dot{V}(t) \le 0$.

Summarizing both cases, we have $\dot{V}(t) \le 0$ if there exists an agent $i$ such that $x_i(t) \notin Z$. Since $x_i(0) \in Z$, then $x_i(t) \in Z$ for all $i$ and all $t$. Then similar to the proof of Theorem 2, it can be proved that there exists a constant $T_0 > 0$ such that $x_i(t) = x_j(t) = x^*(t)$, where $x^*(t)$ is defined in (11), for all $i$ and all $t > T_0$. It is clear that

$$\dot{x}^*(t) = -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\theta_i(t)}{\|\theta_i(t)\|} + gc_i(t)\right)$$

for all $t > T_0$. Define $A_{1k}(t) \triangleq \{i \mid \theta_{ik}(t) = \max_i\{\theta_{ik}(t)\}\}$, $A_{2k}(t) \triangleq \{i \mid \theta_{ik}(t) = \min_i\{\theta_{ik}(t)\}\}$, $\bar{\theta}_k(t) \triangleq \frac{1}{|A_{1k}(t)|}\sum_{i\in A_{1k}(t)}\theta_{ik}(t)$ and $\underline{\theta}_k(t) \triangleq \frac{1}{|A_{2k}(t)|}\sum_{i\in A_{2k}(t)}\theta_{ik}(t)$, where $|A_{1k}(t)| \ge 1$ and $|A_{2k}(t)| \ge 1$ denote, respectively, the cardinality of $A_{1k}(t)$ and $A_{2k}(t)$. It is clear that the $k$th component of each $\dot{\theta}_i(t)$ can be written as

$$\begin{aligned}\dot{\theta}_{ik}(t) =& \sum\nolimits_{j\in N_i(t)} p_{ij}(t)\operatorname{sgn}(\theta_{jk}(t) - \theta_{ik}(t))\\ &+ [\nabla^2 f_i(x^*(t))\dot{x}^*(t)]_k \end{aligned} \tag{31}$$

for $t > T_0$.

Suppose that $\bar{\theta}_k(T_1) \ge \max_{i,x^*(t)\in Z}\left\{\left\|\nabla f_i(x^*(t))\right\|\right\}$ for some $T_1 > T_0$ and $\bar{\theta}_k(t) \ne \underline{\theta}_k(t)$ for $t > T_1$. Let $C_{1k}(t) \triangleq \{(i,j) \in \mathcal{E}(\mathcal{G}(t)) \mid i \in A_{1k}(t), j \notin A_{1k}(t)\}$. Since the graph $\mathcal{G}(t)$ is connected and $\bar{\theta}_k(t) \ne \underline{\theta}_k(t)$, then $C_{1k}(t)$ is nonempty. Moreover, since $x^*(t) \in Z$ and each entry of $\nabla^2 f_i(x^*(t))$ is continuous from Assumption 4, then $d_\mu = \max_{i,x^*(t)\in Z}\left\{\left\|\nabla^2 f_i(x^*(t))\dot{x}^*(t)\right\|\right\}$ is bounded. Let $\alpha(t) = \min_{(i,j)\in C_{1k}(t)} p_{ij}(t)$. Note that $\operatorname{sgn}(\theta_{hk}(t) - \theta_{lk}(t)) \le 0$ for any $l \in A_{1k}(t)$ and any $h \in N_l(t)$. It is clear from (31) that $\dot{\bar{\theta}}_k(t) = \frac{1}{|A_{1k}(t)|}\sum_{i\in A_{1k}(t)}\dot{\theta}_{ik}(t) \le d_\mu - \frac{\alpha(t)}{n}$. From the dynamics of $p_{ij}(t)$, we have $\dot{p}_{ij}(t) = 1$ for any $(i,j) \in C_{1k}(t)$. Note that the number of all parameters $p_{ij}$ is finite, denoted by $n_e$, and it takes at most $nd_\mu$ time for each $p_{ij}$ to increase from 0 to $nd_\mu$ at the rate of 1. Since $\dot{\bar{\theta}}_k(t) \le d_\mu$ and especially $\dot{\bar{\theta}}_k(t) < 0$ when $\alpha(t) > nd_\mu$, we have that it takes at most $nn_e d_\mu$ time for $\alpha(t)$ to increase to $nd_\mu$ when $\bar{\theta}_k(t) \ne \underline{\theta}_k(t)$. Note that when $\bar{\theta}_k(t) = \underline{\theta}_k(t)$, we have $\|\bar{\theta}_k(t)\| = \|\underline{\theta}_k(t)\| = \left\|\frac{1}{n}\sum_{i=1}^n\left[\nabla f_i(x^*(t))\right]_k\right\| \le \max_{i,x^*(t)\in Z}\left\{\left\|\nabla f_i(x^*(t))\right\|\right\}$. Hence, $\bar{\theta}_k(t) < \bar{\theta}_k(T_1) + nn_e d_\mu^2$ for all $t > T_1$. Thus, $\bar{\theta}_k(t)$ is upper bounded for all $t$ and all $k$. In the same way, it can be proved that $\underline{\theta}_k(t)$ is lower bounded for all $t$ and all $k$. Thus, $\|\theta_i(t)\|$ is bounded for all $i$ and all $t$. ■

*Theorem 4:* Suppose that the graph $\mathcal{G}(t)$ is undirected and connected for all $t$ and Assumptions 1, 2 and 4 hold. For system (1) with algorithm (22), all agents reach a consensus and minimize the team objective function (2) in finite time.

