Distributed continuous-time algorithms for nonsmooth extended monotropic optimization problems

Zeng, Xianlin; Yi, Peng; Hong, Yiguang; Xie, Lihua

2018

Zeng, X., Yi, P., Hong, Y., & Xie, L. (2018). Distributed Continuous-Time Algorithms for Nonsmooth Extended Monotropic Optimization Problems. *SIAM Journal on Control and Optimization*, 56(6), 3973-3993. doi:10.1137/17M1118609

<https://hdl.handle.net/10356/86220>

<https://doi.org/10.1137/17M1118609>

© 2018 SIAM. All rights reserved. This paper was published in *SIAM Journal on Control and Optimization* and is made available with permission of SIAM.

Downloaded on 26 Apr 2023 21:43:17 SGT

DISTRIBUTED CONTINUOUS-TIME ALGORITHMS FOR NONSMOOTH EXTENDED MONOTROPIC OPTIMIZATION PROBLEMS*

XIANLIN ZENG[†], PENG YI[‡], YIGUANG HONG[§], AND LIHUA XIE[¶]

Abstract. This paper studies distributed algorithms for the nonsmooth extended monotropic optimization problem, which is a general convex optimization problem with a certain separable structure. The considered nonsmooth objective function is the sum of local objective functions assigned to agents in a multiagent network, with local set constraints and affine equality constraints. Each agent only knows its local objective function, local set constraint, and the information exchanged between neighbors. To solve the constrained convex optimization problem, we propose two novel distributed continuous-time subgradient-based algorithms, with projected output feedback and derivative feedback, respectively. Moreover, we prove the convergence of proposed algorithms to the optimal solutions under some mild conditions and analyze convergence rates, with the help of the techniques of variational inequalities, decomposition methods, and differential inclusions. Finally, we give an example to illustrate the efficacy of the proposed algorithms.

Key words. extended monotropic optimization, distributed algorithms, nonsmooth convex functions, decomposition methods, differential inclusions

AMS subject classifications. 68W15, 34A60, 90C25, 90C06

DOI. 10.1137/17M1118609

1. Introduction. Distributed (convex) optimization problems are important and studied in a variety of fields, including sensor networks, robot teams, and power systems (see [17, 24, 28, 39, 40, 41] and references therein). Such problems are often formulated in terms of a global objective (or cost) function in the form of a sum of individual objective functions, each of which represents the contribution of an agent to the global objective function. Various distributed approaches for efficiently solving convex optimization problems in multiagent networks have been proposed since these distributed algorithms have advantages over centralized ones in large-scale optimization situations [17, 21, 26, 28, 35, 42].

Continuous-time optimization algorithms are increasingly popular due to the well-developed theories of differential equations or inclusions and practical situations when optimization indexes have to be achieved by continuous-time physical systems. Recently, the continuous-time design for distributed optimization has been carried out for optimization problems in smart grids or mechanical systems (see [17, 39, 43]).

*Received by the editors February 27, 2017; accepted for publication (in revised form) September 6, 2018; published electronically November 1, 2018.

<http://www.siam.org/journals/sicon/56-6/M111860.html>

Funding: This work was funded by the NSFC (61603378, 61733018, and 61573344), the National Key Research and Development Program of China (2016YFB0901900), and the China Postdoctoral Science Foundation (2017M620020, 2018T110051).

[†]Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, 100081, Beijing, China (xianlin.zeng@bit.edu.cn).

[‡]Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130 (yipeng@wustl.edu).

[§]Key Laboratory of Systems and Control, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, 100190, China (yghong@iss.ac.cn).

[¶]School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798 (ELHXIE@ntu.edu.sg).

A challenging problem in optimization theory is how to deal with nonsmoothness in optimization problems. In fact, practical problems may naturally be related to nonsmooth objective functions or optimization constraints, for example, the distance function minimization and sparse solution of optimization problems; moreover, the nonsmooth technique may facilitate optimization design or improve performance in some cases, for example, with the introduction of exact penalty functions or other nonsmooth techniques. In the study of distributed optimization, nonsmooth objective functions with various constraints have been investigated from various viewpoints (e.g., in [17, 24, 28, 42]). Subgradients and projections have been widely found for distributed nonsmooth optimization problems with constraints [23, 28, 38, 42], and moreover, novel distributed algorithms have been proposed in the form of differential inclusions to solve such problems in the literature (see [23, 38, 42]).

The monotropic optimization is an important problem with separable and convex objective functions, affine equality constraints, and set constraints. As a generalization of linear programming and network programming, it was first introduced and extensively studied by [31, 32]. The extended monotropic optimization (EMO) problem was then studied in [4], but not in a distributed manner. In fact, in many applications such as wireless communication, sensor networks, neural computation, and networked robotics, the optimization problems can be converted to EMO problems [4]. Therefore, distributed algorithms and decomposition methods for EMO problems become more and more important, as pointed out in [9]. However, very few distributed optimization algorithms for EMO were proposed, due to the complexity of EMO problems.

This paper aims to investigate distributed EMO problems with nonsmooth objective functions. Based on modified Lagrangian functions and decomposition methods, our distributed projection-based algorithms are given to achieve the optimal solutions of the problems. The analysis of the proposed algorithms is carried out by virtue of differential inclusions to tackle nonsmooth objective functions. The main technical contributions of the paper can be summarized as follows. First, the distributed EMO problem is considered, and two distributed continuous-time algorithms are proposed, with projected output feedback and derivative feedback, respectively, to deal with the nonsmoothness and set constraints of the considered EMO problem. Second, based on the Lagrangian function method along with projections, the proposed algorithms are proved to have bounded states and solve the distributed EMO problems with any initial condition, with the help of the stability theory of differential inclusions and nonsmooth analysis techniques. Third, convergence rates of proposed algorithms are analyzed.

The rest of the paper is organized as follows. Section 2 introduces the preliminary knowledge related to graph theory, nonsmooth analysis, convex optimization, and projection operators. Next, section 3 formulates a class of distributed EMO problems with nonsmooth objective functions, while section 4 proposes two distributed algorithms based on projected output feedback and derivative feedback. Then section 5 shows the convergence of the proposed algorithms with nonsmooth analysis and estimates convergence rates of proposed algorithms. Following that, section 6 gives a numerical simulation to illustrate the theoretical results. Finally, section 7 presents some concluding remarks.

2. Mathematical preliminaries. In this section, we introduce relevant notation, concepts, and preliminaries on graph theory, differential inclusions, convex analysis, and projection operators.

2.1. Notation. $\mathbb{R}$ denotes the set of real numbers; $\mathbb{R}^n$ denotes the set of n -dimensional real column vectors; $\mathbb{R}^{n \times m}$ denotes the set of n -by- m real matrices; I_n denotes the $n \times n$ identity matrix; and $(\cdot)^T$ denotes the transpose. Denote $\text{rank } A$ as the rank of the matrix A , $\text{range}(A)$ as the range of A , $\ker(A)$ as the kernel of A , $\text{diag}\{A_1, \dots, A_n\}$ as the block diagonal matrix of $A_1, \dots, A_n$, 1_n ($1_{n \times q}$) as the $n \times 1$ vector ($n \times q$ matrix) with all elements of 1, 0_n ($0_{n \times q}$) as the $n \times 1$ vector ($n \times q$ matrix) with all elements of 0, and $A \otimes B$ as the Kronecker product of matrices A and B . $A > 0$ ($A \geq 0$) means that matrix $A \in \mathbb{R}^{n \times n}$ is positive definite (positive semidefinite). Furthermore, $\|\cdot\|$ stands for the Euclidean norm; $\overline{\mathcal{S}}$ ($\mathcal{S}^\circ$) stands for the closure (interior) of the subset $\mathcal{S} \subset \mathbb{R}^n$; $\mathcal{B}_\epsilon(x)$, $x \in \mathbb{R}^n$, $\epsilon > 0$, stands for the open ball centered at x with radius ϵ . $\text{dist}(p, \mathcal{M})$ denotes the distance from a point p to the set $\mathcal{M}$ (that is, $\text{dist}(p, \mathcal{M}) \triangleq \inf_{x \in \mathcal{M}} \|p - x\|$), and $x(t) \rightarrow \mathcal{M}$ as $t \rightarrow \infty$ denotes that $x(t)$ approaches the set $\mathcal{M}$ (that is, for each $\epsilon > 0$, there is $T > 0$ such that $\text{dist}(x(t), \mathcal{M}) < \epsilon$ for all $t > T$).

2.2. Graph theory. A weighted undirected graph is described by $\mathcal{G}(\mathcal{V}, \mathcal{E}, A)$ or $\mathcal{G}$, where $\mathcal{V} = \{1, \dots, n\}$ is the set of nodes, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges, and $A = [a_{i,j}] \in \mathbb{R}^{n \times n}$ is the *weighted adjacency matrix* such that $a_{i,j} = a_{j,i} > 0$ if $\{j, i\} \in \mathcal{E}$ and $a_{i,j} = 0$ otherwise. The *weighted Laplacian matrix* is $L_n = D - A$, where $D \in \mathbb{R}^{n \times n}$ is diagonal with $D_{i,i} = \sum_{j=1}^n a_{i,j}$, $i \in \{1, \dots, n\}$. In this paper, we call L_n the Laplacian matrix and A the adjacency matrix of $\mathcal{G}$ for convenience when there is no confusion. Specifically, if the weighted undirected graph $\mathcal{G}$ is connected, then $L_n = L_n^T \geq 0$, $\text{rank } L_n = n - 1$, and $\ker(L_n) = \{k1_n : k \in \mathbb{R}\}$ [18].

2.3. Differential inclusion. Following [2], a differential inclusion is given by

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

where $\mathcal{F}$ is a set-valued map from points in $\mathbb{R}^q$ to the nonempty, compact, convex subsets of $\mathbb{R}^q$. For each state $x \in \mathbb{R}^q$, system (1) specifies a *set* of possible evolutions rather than a single one. A solution of (1) defined on $[0, \tau] \subset [0, \infty)$ is an absolutely continuous function $x : [0, \tau] \rightarrow \mathbb{R}^q$ such that (1) holds for almost all $t \in [0, \tau]$ for $\tau > 0$. The solution $t \mapsto x(t)$ to (1) is a *right maximal solution* if it cannot be extended forward in time. Suppose that all right maximal solutions to (1) exist on $[0, \infty)$. A set $\mathcal{M}$ is said to be *weakly invariant* (resp., *strongly invariant*) with respect to (1) if, for every $x_0 \in \mathcal{M}$, $\mathcal{M}$ contains a maximal solution (resp., all maximal solutions) of (1). A point $z \in \mathbb{R}^q$ is a *positive limit point* of a solution $\phi(t)$ to (1) with $\phi(0) = x_0 \in \mathbb{R}^q$ if there exists a sequence $\{t_k\}_{k=1}^\infty$ with $t_k \rightarrow \infty$ and $\phi(t_k) \rightarrow z$ as $k \rightarrow \infty$. The set $\omega(\phi(\cdot))$ of all such positive limit points is the *positive limit set* for the trajectory $\phi(t)$ with $\phi(0) = x_0 \in \mathbb{R}^q$.

An equilibrium point of (1) is a point $x_e \in \mathbb{R}^q$ such that $0_q \in \mathcal{F}(x_e)$. It is easy to see that x_e is an equilibrium point of (1) if and only if the constant function $x(\cdot) = x_e$ is a solution of (1). An equilibrium point $z \in \mathbb{R}^q$ of (1) is *Lyapunov stable* if, for every $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that, for every initial condition $x(0) = x_0 \in \mathcal{B}_\delta(z)$, every solution $x(t) \in \mathcal{B}_\varepsilon(z)$ for all $t \geq 0$.