*Proof:* Under Assumptions 1, 2 and 4, Lemma 7 holds. Hence, $x_i(t) \in Z$ for all $t$ and a closed bounded region $Z$ and there exists a constant $T_0 > 0$ such that $x_i(t) = x_j(t) = x^*(t)$, where $x^*(t)$ is defined in (11), for all $i$ and all $t > T_0$. Moreover, from Lemma 7, each $\|\theta_i(t)\|$ is bounded for all $i$ and all $t$. Then, similar to the proof of Theorem 2, it can be proved that all $\theta_i(t)$ reach a consensus in finite time. That is, there exists a number $T_1 > T_0$ such that $\theta_i(t) = \theta_j(t) \triangleq \theta^*(t)$ for all $t > T_1$.

As a result, we have

$$\dot{x}^*(t) = -\theta^*(t)/\|\theta^*(t)\|$$

for all $t > T_1$. Recalling (29), from Lemma 6, the team objective function (2) will be minimized in finite time. ■

*Remark 6:* Due to the existence and the nonlinearity of the objective functions, the existing approaches for the distributed finite-time consensus problem (e.g., [17], [18]) cannot be extended directly to the distributed finite-time optimization problems, which need to consider the finite-time convergence of the consensus of the agents and the finite-time convergence of the objective functions simultaneously. Although some results have been obtained in our previous works in [19], [20] for the distributed finite-time optimization problem, they are limited to a special class of convex objective functions that have a quadratic-like form and the approaches cannot be applied to more general convex objective functions.

## IV. Distributed Continuous-Time Optimization With a Common Convex Constraint Set

In this section, we will extend the results in Sections III.C and III.D and design algorithms for system (1) to solve a distributed optimization problem with a common convex constraint set as follows

$$\begin{aligned} &\text{minimize } \sum\nolimits_{i=1}^{n} f_i(x_i)\\ &\text{subject to } x_i = x_j \in \mathcal{H} \subset \mathbb{R}^m, \end{aligned} \tag{32}$$

where $\mathcal{H}$ is a closed convex set.

### A. Distributed Optimization Algorithm With Nonuniform Gradient Gains

In this subsection, we extend Theorem 3 to the problem (32) for general convex local objective functions. Let $\mathcal{X} \subset \mathbb{R}^m$ denote the optimal set of the problem (32).

When $\mathcal{H}$ is a closed bounded convex set, the algorithm is given by

$$\begin{aligned} \dot{q}_i(t) =&\ \arctan(e^{\|x_i(t)\|}), q_i(0) > 0, \\ u_i(t) =&\ \sum\nolimits_{j\in N_i(t)} \mathrm{sgn}\big(x_j(t) - x_i(t)\big) - gr_i(t) - gc_i(t) \\ gr_i(t) =&\ \frac{\nabla f_i(x_i(t))}{\sqrt{q_i(t)}} \\ gc_i(t) =&\ \frac{\gamma_i(t)[x_i(t) - P_{\mathcal{H}}(x_i(t))]}{\|x_i(t) - P_{\mathcal{H}}(x_i(t))\|} \end{aligned} \tag{33}$$

for all $i$, where $\gamma_i(t) > |N_i(t)| + 1$ and $|N_i(t)|$ denotes the cardinality of $N_i(t)$.

*Lemma 8:* Under Assumption 1, $\mathcal{X}$ is a nonempty closed bounded convex set.

*Proof:* When $\mathcal{H}$ is a bounded closed convex set, from the property of continuous functions on closed bounded sets and the convexity of the functions $f_i(s)$ and the set $\mathcal{H}$, it is easy to see that $\mathcal{X}$ is a nonempty closed bounded convex set.

Under Assumption 1, Lemma 4 holds. Thus, $\lim_{\|y\|\to+\infty} f_i(y) = +\infty$ and $\lim_{\|y\|\to+\infty} \sum_{i=1}^n f_i(y) = +\infty$. Since $\sum_{i=1}^n f_i(y)$ is lower bounded in $\mathcal{H}$, its infimum exists. Similar to the proof of Lemma 4, it can be proved that $\mathcal{X}$ is a nonempty closed bounded convex set, when $\mathcal{H}$ is an unbounded closed convex set. ■

*Theorem 5:* Suppose that the graph $\mathcal{G}(t)$ is undirected and connected for all $t$ and Assumptions 1 and 2 hold. For system (1) with algorithm (33), all agents reach a consensus in finite time and minimize the team objective function (32) as $t \to +\infty$.