Let $V : \mathbb{R}^q \rightarrow \mathbb{R}$ be a locally Lipschitz continuous function and ∂V be the *Clarke generalized gradient* [12] of $V(x)$ at x . The *set-valued Lie derivative* [12] $\mathcal{L}_{\mathcal{F}}V : \mathbb{R}^q \rightarrow \mathfrak{B}(\mathbb{R})$ of V with respect to (1) is defined as $\mathcal{L}_{\mathcal{F}}V(x) \triangleq \{a \in \mathbb{R} : \text{there exists } v \in \mathcal{F}(x) \text{ such that } p^T v = a \text{ for all } p \in \partial V(x)\}$. In the case when $\mathcal{L}_{\mathcal{F}}V(x)$ is nonempty, we use $\max \mathcal{L}_{\mathcal{F}}V(x)$ to denote the largest element of $\mathcal{L}_{\mathcal{F}}V(x)$. Recall from [3] that if $\phi(\cdot)$ is a solution of (1) and $V : \mathbb{R}^q \rightarrow \mathbb{R}$ is locally Lipschitz and *regular* (see [12, p. 39]), then $\dot{V}(\phi(t))$ exists almost everywhere, and $\dot{V}(\phi(t)) \in \mathcal{L}_{\mathcal{F}}V(\phi(t))$ almost everywhere.

Next, we introduce a version of the invariance principle (Theorem 2 of [13]), which is based on nonsmooth regular functions.

LEMMA 2.1. *For the differential inclusion (1), we assume that $\mathcal{F}$ is upper semi-continuous and locally bounded, and $\mathcal{F}(x)$ takes nonempty, compact, and convex values. Let $V : \mathbb{R}^q \rightarrow \mathbb{R}$ be a locally Lipschitz and regular function, $\mathcal{S} \subset \mathbb{R}^q$ be compact and strongly invariant for (1), $\phi(\cdot)$ be a solution of (1),*

$$\mathcal{R} = \{x \in \mathbb{R}^q : 0 \in \mathcal{L}_{\mathcal{F}}V(x)\},$$

and $\mathcal{M}$ be the largest weakly invariant subset of $\overline{\mathcal{R}} \cap \mathcal{S}$, where $\overline{\mathcal{R}}$ is the closure of $\mathcal{R}$. If $\max \mathcal{L}_{\mathcal{F}}V(x) \leq 0$ for all $x \in \mathcal{S}$, then $\text{dist}(\phi(t), \mathcal{M}) \rightarrow 0$ as $t \rightarrow +\infty$.

2.4. Convex analysis. A function $\psi : \mathbb{R}^q \rightarrow \mathbb{R}$ is *convex* if $\psi(\lambda x + (1 - \lambda)y) \leq \lambda\psi(x) + (1 - \lambda)\psi(y)$ for all $x, y \in \mathbb{R}^q$ and $\lambda \in [0, 1]$. A function $\psi : \mathbb{R}^q \rightarrow \mathbb{R}$ is *strictly convex* whenever $\psi(\lambda x + (1 - \lambda)y) < \lambda\psi(x) + (1 - \lambda)\psi(y)$ for all $x, y \in \mathbb{R}^q$, $x \neq y$, and $\lambda \in (0, 1)$. Let $\psi : \mathbb{R}^q \rightarrow \mathbb{R}$ be a convex function. The *subdifferential* [14, p. 544] $\partial_{\text{sub}}\psi$ of ψ at $x \in \mathbb{R}^q$ is defined by $\partial_{\text{sub}}\psi(x) \triangleq \{p \in \mathbb{R}^q : \langle p, y - x \rangle \leq \psi(y) - \psi(x) \text{ for all } y \in \mathbb{R}^q\}$, and the elements of $\partial_{\text{sub}}\psi(x)$ are called *subgradients* of ψ at point x . Recall from [14, p. 607] that continuous convex functions are locally Lipschitz continuous and regular, and their subdifferentials and Clarke generalized gradients coincide. Thus, the framework for stability theory of differential inclusions can be applied to the theoretical analysis in this paper.

The following result can be easily verified by the property of strictly convex functions.

LEMMA 2.2. *If $f : \mathbb{R}^q \rightarrow \mathbb{R}$ is a continuous strictly convex function, then*

$$(2) \quad (g_x - g_y)^T(x - y) > 0$$

for all $x \neq y$, where $g_x \in \partial f(x)$ and $g_y \in \partial f(y)$.

2.5. Projection operator. Define $P_{\Omega}(\cdot)$ as a projection operator given by $P_{\Omega}(u) = \arg \min_{v \in \Omega} \|u - v\|$, where $\Omega \subset \mathbb{R}^n$ is closed and convex. A basic property [20] of a projection $P_{\Omega}(\cdot)$ on a closed convex set $\Omega \subset \mathbb{R}^n$ is

$$(3) \quad (u - P_{\Omega}(u))^T(v - P_{\Omega}(u)) \leq 0 \quad \forall u \in \mathbb{R}^n, \quad \forall v \in \Omega.$$

Using (3), the following results can be easily verified.

LEMMA 2.3 [16, Theorem 1.5.5, p. 77]. *If $\Omega \subset \mathbb{R}^n$ is closed and convex, then $(P_{\Omega}(x) - P_{\Omega}(y))^T(x - y) \geq \|P_{\Omega}(x) - P_{\Omega}(y)\|^2$ for all $x, y \in \mathbb{R}^n$.*

LEMMA 2.4 [23, Lemma 4]. *Let $\Omega \subset \mathbb{R}^n$ be closed and convex, and define $V : \mathbb{R}^n \rightarrow \mathbb{R}$ as $V(x) = \frac{1}{2}(\|x - P_{\Omega}(y)\|^2 - \|x - P_{\Omega}(x)\|^2)$, where $y \in \mathbb{R}^n$. Then $V(x) \geq \frac{1}{2}\|P_{\Omega}(x) - P_{\Omega}(y)\|^2$, $V(x)$ is differentiable and convex with respect to x , and $\nabla V(x) = P_{\Omega}(x) - P_{\Omega}(y)$.*

3. Problem description. In this section, we present the distributed EMO problem with nonsmooth objective functions and give the optimality condition for the problem.

Consider a network of n agents interacting over a graph $\mathcal{G}$. There are a local objective function $f^i : \Omega_i \rightarrow \mathbb{R}$ and a local feasible constraint set $\Omega_i \subset \mathbb{R}^{q_i}$ for all $i \in \{1, \dots, n\}$. Let $x_i \in \Omega_i \subset \mathbb{R}^{q_i}$ and denote $x \triangleq [x_1^T, \dots, x_n^T]^T \in \Omega \triangleq \prod_{i=1}^n \Omega_i \subset \mathbb{R}^{\sum_{i=1}^n q_i}$. The global objective function of the network is $f(x) = \sum_{i=1}^n f^i(x_i)$, $x \in \Omega \subset \mathbb{R}^{\sum_{i=1}^n q_i}$.