*Proof:* Note that $\pi/4 \le \arctan(e^{\|x_i(s)\|}) \le \pi/2$ for all $s$ and all $i$. There exist a constant $T > 0$ such that $2\sqrt{t} > \sqrt{q_i(t)} > \frac{\sqrt{t}}{2}$ for all $i$ and all $t > T$. Consider the Lyapunov function candidate $V(t) = \frac{1}{2}\|x_i(t) - z\|^2$ for $z \in \mathcal{H}$ and all $t$. Under Assumption 1, it follows from Lemma 4 that all $X_i$ and $X$ are nonempty closed bounded convex sets for all $i$. Let $Y$ be a closed bounded convex set such that $x_i(T) \in Y$, $X \subset Y$, $X_i \subset Y$ and $f_i(x_i(t)) - f_i(z) \ge 4\sum_{j=1,j\neq i}^n [f_j(z) - f_j(z_j)]$ for all $i$, all $z_j \in X_j$ and all $x_i(t) \notin Y$. It follows that $\frac{1}{\sqrt{q_i(t)}}[f_i(x_i(t)) - f_i(z)] \ge \sum_{j=1,j\neq i}^n \frac{1}{\sqrt{q_j(t)}}[f_j(z) - f_j(z_j)]$ for all $t > T$, all $i$, all $z_j \in X_j$ and all $x_i(t) \notin Y$. Moreover, since $z \in \mathcal{H}$, from Lemma 2, we have $(x_i(t) - z)^T \frac{\gamma_i(t)[x_i(t) - P_{\mathcal{H}}(x_i(t))]}{\|x_i(t) - P_{\mathcal{H}}(x_i(t))\|} \ge 0$. Then similar to the proof of Theorem 2, it can be proved that all agents remain in a bounded region and each $\big\|\nabla f_i(x_i(t))\big\|$ is bounded for all $i$ and all $t$. Then there exist two constants $T_0 > T$ and $\mu_c > 0$ such that $2\sqrt{t} > \sqrt{q_i(t)} > \frac{\sqrt{t}}{2} > \mu_c > 8n\big\|\nabla f_i(x_i(t))\big\|$ for all $i$ and all $t > T_0$.

Consider the Lyapunov function candidate

$$V_i(t) = \frac{1}{2}\|x_i(t) - P_{\mathcal{H}}(x_i(t))\|^2$$

for all $i$. Calculating $\dot{V}_i(t)$, we have for all $t > T_0$,

$$\begin{aligned} \dot{V}_i(t) =&\ -(x_i(t) - P_{\mathcal{H}}(x_i(t)))^T \big[gr_i(t) + gc_i(t) \\ &- \sum\nolimits_{j\in N_i(t)} \mathrm{sgn}\,(x_j(t) - x_i(t))\,\big] \\ \le&\ -\|x_i(t) - P_{\mathcal{H}}(x_i(t))\| \left[\gamma_i(t) - |N_i(t)| - \frac{1}{8n}\right] \\ \le&\ -\frac{7}{8}\|x_i(t) - P_{\mathcal{H}}(x_i(t))\| \\ \le&\ -\frac{7}{8}\sqrt{2V_i(t)} \end{aligned}$$

where the second inequality holds since $\gamma_i(t) > |N_i(t)| + 1$ and $n \ge 1$. It follows that $\frac{\dot{V}_i(t)}{\sqrt{2V_i(t)}} \le -\frac{7}{8}$. Integrating both sides of this inequality from $T_0$ to $t$, we have $2\sqrt{V_i(t)}/\sqrt{2} - 2\sqrt{V_i(T_0)}/\sqrt{2} \le -\frac{7}{8}(t - T_0)$. Thus, $V_i(t)$ vanishes to zero in finite time. That is, there exist a constant $T_1 > T_0$ such that $x_i(t) \in \mathcal{H}$ and $\dot{x}_i(t) = \sum_{j\in N_i(t)} \mathrm{sgn}\big(x_j(t) - x_i(t)\big) - gr_i(t)$ for all $i$ and all $t > T_1$. Since $\sqrt{q_i(t)} > 8n\big\|\nabla f_i(x_i(t))\big\|$ for all $i$ and all $t > T_0$, similar to the proof of Proposition 1, it can be proved that all agents reach a consensus in finite time. That is, there exists a constant $T_2 > T_1$ such that $x_i(t) = x^*(t)$, where $x^*(t)$ is defined in (11), for all $i$ and all $t > T_2$. For $t > T_2$, we have

$$\dot{x}^*(t) = -\tfrac{1}{n}\sum\nolimits_{i=1}^n gr_i(t).$$

Now, we prove that the team objective function (32) can be minimized as $t \to +\infty$. Under Assumption 1, Lemma 8 holds and hence $\mathcal{X}$ is a nonempty closed bounded convex set. Consider the Lyapunov function candidate

$$\bar{V}(t) = \frac{1}{2}\|x^*(t) - P_{\mathcal{X}}(x^*(t))\|^2$$

for all $t > T_2$. After some calculations, we have

$$\begin{aligned} &\dot{\bar{V}}(t) \\ &\quad = -[x^*(t) - P_{\mathcal{X}}(x^*(t))]^T \frac{1}{n}\sum\nolimits_{i=1}^n \frac{\nabla f_i(x^*(t))}{\sqrt{q_i(t)}} \\ &\quad = -[x^*(t) - P_{\mathcal{X}}(x^*(t))]^T \frac{1}{n}\sum\nolimits_{i=1}^n \frac{\nabla f_i(x^*(t))}{\sqrt{q^*(t)}} \frac{\sqrt{q^*(t)}}{\sqrt{q_i(t)}} \end{aligned}$$

where $q^*(t)$ is defined in the proof of Theorem 3. Similar to the proof of Theorem 3, we can let $T_2$ be sufficiently large such that $2\sqrt{t} > \sqrt{q^*(t)} > \frac{\sqrt{t}}{2}$ and $\left|1/\sqrt{1 + \frac{q_i(T_0)}{q^*(t)}} - 1\right| < \epsilon/\mu_c$ for any small $\epsilon > 0$, all $i$ and all $t > T_2$. Since $\frac{q^*(t)}{q_i(t)} = 1/(1 + \frac{q_i(T_0)}{q^*(t)})$, we have

$$\begin{aligned}\dot{\bar{V}}(t) &\le -[x^*(t) - P_{\mathcal{X}}(x^*(t))]^T \frac{1}{n}\sum_{i=1}^n \frac{\nabla f_i(x^*(t))}{\sqrt{q^*(t)}} \\ &\quad + \|x^*(t) - P_{\mathcal{X}}(x^*(t))\| \frac{\epsilon}{n\sqrt{q^*(t)}} \\ &\le -\frac{1}{n\sqrt{q^*(t)}}\sum_{i=1}^n [f_i(x^*(t)) - f_i(P_{\mathcal{X}}(x^*(t)) \\ &\quad - \epsilon\|x^*(t) - P_{\mathcal{X}}(x^*(t))\|],\end{aligned}$$

for all $t > T_2$, where the second inequality has used the convexity of the functions $f_i(x^*(t))$, i.e.,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n [f_i(P_{\mathcal{X}}(x^*(t)) - f_i(x^*(t))] \\ &\quad \ge \frac{1}{n}\sum_{i=1}^n \nabla f_i(x^*(t))^T [P_{\mathcal{X}}(x^*(t)) - x^*(t)].\end{aligned}$$

Since all agents remain in a bounded region and $\mathcal{X}$ is bounded, $\|x^*(t) - P_{\mathcal{X}}(x^*(t))\|$ is bounded. That is, $\|x^*(t) - P_{\mathcal{X}}(x^*(t))\| < \mu_p$ for some constant $\mu_p > 0$. Let $E = \{s \in \mathbb{R}^m \mid \|s - P_{\mathcal{X}}(s)\| \le l_1\}$ for some constant $0 < l_1 \le \max_{s\in\mathcal{H}} \|s - P_{\mathcal{X}}(s)\|$ and $\rho = \min_{s\in\mathcal{H}\cap\bar{\partial}E} \sum_{i=1}^n [f_i(s) - f_i(P_{\mathcal{X}}(s))]$, where $\bar{\partial}E$ denotes the boundary of $E$. Since $P_{\mathcal{X}}(s) \in \mathcal{X}$, from the definition of $\mathcal{X}$, we have $\rho > 0$. From Lemma 3, $\sum_{i=1}^n [f_i(s) - f_i(P_{\mathcal{X}}(s))] > \rho$ for any $s \notin E$ and $s \in \mathcal{H}$. Let $T_2$ be further large for any given $l_1 > 0$ such that $\epsilon < \frac{\rho}{4\mu_p}$. Recall that $2\sqrt{t} > \sqrt{q^*(t)} > \frac{\sqrt{t}}{2}$ for all $t > T_2$. It follows that for any $t > T_2$, $\dot{\bar{V}}(t) \le -\frac{1}{2n\sqrt{t}}\rho + \frac{2}{n\sqrt{t}}\mu_p\epsilon < 0$. Integrating both sides of this inequality from $T_2$ to $t$, we have $\bar{V}(t) \le (-\rho + 4\mu_p\epsilon)(\sqrt{t} - \sqrt{T_2})/n$. This implies that there exists a constant $T_3 > T_2$ for any $l_1 > 0$ such that $\|x^*(t) - P_{\mathcal{X}}(x^*(t))\| < l_1$ and $x^*(t) \in E \cap \mathcal{H}$ for all $t > T_3$. In view of the arbitrariness of $l_1$, letting $l_1 \to 0$, we have $\lim_{t\to+\infty} \|x^*(t) - P_{\mathcal{X}}(x^*(t))\| = 0$. That is, the team objective function (32) is minimized as $t \to +\infty$. ■

*Remark 7:* On the boundary of $\mathcal{H}$, there might rise a switching surface due to the term $\frac{\gamma_i(t)[x_i(t)-P_{\mathcal{H}}(x_i(t))]}{\|x_i(t)-P_{\mathcal{H}}(x_i(t))\|}$. But the term $\frac{\gamma_i(t)[x_i(t)-P_{\mathcal{H}}(x_i(t))]}{\|x_i(t)-P_{\mathcal{H}}(x_i(t))\|}$ does not decrease but increases the convergence rate of the Lyapunov function $\bar{V}(t) = \frac{1}{2}\|P_{\mathcal{X}}(x^*(t)) - x^*(t)\|^2$ for all $t > T_2$. This is because at the switching surface, the angle between the vectors $P_{\mathcal{X}}(x^*(t)) - x^*(t)$ and $-\frac{1}{n}\sum_{i=1}^n \frac{\gamma_i(t)[x^*(t)-P_{\mathcal{H}}(x^*(t))]}{\|x^*(t)-P_{\mathcal{H}}(x^*(t))\|}$ is no larger than $\frac{\pi}{2}$, i.e., $-[P_{\mathcal{X}}(x^*(t)) - x^*(t)]^T \frac{1}{n}\sum_{i=1}^n \frac{\gamma_i(t)[x^*(t)-P_{\mathcal{H}}(x^*(t))]}{\|x^*(t)-P_{\mathcal{H}}(x^*(t))\|} \ge 0$, from Lemma 2.

*Remark 8:* The approach in [10] is to analyze the convergence of the largest distance from the agents to the constraint set so as to yield a contradiction to prove the optimal convergence. Due to the unboundedness of the local objective functions and the nonuniformity of the gradient gains, the approach in [10] cannot be applied in this paper. The approach in this paper is to analyze the convergence rates of the consensus, the distance to the constraint set, and the optimization by fully exploiting the convexity of the objective functions and the constraint set, and it can be used to deal with the case of the unbounded closed convex set.