Here we formulate the distributed EMO problem as follows:

$$(4a) \quad \min f(x), \quad f(x) = \sum_{i=1}^n f^i(x_i),$$

$$(4b) \quad Wx = \sum_{i=1}^n W_i x_i = d_0, \quad x_i \in \Omega_i \subset \mathbb{R}^{q_i}, \quad i \in \{1, \dots, n\},$$

where $W_i \in \mathbb{R}^{m \times q_i}$, $i \in \{1, \dots, n\}$, and $W = [W_1, \dots, W_n] \in \mathbb{R}^{m \times \sum_{i=1}^n q_i}$. In this problem, agent i has its state $x_i \in \Omega_i \subset \mathbb{R}^{q_i}$, objective function $f_i(x_i)$, set constraint $\Omega_i \subset \mathbb{R}^{q_i}$, constraint matrix $W_i \in \mathbb{R}^{m \times q_i}$, and information from neighboring agents.

The goal of the distributed EMO is to solve the problem in a distributed manner. In a distributed optimization algorithm, each agent in the graph $\mathcal{G}$ only uses its own local cost function, its local set constraint, the decomposed information of the global equality constraint, and the shared information of its neighbors through constant local communications. The special case of problem (4), where each component x_i is one-dimensional (that is, $q_i = 1$), is called the monotropic programming problem and has been introduced and studied extensively in [31, 32].

Remark 3.1. The distributed EMO problem (4) covers many problems in the areas of optimization and machine learning because of the general expression. For example, it generalizes the optimization model in resource allocation problems [39, 40] by allowing nonsmooth objective functions and a more general equality constraint. It also covers the model proposed in [25] and generalizes the model in the distributed constrained optimal consensus problem [30] by allowing heterogeneous constraints. Furthermore, this problem may be viewed as a different formulation of distributed computing linear algebraic equations widely investigated in [1, 27, 34, 44].

For illustration, we show concrete application examples where the constraint (4b) arises.

- Consider the resource allocation problem in power grids of the form (4). In (4b), $d_0 > 0$ is the total amount of energy to be generated, $x_i \in \Omega_i \subset \mathbb{R}^{q_i}$ is a vector of p_i different resources allocated to agent i , Ω_i is the feasible set for agent i , and $W_i \in \mathbb{R}^{1 \times p_i}$ is the vector of energy generated by per unit of resources.
- Consider the minimum (l_1) norm solution of underdetermined linear equation $Wx = b$, $x \in \Omega = \prod_{i=1}^n \Omega_i$ (see [44]), which is an important problem in signal processing [15]. In (4b), the information of $W = [W_1, \dots, W_n]$ and $x \triangleq [x_1^T, \dots, x_n^T]^T$ is distributed such that agent i knows W_i and solves $x_i \in \Omega_i$ for $i \in \{1, \dots, n\}$.

Remark 3.2. This paper considers nonsmooth objective functions in the problem formulation, which frequently occur in engineering and science problems. For example, the objective functions of resource allocation in power grids may be nonsmooth (see [11, 29, 37]); finding the sparse solutions of linear algebraic equations (see [15, 44]) in signal processing naturally arrives at nonsmooth optimization problems; the LASSO problems and compressed sensing in machine learning and data science (see [8]) involve nonsmooth norms.

To ensure the well-posedness of the problem, the following assumption for problem (4) is needed, which is quite standard.

Assumption 3.3.

1. The weighted graph $\mathcal{G}$ is connected and undirected.
2. For all $i \in \{1, \dots, n\}$, f^i is strictly convex on an open set containing Ω_i , and $\Omega_i \subset \mathbb{R}^{q_i}$ is closed and convex.
3. (Slater's constraint condition) There exists $x \in \Omega^\circ$ satisfying the constraint $Wx = d_0$, where Ω° is the interior of Ω .

LEMMA 3.4. *Under Assumption 3.3, $x^* \in \Omega$ is an optimal solution of (4) if and only if there exist $\lambda_0^* \in \mathbb{R}^m$ and $g(x^*) \in \partial f(x^*)$ such that*

$$(5) \quad x^* = P_\Omega(x^* - g(x^*) + W^T \lambda_0^*),$$

$$(6) \quad Wx^* = d_0.$$

Proof. Consider problem (4). By the KKT optimality condition (Theorem 3.34 of [33]), $x^* \in \Omega$ is an optimal solution of (4) if and only if there exist $\lambda_0^* \in \mathbb{R}^m$ and $g(x^*) \in \partial f(x^*)$ such that (6) holds and

$$(7) \quad -g(x^*) + W^T \lambda_0^* \in \mathcal{N}_\Omega(x^*),$$

where $\mathcal{N}_\Omega(x^*)$ is the normal cone of Ω at an element $x^* \in \Omega$. Note that (7) holds if and only if (5) holds. Thus, the proof is completed. $\square$

4. Optimization algorithms. In this section, we propose two distributed optimization algorithms to solve the EMO problem with nonsmooth objective functions. To our best knowledge, there are no distributed algorithms for such problems.

The resource allocation problem, a special case of the EMO problem, was studied for problems with smooth objective functions in [40]. An intuitive extension of the continuous-time algorithm given in [40] to nonsmooth EMO cases may be written as

$$(8) \quad \begin{cases} \dot{x}_i(t) & \in \left\{ p : p = P_{\Omega_i}[x_i(t) - g_i(x_i(t)) + W_i^T \lambda_i(t)] - x_i(t), \right. \\ & \left. g_i(x_i(t)) \in \partial f^i(x_i(t)) \right\}, \\ \dot{\lambda}_i(t) & = d_i - W_i x_i(t) - \sum_{j=1}^n a_{i,j}(\lambda_i(t) - \lambda_j(t)) \\ & \quad - \sum_{j=1}^n a_{i,j}(z_i(t) - z_j(t)), \\ \dot{z}_i(t) & = \sum_{j=1}^n a_{i,j}(\lambda_i(t) - \lambda_j(t)), \end{cases}$$

where $t \geq 0$, $i \in \{1, \dots, n\}$, $x_i(0) = x_{i0} \in \Omega_i \subset \mathbb{R}^{q_i}$, $\lambda_i(0) = \lambda_{i0} \in \mathbb{R}^m$, $z_i(0) = z_{i0} \in \mathbb{R}^m$, $\sum_{i=1}^n d_i = d_0$, and $a_{i,j}$ is the (i, j) th element of the adjacency matrix of graph $\mathcal{G}$.