*Remark 9:* In Theorems 1, 2, 3 and 5, we assume that each local objective function $f_i(x)$ is differentiable for discussion convenience. The results obtained in these theorems can be extended to more general nondifferentiable convex functions by using a minimum norm subgradient, denoted by $ls_i(s)$. That is, $ls_i(s) = \arg\min_{z\in\partial f_i(s)} \|z\|$, where $\partial f_i(s)$ denotes the subgradient set of $f_i(s)$ at $s$. However, it should be noted that from the convexity of the convex function $f_i(s)$, it can be proved that $f_i(s)$ is minimized if and only if $ls_i(s) = 0$. It should also be noted that when the minimum norm subgradients are used, after all agents reach a consensus, the Lyapunov function $\|x^*(t) - x_e\|^2$ for $x_e \in X$ or $x_e \in \mathcal{X}$ should be used instead to prove that all agents minimize the team objective function as $t \to +\infty$.

*Remark 10:* Since the solution sets for linear inequalities or equalities in the form of $h(x) \ge 0$ or $h(x) = 0$ are usually closed convex sets, Theorem 5 might be used to deal with the distributed optimization problem with linear inequality or equality constraints if its optimal set is bounded.

## B. Distributed Finite-Time Optimization Algorithm

In this subsection, we extend Theorem 4 to the problem (32) for general convex local objective functions. The algorithm is given by

$$\begin{aligned}\dot{\psi}_i(t) &= \sum_{j\in N_i(t)} p_{ij}(t)\text{sgn}(\theta_j(t) - \theta_i(t)), \\ \theta_i(t) &= \psi_i(t) + \nabla f_i(x_i(t)), \psi_i(0) = 0, \\ \dot{p}_{ij}(t) &= \begin{cases}\text{sgn}(\max_{s\in[t-c_0,t]} \|\theta_j(s) - \theta_i(s)\|), \text{ if } (i,j) \in \mathcal{G}(t), \\ 0, \text{ otherwise},\end{cases} \\ p_{ij}(0) &= p_{ji}(0) = 0, \\ u_i(t) &= \sum_{j\in N_i(t)} q_{ij}(t)\text{sgn}(x_j(t) - x_i(t)) - \frac{\theta_i(t)}{\|\theta_i(t)\|} \\ &\quad - \frac{\gamma_i(x_i(t) - P_{\mathcal{H}}(x_i(t)))}{\|x_i(t) - P_{\mathcal{H}}(x_i(t))\|}, \\ \dot{q}_{ij}(t) &= \begin{cases}\text{sgn}(\max_{s\in[t-c_0,t]} \|x_j(s) - x_i(s)\|), \text{ if } (i,j) \in \mathcal{G}(t), \\ 0, \text{ otherwise},\end{cases} \\ q_{ij}(0) &= q_{ji}(0) = 0,\end{aligned} \tag{34}$$

where $c_0 > 0$ is an arbitrary constant, $\gamma_i > 1$, and $\theta_i(t)$ and $\psi_i(t)$ are the internal states of the dynamic averaging estimator for all $i$.

*Assumption 5:* Suppose that each $[\nabla^2 f_i(s)]_{jk} = \frac{\partial^2 f(s)}{\partial s_j \partial s_k}$ is continuous with respect to $s$, and either one of the following conditions holds:

(a) There exists a scalar $\delta > 0$ and a vector $\bar{s} \in \mathcal{X}$ such that $\{\xi | \|\xi - \bar{s}\| \le \delta\} \subset \mathcal{X}$.

(b) There is a neighborhood of $\mathcal{X}$, denoted by $S$, and a uniform constant $0 < c_s \le 1$ such that $(s - P_{\mathcal{X}}(s))^T \frac{1}{n}\sum_{i=1}^n \nabla f_i(s) \ge c_s \|\frac{1}{n}\sum_{i=1}^n \nabla f_i(s)\| \|s - P_{\mathcal{X}}(s)\|$ for all $s \in S$.

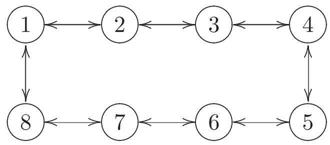



Fig. 1. One undirected graph.

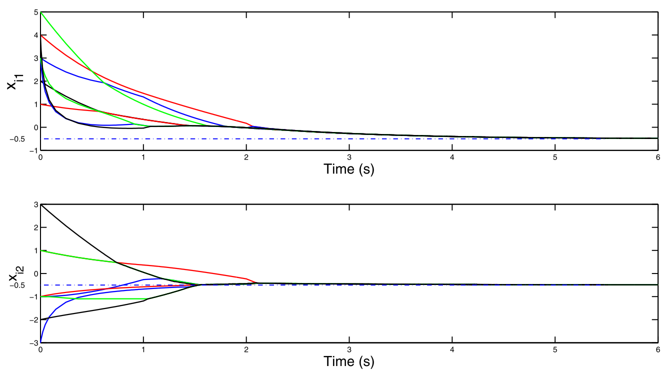



Fig. 2. State trajectories of all agents using (14).

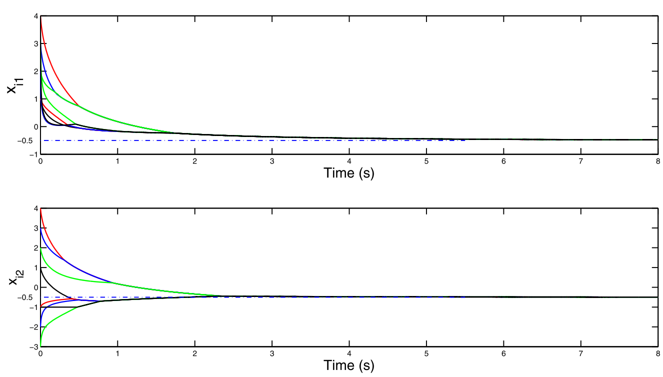



Fig. 3. State trajectories of all agents using (19).

*Theorem 6:* Suppose that the graph $\mathcal{G}(t)$ is undirected and connected for all $t$ and Assumptions 1, 2 and 5 hold. For system (1) with algorithm (34), all agents reach a consensus and minimize the team objective function (32) in finite time.

*Proof:* This theorem can also be proved based on the ideas of the proofs of Theorems 4 and 5 and hence its proof is omitted. ∎

## V. SIMULATIONS

Consider a multi-agent system with 8 continuous-time agents in $\mathbb{R}^2$. For the algorithms (14), (19), (22), (33) and (34), the communication graph is randomly switched among connected graphs, the union of which is shown in Fig. 1. The local objective functions are adopted as

$$f_1(x_1) = \frac{1}{2}x_{11}^2 + \frac{1}{2}x_{12}^2,$$

$$f_2(x_2) = \frac{1}{2}(x_{21}+1)^2 + \frac{1}{2}x_{22}^2,$$

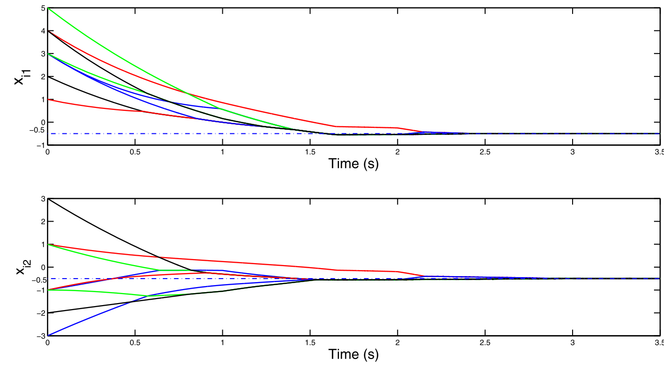



Fig. 4. State trajectories of all agents using (22).

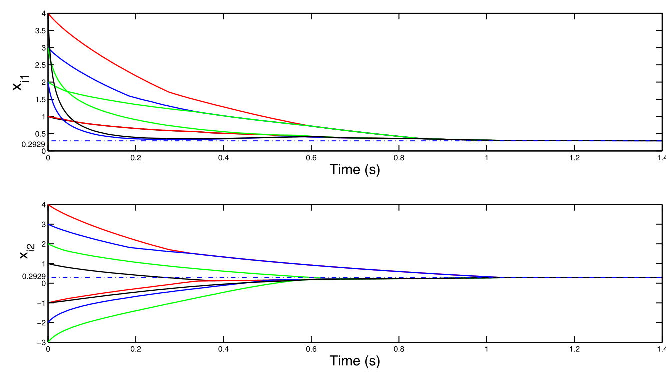



Fig. 5. State trajectories of all agents using (33).

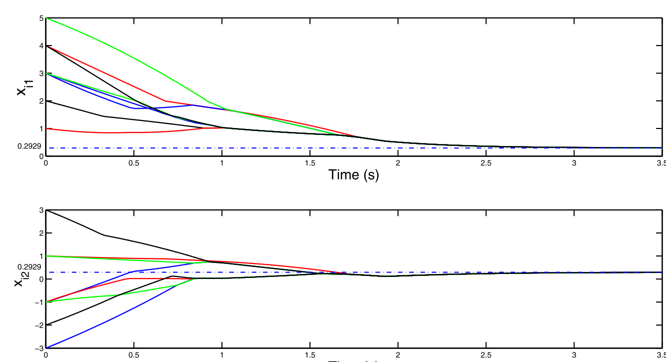



Fig. 6. State trajectories of all agents using (34).