However, this algorithm involves the projection of a subdifferential set (from $\partial f^i(x_i)$ to Ω_i) and may be a nonconvex differential inclusion. As a result, the existence of trajectories to (8) is not guaranteed due to the nonconvexity of the algorithm and the convergence analysis of the algorithm is very hard because of the nonsmoothness. To overcome the technical challenges, we propose two different ideas to construct effective algorithms for the EMO problem in the following two subsections.

4.1. Distributed projected output feedback algorithm. The first idea is to use an auxiliary variable to avoid the projection of a subdifferential set in the algorithm for EMO problem (4). In other words, we propose a distributed algorithm based on projected output feedbacks, and the projected output feedback of the auxiliary variable is adopted to track the optimal solution. To be strict, we propose the continuous-time algorithm of agent i as follows:

$$(9) \quad \begin{cases} \dot{y}_i(t) & \in \left\{ p : p = -y_i(t) + x_i(t) - g_i(x_i(t)) + W_i^T \lambda_i(t), \right. \\ & \left. g_i(x_i(t)) \in \partial f^i(x_i(t)) \right\}, \\ \dot{\lambda}_i(t) & = d_i - W_i x_i(t) - \sum_{j=1}^n a_{i,j} (\lambda_i(t) - \lambda_j(t)) \\ & \quad - \sum_{j=1}^n a_{i,j} (z_i(t) - z_j(t)), \\ \dot{z}_i(t) & = \sum_{j=1}^n a_{i,j} (\lambda_i(t) - \lambda_j(t)), \\ x_i(t) & = P_{\Omega_i}(y_i(t)), \end{cases}$$

with the auxiliary variable $y_i(t)$ for $t \geq 0$, $y_i(0) = y_{i0} \in \mathbb{R}^{q_i}$, $i \in \{1, \dots, n\}$, and other notation is kept the same as that for (8). The term $x_i(t) = P_{\Omega_i}(y_i(t))$ is viewed as “projected output feedback,” which is inspired by [23]. In this way, we avoid the technical difficulties resulting from the projection of the subdifferential.

Let $x \triangleq [x_1^T, \dots, x_n^T]^T \in \Omega \subset \mathbb{R}^{\sum_{i=1}^n q_i}$, $y \triangleq [y_1^T, \dots, y_n^T]^T \in \mathbb{R}^{\sum_{i=1}^n q_i}$, $\lambda \triangleq [\lambda_1^T, \dots, \lambda_n^T]^T \in \mathbb{R}^{nm}$, $d \triangleq [d_1^T, \dots, d_n^T]^T \in \mathbb{R}^{nm}$, and $z \triangleq [z_1^T, \dots, z_n^T]^T \in \mathbb{R}^{nm}$, where $\Omega \triangleq \prod_{i=1}^n \Omega_i$. Let $W = [W_1, \dots, W_n] \in \mathbb{R}^{m \times \sum_{i=1}^n q_i}$ and $\bar{W} = \text{diag}\{W_1, \dots, W_n\} \in \mathbb{R}^{nm \times \sum_{i=1}^n q_i}$. Define the modified Lagrangian function $\hat{L} : \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ as

$$(10) \quad \hat{L}(x, z, \lambda) = f(x) + \lambda^T (d - \bar{W}x - Lz) - \frac{1}{2} \lambda^T L \lambda,$$

where $L = L_n \otimes I_m \in \mathbb{R}^{nm \times nm}$ and $L_n \in \mathbb{R}^{n \times n}$ is the Laplacian matrix of graph $\mathcal{G}$.

Then (9) can be written in a compact form:

$$(11) \quad \begin{bmatrix} \dot{y}(t) \\ \dot{\lambda}(t) \\ \dot{z}(t) \end{bmatrix} \in \mathcal{F}(y(t), \lambda(t), z(t)),$$

$$(12) \quad x(t) = P_{\Omega}(y(t)),$$

$$(13) \quad \mathcal{F}(y, \lambda, z) \triangleq \left\{ \begin{bmatrix} -y + x - p_x \\ \nabla_{\lambda} \hat{L}(x, z, \lambda) \\ -\nabla_z \hat{L}(x, z, \lambda) \end{bmatrix} : p_x \in \partial_x \hat{L}(x, z, \lambda), x = P_{\Omega}(y) \right\},$$

where $y(0) = y_0 \in \mathbb{R}^{\sum_{i=1}^n q_i}$, $\lambda(0) = \lambda_0 \in \mathbb{R}^{nm}$, $z(0) = z_0 \in \mathbb{R}^{nm}$, and function $\hat{L}(\cdot, \cdot, \cdot)$ is defined in (10).

Remark 4.1. In this algorithm, $x(t) = P_{\Omega}(y(t))$ is used to estimate the optimal solution of the EMO problem. Moreover, $x(t)$ stays in the constraint set Ω for every $t \geq 0$, though $y(t)$ may be out of Ω . Later, we will show that this algorithm can solve the EMO problem with nonsmooth objective functions and private constraints.

4.2. Distributed derivative feedback algorithm. The second idea to facilitate the convergence analysis is to make a copy of the projection term by using the derivative feedback so as to eliminate the “trouble” caused by the projection term somehow. To be strict, we propose the algorithm for the EMO problem (4) as follows:

$$(14) \quad \begin{cases} \dot{x}_i(t) & \in \left\{ p : p = P_{\Omega_i}[x_i(t) - g_i(x_i(t)) + W_i^T \lambda_i(t)] - x_i(t), \right. \\ & \left. g_i(x_i(t)) \in \partial f^i(x_i(t)) \right\}, \\ \dot{\lambda}_i(t) & = d_i - W_i x_i(t) - \sum_{j=1}^n a_{i,j} (\lambda_i(t) - \lambda_j(t)) \\ & \quad - \sum_{j=1}^n a_{i,j} (z_i(t) - z_j(t)) - W_i \dot{x}_i(t), \\ \dot{z}_i(t) & = \sum_{j=1}^n a_{i,j} (\lambda_i(t) - \lambda_j(t)), \end{cases}$$

where all the notation remains the same as in (8). Note that there is a derivative term $\dot{x}_i(t)$, viewed as “derivative feedback,” on the right-hand side of the second equation. The derivative term can be found to be effective in the cancellation of the “trouble” term $P_{\Omega_i}[x_i(t) - g_i(x_i(t)) + W_i^T \lambda_i(t)]$ in the analysis as demonstrated later.