$$f_3(x_3) = \frac{1}{2}x_{31}^2 + \frac{1}{2}(x_{32}+1)^2,$$

$$f_4(x_4) = \frac{1}{2}(x_{41}+1)^2 + \frac{1}{2}(x_{42}+1)^2,$$

$$f_5(x_5) = \frac{1}{4}x_{51}^4 + \frac{1}{4}x_{52}^4,$$

$$f_6(x_6) = \frac{1}{4}(x_{61}+1)^4 + \frac{1}{4}x_{62}^4,$$

$$f_7(x_7) = \frac{1}{4}x_{71}^4 + \frac{1}{4}(x_{72}+1)^4$$

and

$$f_8(x_8) = \frac{1}{4}(x_{81} + 1)^4 + \frac{1}{4}(x_{82} + 1)^4,$$

where $x_{i1}$ and $x_{i2}$ denote the 1st and 2nd components of $x_i$. The constrained convex set is adopted as $\mathcal{H} = \{s \in \mathbb{R}^2 \mid \|s - [1, 1]^T\| \leq 1\}$. According to Lemma 1, by calculating the solution of $\sum_{i=1}^n \nabla f_i(s) = 0$, we have that the team objective function (2) is minimized if and only if $s = [-0.5, -0.5]^T$. From the convexity of the function $\sum_{i=1}^n f_i(s)$, its local optimal point is also its global optimal point when there are no constraints. Since $[-0.5, -0.5]^T \notin \mathcal{H}$, the team objective function (32) must have at least one optimal point at the boundary of $\mathcal{H}$. By calculating the values of $\sum_{i=1}^n f_i(s)$ along the boundary of $\mathcal{H}$, we have that $s^* \doteq [0.2929, 0.2929]^T$ is one of the optimal points of the team objective function (32). Note that $\sum_{i=1}^n \nabla f_i(s)$ at $s^*$ is approximately equal to $[7.5443, 7.5443]^T$ and orthogonal to the tangent line of $\mathcal{H}$ at $s^*$. Thus, the angle between the vectors $\sum_{i=1}^n \nabla f_i(s^*)$ and $y - s^*$ is smaller than $\pi/2$, i.e., $\sum_{i=1}^n \nabla f_i(s^*)^T (y - s^*) > 0$, for any $y \in \mathcal{H} - \{s^*\}$. Hence from the convexity of the functions $f_i$, $\sum_{i=1}^8 f_i(y) - \sum_{i=1}^8 f_i(s^*) \geq \sum_{i=1}^n \nabla f_i(s^*)^T (y - s^*) > 0$ for any $y \in \mathcal{H} - \{s^*\}$. That is, the team objective function (32) is minimized if and only if $s = s^*$. The simulation results are shown in Figs. 2–6. We use dash-dot lines to denote the two components of the optimal state. Specifically, for the algorithms (14) and (19), consensus is reached, respectively, at about 2.1 s and 2.2 s and the team objective function (2) is minimized as $t \to +\infty$. For algorithm (22), consensus is reached at about 2.2 s and the team objective function (2) is minimized at about 2.5 s. For the algorithm (33), consensus is reached at about 1 s and the team objective function (32) is minimized as $t \to +\infty$. For algorithm (34), consensus is reached at about 1.7 s and the team objective function (32) is minimized at about 2.6 s. Clearly, all these simulation results are consistent with the obtained theorems.

## VI. Conclusion

In this paper, a distributed continuous-time optimization problem was studied with the consideration of nonuniform gradient gains, finite-time convergence, and a common convex constraint set. Six distributed algorithms were given. The first three and the fifth dealt with a distributed gradient optimization problem for general differentiable convex local objective functions. The fourth and the sixth dealt with a distributed finite-time optimization problem using a combination of a distributed tracking algorithm and a distributed dynamic averaging estimator. For the first three and the fifth algorithms, it has been shown that the agents reach a consensus in finite time while minimizing the team objective function as time evolves. In particular, it has been shown that the third and the fifth algorithms can be used to deal with general differentiable convex local objective functions with nonuniform gradient gains, and their gradient gains are state-dependent and need not to be known in advance. For the fourth and the sixth algorithms, it has been shown that all agents reach a consensus while minimizing the team objective function in finite time. In addition, it has been shown that the last two algorithms can be used to deal with a distributed continuous-time optimization problem with a common convex constraint set. Our future work will be directed towards the case of directed graphs with nonuniform convex constraint sets.

## References


[1] A. Nedić, A. Ozdaglar, and P. A. Parrilo, “Constrained consensus and optimization in multi-agent networks,” *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, 2010.

[2] A. Nedić and A. Ozdaglar, “Distributed subgradient methods for multi-agent optimization,” *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, 2009.

[3] A. Nedić and A. Ozdaglar, “Distributed optimization over time-varying directed graphs,” *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 601–615, 2015.

[4] E. Wei, A. Ozdaglar, and A. Jadbabaie, “A distributed Newton method for network utility maximization,” in *Proc. IEEE Conf. Decision Control & Eur. Control Conf.*, Orlando, FL, Dec. 2011, pp. 6612–6617.

[5] K. Srivastava and A. Nedić, “Distributed asynchronous constrained stochastic optimization,” *IEEE J. Selected Topics Signal Processing*, vol. 5, no. 4, pp. 772–790, 2011.

[6] M. Zhu and S. Martínez, “An approximate dual subgradient algorithm for multi-agent non-convex optimization,” *IEEE Trans. Autom. Control*, vol. 58, no. 6, pp. 1534–1539, 2013.

[7] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson, “Subgradient methods and consensus algorithms for solving convex optimization problems,” in *Proc. IEEE Conf. Decision Control*, Cancun, Mexico, Dec. 2008, pp. 4185–4190.

[8] J. Lu, C. Y. Tang, P. Regier, and T. D. Bow, “Gossip algorithms for convex consensus optimization over networks,” *IEEE Trans. Autom. Control*, vol. 56, no. 12, pp. 2917–2923, 2011.

[9] J. Lu and C. Y. Tang, “Zero-gradient-sum algorithms for distributed convex optimization: The continuous-time case,” *IEEE Trans. Autom. Control*, vol. 57, no. 9, pp. 2348–2354, 2012.

[10] G. Shi, K. H. Johansson, and Y. Hong, “Reaching an optimal consensus: Dynamical systems that compute intersections of convex sets,” *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 610–622, 2013.

[11] K. Kvaternik and L. Pavel, “A continuous-time decentralized optimization scheme with positivity constraints,” in *Proc. IEEE Conf. Decision Control*, Maui, HI, Dec. 2012, pp. 6801–6807.