Denote x , λ , d , z , W , and $\bar{W}$ as in (13). Then (14) can be written in a compact form,

$$(15) \quad \begin{bmatrix} \dot{x}(t) \\ \dot{\lambda}(t) \\ \dot{z}(t) \end{bmatrix} \in \mathcal{F}(x(t), \lambda(t), z(t)),$$

$$(16) \quad \mathcal{F}(x, \lambda, z) \triangleq \left\{ \begin{bmatrix} p \\ \nabla_{\lambda} \hat{L}(x + p, z, \lambda) \\ -\nabla_z \hat{L}(x, z, \lambda) \end{bmatrix} : p \in P_{\Omega}[x - \partial_x \hat{L}(x, z, \lambda)] - x \right\},$$

where $x(0) = x_0 \in \Omega$, $\lambda(0) = \lambda_0 \in \mathbb{R}^{nm}$, $z(0) = z_0 \in \mathbb{R}^{nm}$, and function $\hat{L}(\cdot, \cdot, \cdot)$ is defined in (10).

4.3. Algorithm comparison. In addition to the different techniques used in the design of algorithms (9) and (14), the proposed algorithms are also different in the following aspects:

- The application situations may be different. Although the convergence of both algorithms is based on the strict convexity assumption of the objective functions, algorithm (14) can also solve the EMO problem with only convex objective functions (which may have a continuum set of optimal solutions) when the objective functions are differentiable (see Corollary 5.12 in section 5).
- The dynamic performances may be different. Because algorithm (14) directly changes $x(t) \in \Omega$ to estimate the optimal solution, it may show a faster response speed of $x(t)$ than that of algorithm (9) (see simulation results in section 6).

Furthermore, algorithms (9) and (14) are essentially different from existing ones. Compared with the algorithm in [10], our algorithms need not exchange information of subgradients among the agents. Unlike the algorithm given in [39], ours use different techniques (i.e., the projected output feedback in (9) or derivative feedback in (14)) to estimate the optimal solution. Moreover, our algorithms have two advantages compared with previous methods.

- Agent i of the proposed algorithms knows W_i , which is composed of a subset of columns in W . This is different from some existing results with assuming that each agent knows a subset of rows of the equality constraints [1, 27, 34, 44]. If n is a sufficiently large number and m is relatively small, the proposed designs make the computation load at each node relatively small compared with previous algorithms in [1, 27, 34, 44].
- Agent i in the proposed algorithms exchanges information of $\lambda_i \in \mathbb{R}^m$ and $z_i \in \mathbb{R}^m$ with its neighbors. Compared with algorithms which require exchanging $x_i \in \mathbb{R}^{q_i}$, this design can greatly reduce the communication cost when q_i is much larger than m for $i \in \{1, \dots, n\}$ and the information of $x_i \in \mathbb{R}^{q_i}$ is kept confidential.

Remark 4.2. The results proposed in this paper may also be applied to the problem in [44] with a different formulation. Specifically, each agent knows a few rows of matrix and vector information in the formulation of [44], while each agent knows a

few columns of the matrix information in the formulation of this paper. Additionally, the algorithm in [44] uses sign functions and the sign consensus design to achieve the finite-time convergence.

5. Convergence analysis. In this section, we use the stability analysis of differential inclusions to prove the correctness and convergence of our proposed algorithms.

5.1. Convergence analysis of the distributed projected output feedback algorithm. Consider algorithm (9) (or (11)). Let $(y^*, \lambda^*, z^*) \in \mathbb{R}^{\sum_{i=1}^n q_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ be an equilibrium of (9). Then

$$(17a) \quad 0_{\sum_{i=1}^n q_i} \in \{p : p = -y^* + x^* - g(x^*) + \bar{W}^T \lambda^*, x^* = P_\Omega(y^*), g(x^*) \in \partial f(x^*)\},$$

$$(17b) \quad 0_{nm} = d - \bar{W}x^* - Lz^*, \quad x^* = P_\Omega(y^*),$$

$$(17c) \quad 0_{nm} = L\lambda^*.$$

The following result reveals the relationship between the equilibrium points of algorithm (9) and the solutions of problem (4).

THEOREM 5.1. *Suppose Assumption 3.3 holds. If $(y^*, \lambda^*, z^*) \in \mathbb{R}^{\sum_{i=1}^n q_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ is an equilibrium of (9), then $x^* = P_\Omega(y^*)$ is a solution to problem (4). Conversely, if $x^* \in \Omega$ is a solution to problem (4), then there exists $(y^*, \lambda^*, z^*) \in \mathbb{R}^{\sum_{i=1}^n q_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ such that (y^*, λ^*, z^*) is an equilibrium of (9) with $x^* = P_\Omega(y^*)$.*

Proof. (i) Suppose $(y^*, \lambda^*, z^*) \in \mathbb{R}^{\sum_{i=1}^n q_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ is an equilibrium of (9). Left-multiply both sides of (17b) by $1_n^T \otimes I_m$. It follows that

$$(18) \quad \begin{aligned} 1_n^T \otimes I_m (d - \bar{W}x^* - Lz^*) &= \sum_{i=1}^n (d_i - W_i x_i^*) - (1_n^T \otimes I_m) Lz^* \\ &= d_0 - Wx^* - (1_n^T \otimes I_m) Lz^* = 0_m. \end{aligned}$$

It follows from the properties of the Kronecker product and $L_n^T 1_n = 0_n$ that

$$(19) \quad (1_n^T \otimes I_m) L = (1_n^T \otimes I_m) (L_n \otimes I_m) = (1_n^T L_n) \otimes (I_m) = 0_{m \times nm}.$$

In light of (18) and (19), (6) holds.

Next, it follows from (17c) that there exists $\lambda_0^* \in \mathbb{R}^m$ such that $\lambda^* = 1_n \otimes \lambda_0^*$. By taking into consideration (17a) and $\lambda^* = 1_n \otimes \lambda_0^*$, there exists $g(x^*) \in \partial f(x^*)$ such that $x^* - g(x^*) + \bar{W}^T (1_n \otimes \lambda_0^*) = x^* - g(x^*) + W^T \lambda_0^* = y^*$. Since $x^* = P_\Omega(y^*)$, it follows that (5) holds.

By virtue of (5), (6), and Lemma 3.4, x^* is the solution to problem (4).

(ii) Conversely, suppose x^* is the solution to problem (4). According to Lemma 3.4, there exist $\lambda_0^* \in \mathbb{R}^q$ and $g(x^*) \in \partial f(x^*)$ such that (5) and (6) hold. Define $\lambda^* = 1_n \otimes \lambda_0^*$. As a result, (17c) holds.

Take any $v \in \mathbb{R}^m$ and let $\mathbf{v} = 1_n \otimes v$. Since (6) holds, $(d - \bar{W}x^*)^T \mathbf{v} = (\sum_{i=1}^n (d_i - W_i x_i^*))^T v = (d_0 - Wx^*)^T v = 0$. Due to the properties of the Kronecker product and $L_n 1_n = 0_n$, $L\mathbf{v} = (L_n \otimes I_m)(1_n \otimes v) = (L_n 1_n) \otimes (I_m v) = 0_n \otimes v = 0_{nm}$, and hence, $\mathbf{v} \in \ker(L)$. Note that $\ker(L)$ and $\text{range}(L)$ form an orthogonal decomposition of $\mathbb{R}^{nm}$ by the fundamental theorem of linear algebra [36]. It follows from $(d - \bar{W}x^*)^T \mathbf{v} = 0$ and $\mathbf{v} \in \ker(L)$ that $d - \bar{W}x^* \in \text{range}(L)$. Hence, there exists $z^* \in \mathbb{R}^{nm}$ such that (17b) holds.

Because $W^T \lambda_0^* = \bar{W}^T (1_n \otimes \lambda_0^*) = \bar{W}^T \lambda^*$, (5) implies $x^* = P_\Omega(x^* - g(x^*) + \bar{W}^T (1_n \otimes \lambda_0^*)) = P_\Omega(x^* - g(x^*) + \bar{W}^T \lambda^*)$ for some $g(x^*) \in \partial f(x^*)$. Choose $y^* = x^* - g(x^*) + \bar{W}^T \lambda^*$. Then (17a) holds.

To sum up, if $x^* \in \Omega$ is a solution to problem (4), there exists $(y^*, \lambda^*, z^*) \in \mathbb{R}^{\sum_{i=1}^n q_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ such that (17) holds and $x^* = P_\Omega(y^*)$. Hence, (y^*, λ^*, z^*) is an equilibrium of (9) with $x^* = P_\Omega(y^*)$. $\square$

Remark 5.2. Following the proof of Theorem 5.1, one can easily prove that a point $(x^*, \lambda^*, z^*) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ is a saddle point of $\hat{L}$ (that is, $\hat{L}(x, z, \lambda^*) \geq \hat{L}(x^*, z^*, \lambda^*) \geq \hat{L}(x^*, z^*, \lambda)$ for all $x \in \Omega$ and $\lambda, z \in \mathbb{R}^{nm}$) if and only if x^* is a solution to problem (4).

Let x^* be the solution to problem (4). It follows from Theorem 5.1 that there exist y^*, λ^* , and z^* such that (y^*, λ^*, z^*) is an equilibrium of (9) with $x^* = P_\Omega(y^*)$. Define the function

$$(20) \quad V(y, \lambda, z) \triangleq \frac{1}{2}(\|y - P_\Omega(y^*)\|^2 - \|y - P_\Omega(y)\|^2) + \frac{1}{2}\|\lambda - \lambda^*\|^2 + \frac{1}{2}\|z - z^*\|^2.$$

LEMMA 5.3. Consider algorithm (9). Under Assumption 3.3 with $V(y, \lambda, z)$ defined in (20), if $a \in \mathcal{L}_F V(y, \lambda, z)$, then there exist $g(x) \in \partial f(x)$ and $g(x^*) \in \partial f(x^*)$ with $x = P_\Omega(y)$ and $x^* = P_\Omega(y^*)$ such that $a \leq -(x - x^*)^T(y - y^*) + \|x - x^*\|^2 - (x - x^*)^T(g(x) - g(x^*)) - \lambda^T L \lambda \leq 0$.

Proof. It follows from Lemma 2.4 that the gradient of $V(y, \lambda, z)$ with respect to y is $\nabla_y V(y, \lambda, z) = x - x^*$, where $x = P_\Omega(y)$ and $x^* = P_\Omega(y^*)$. The gradients of $V(y, \lambda, z)$ with respect to λ and z are $\nabla_\lambda V(y, \lambda, z) = \lambda - \lambda^*$ and $\nabla_z V(y, \lambda, z) = z - z^*$, respectively.

The function $V(y, \lambda, z)$ along the trajectories of (9) satisfies that

$$\begin{aligned} \mathcal{L}_F V(y, \lambda, z) = & \left\{ a \in \mathbb{R} : a = \nabla_y V(y, \lambda, z)^T(-y + x - g(x) + \bar{W}^T \lambda) \right. \\ & + \nabla_\lambda V(y, \lambda, z)^T(d - \bar{W}x - L\lambda - Lz) + \nabla_z V(y, \lambda, z)^T L\lambda, \\ & \left. g(x) \in \partial f(x), x = P_\Omega(y) \right\}. \end{aligned}$$

Suppose $a \in \mathcal{L}_F V(y, \lambda, z)$. There is $g(x) \in \partial f(x)$ such that

$$(21) \quad \begin{aligned} a = & (x - x^*)^T(-y + x - g(x) + \bar{W}^T \lambda) \\ & + (\lambda - \lambda^*)^T(d - \bar{W}x - L\lambda - Lz) + (z - z^*)^T L\lambda, \end{aligned}$$

where $x = P_\Omega(y)$.