[12] J. Wang and N. Elia, “A control perspective for centralized and distributed convex optimization,” in *Proc. IEEE Conf. Decision Control & Eur. Control Conf.*, Orlando, FL, Dec. 2011, pp. 3800–3805.

[13] B. Gharesifard and J. Cortés, “Distributed continuous-time convex optimization on weight-balanced digraphs,” *IEEE Trans. Autom. Control*, vol. 59, no. 3, pp. 781–786, 2014.

[14] S. S. Kia and J. Cortés, “Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication,” *Automatica*, vol. 55, no. 5, pp. 254–264, 2015.

[15] P. Lin, W. Ren, and Y. Song, “Distributed multi-agent optimization subject to nonidentical constraints under local communication,” *Automatica*, vol. 65, no. 3, pp. 120–131, 2016.

[16] Z. Qiu, S. Liu, and L. Xie, “Distributed constrained optimal consensus of multi-agent systems,” *Automatica*, vol. 68, no. 6, pp. 209–215, 2016.

[17] L. Wang and F. Xiao, “Finite-time consensus problems for networks of dynamic agents,” *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 950–955, 2010.

[18] J. M. Hendrickx, G. Shi, and K. H. Johansson, “Finite-time consensus using stochastic matrices with positive diagonals,” *IEEE Trans. Autom. Control*, vol. 60, no. 4, pp. 1070–1073, 2015.

[19] P. Lin and W. Ren, “Distributed shortest distance consensus problem in multi-agent systems,” in *Proc. IEEE Conf. Decision Control*, Maui, HI, Dec. 2012, pp. 4696–4701.

[20] P. Lin, W. Ren, Y. Song, and J. A. Farrell, “Distributed optimization with the consideration of adaptivity and finite-time convergence,” in *Proc. Amer. Control Conf.*, Portland, OR, Jun. 2014, pp. 3177–3182.

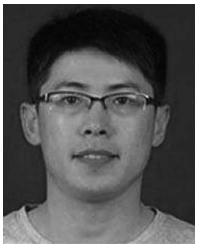
[21] F. Chen, Y. Cao, and W. Ren, “Distributed average tracking of multiple time-varying reference signals with bounded derivatives,” *IEEE Trans. Autom. Control*, vol. 57, no. 12, pp. 3169–3174, 2012.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*, London, U.K., Cambridge University Press, 2004.

[23] F. Facchinei and J. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, New York, Springer-Verlag, 2003.

[24] A. F. Filippov, *Differential Equations With Discontinuous Righthand Sides*. Amsterdam, The Netherlands: Kluwer Academic, 1988.

[25] J. Cortés, "Finite-time convergent gradient flows with applications to network consensus," *Automatica*, vol. 42, no. 11, pp. 1993–2000, 2006.



**Peng Lin** received the B.S. degree in automation from Xidian University, Xi'an, China, in 2004, and the Ph.D. degree in control theory and applications from Beihang University, Beijing, China, in 2010.

He is currently a Professor with the School of Information Science and Engineering, Central South University, Changsha 410083, China. His research interests include cooperative control, optimization theory, switching control and robust control.

**Wei Ren** (F'16) received the Ph.D. degree in electrical engineering from Brigham Young University, Provo, UT, in 2004.

He is currently a Professor with the Department of Electrical and Computer Engineering, University of California, Riverside. From 2004 to 2005, he was a Postdoctoral Research Associate with the Department of Aerospace Engineering, University of Maryland, College Park. He was an Assistant Professor (2005–2010) and an Associate Professor (2010–2011) with the Department of Electrical and Computer Engineering, Utah State University. He is an author of two books, Distributed Coordination of Multi-agent Networks (Springer-Verlag, 2011) and Distributed Consensus in Multi-vehicle Cooperative Control (Springer-Verlag, 2008). His research focuses on distributed control of multi-agent systems and autonomous control of unmanned vehicles.

Dr. Ren received the National Science Foundation CAREER Award in 2008. He is currently an Associate Editor for Automatica, Systems and Control Letters, and IEEE Transactions on Control of Network Systems.

**Jay A. Farrell** (F'08) received B.S. degrees in physics and electrical engineering from Iowa State University, and the M.S. and Ph.D. degrees in electrical engineering from the University of Notre Dame, Notre Dame, IN.

He is a Professor and two time Chair of the Department of Electrical and Computer Engineering at the University of California, Riverside. and author of over 200 technical publications. He is author of the book "Aided Navigation: GPS with High Rate Sensors" (McGraw-Hill 2008). He is also co-author of the books "The Global Positioning System and Inertial Navigation" (McGraw-Hill, 1998) and "Adaptive Approximation Based Control: Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches" (Wiley, 2006).

Dr. Farrell received the Engineering Vice President's Best Technical Publication Award in 1990, and Recognition Awards for Outstanding Performance and Achievement in 1991 and 1993. He was named a GNSS Leader to Watch for 2009–2010 by GPS World Magazine in May 2009 and a winner of the Connected Vehicle Technology Challenge by the U.S. Department of Transportation's (DOT's) Research and Innovative Technology Administration in July 2011. He is a Fellow of AAAS, a Distinguished Member of IEEE CSS. He has served the IEEE Control Systems Society (CSS) as Finance Chair for three IEEE CDC's (1995, 2001, and 2003), on the Board of Governors for two terms (2003-2006, 2012-2014), as Vice President Finance and Vice President of Technical Activities, as CSS General Vice Chair of IEEE CDC-ECC 2011, as General Chair of IEEE CDC 2012, and as President in 2014.