Because (y^*, λ^*, z^*) is an equilibrium of (9) with $x^* = P_\Omega(y^*)$, there exists $g(x^*) \in \partial f(x^*)$ such that

$$(22) \quad \begin{cases} 0_{nm} & = L\lambda^* \\ d & = \bar{W}x^* + Lz^*, \\ 0_{\sum_{i=1}^n q_i} & = -y^* + x^* - g(x^*) + \bar{W}^T \lambda^*. \end{cases}$$

It follows from (21) and (22) that

$$\begin{aligned} a = & (x - x^*)^T[(-y + x - g(x) + \bar{W}^T \lambda) - (-y^* + x^* - g(x^*) + \bar{W}^T \lambda^*)] \\ & + (\lambda - \lambda^*)^T(\bar{W}x^* + Lz^* - \bar{W}x - L\lambda - Lz) + (z - z^*)^T L\lambda \\ = & -(x - x^*)^T(y - y^*) + \|x - x^*\|^2 - (x - x^*)^T(g(x) - g(x^*)) \\ & + (x - x^*)^T \bar{W}^T(\lambda - \lambda^*) - (x - x^*)^T \bar{W}^T(\lambda - \lambda^*) \\ & - \lambda^T L\lambda - \lambda^T L(z - z^*) + (z - z^*)^T L\lambda \\ (23) \quad = & -(x - x^*)^T(y - y^*) + \|x - x^*\|^2 - (x - x^*)^T(g(x) - g(x^*)) - \lambda^T L\lambda. \end{aligned}$$

Since $x = P_\Omega(y)$ and $x^* = P_\Omega(y^*)$, we obtain from Lemma 2.3 that $-(x - x^*)^T(y - y^*) + \|x - x^*\|^2 \leq 0$. The convexity of f implies that $(x - x^*)^T(g(x) - g(x^*)) \geq 0$. In addition, $L = L_n \otimes I_q \geq 0$ since $L_n \geq 0$. Hence, $a \leq -(x - x^*)^T(y - y^*) + \|x - x^*\|^2 - (x - x^*)^T(g(x) - g(x^*)) - \lambda^T L \lambda \leq 0$. $\square$

The following result shows the correctness of the proposed algorithm.

THEOREM 5.4. *Consider algorithm (9). If Assumption 3.3 holds, then*

- (i) *every solution $(y(t), x(t), \lambda(t), z(t))$ is bounded;*
- (ii) *for every solution, $x(t)$ converges to the optimal solution to problem (4).*

Proof. (i) Let $V(y, \lambda, z)$ be as defined in (20). It follows from Lemma 5.3 that

$$(24) \quad \max_{\mathcal{F}} V(y, \lambda, z) \leq \max\{-(x - x^*)^T(g(x) - g(x^*)) - \lambda^T L \lambda : g(x) \in \partial f(x)\} \leq 0.$$

Note that $V(y, \lambda, z) \geq \frac{1}{2}\|x - x^*\|^2 + \frac{1}{2}\|\lambda - \lambda^*\|^2 + \frac{1}{2}\|z - z^*\|^2$ in view of Lemma 2.4. It follows from (24) that $(x(t), \lambda(t), z(t))$, $t \geq 0$, is bounded. Because $\partial f(x)$ is compact, there exists $m = m(y_0, \lambda_0, z_0) > 0$ such that

$$(25) \quad \|x(t) - g(x(t)) + \bar{W}^T \lambda(t)\| < m$$

for all $g(x(t)) \in \partial f(x(t))$ and all $t \geq 0$.

Define $X : \mathbb{R}^{\sum_{i=1}^n q_i} \rightarrow \mathbb{R}$ by $X(y) = \frac{1}{2}\|y\|^2$. The function $X(y)$ along the trajectories of (9) satisfies that

$$\mathcal{L}_{\mathcal{F}} X(y) = \{y^T(-y + x - g(x) + \bar{W}^T \lambda) : g(x) \in \partial f(x)\}.$$

Note that $y^T(t)(-y(t) + x(t) - g(x(t)) + \lambda(t)) \leq -\|y(t)\|^2 + m\|y(t)\|$, where $t \geq 0$, m is defined by (25), and $g(x(t)) \in \partial f(x(t))$. Hence,

$$\max_{\mathcal{F}} \mathcal{L}_{\mathcal{F}} X(y(t)) \leq -\|y(t)\|^2 + m\|y(t)\| = -2X(y(t)) + m\sqrt{2X(y(t))}.$$

It can be easily verified that $X(y(t))$, $t \geq 0$, is bounded, so is $y(t)$, $t \geq 0$.

To sum up, part (i) is proved.

(ii) Let $\mathcal{R} = \{(y, \lambda, z) \in \mathbb{R}^{\sum_{i=1}^n q_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : 0 \in \mathcal{L}_{\mathcal{F}} V(y, \lambda, z)\} \subset \{(y, \lambda, z) \in \mathbb{R}^{\sum_{i=1}^n q_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : 0 = \min_{g(x) \in \partial f(x), g(x^*) \in \partial f(x^*)} (x - x^*)^T(g(x) - g(x^*)), L\lambda = 0_{nm}, x = P_\Omega(y), x^* = P_\Omega(y^*)\}$. Note that $(x - x^*)^T(g(x) - g(x^*)) > 0$ if $x \neq x^*$ because of the strict convexity assumption of f and Lemma 2.2. Hence, $\mathcal{R} \subset \{(y, \lambda, z) \in \mathbb{R}^{\sum_{i=1}^n q_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : x = P_\Omega(y) = x^*, L\lambda = 0_{nm}\}$. Let $\mathcal{M}$ be the largest weakly invariant subset of $\bar{\mathcal{R}}$. It follows from Lemma 2.1 that $(y(t), \lambda(t), z(t)) \rightarrow \mathcal{M}$ as $t \rightarrow \infty$. Hence, $x(t) \rightarrow x^*$ as $t \rightarrow \infty$. $\square$

Define the average values for $x(t)$, $\lambda(t)$, and $z(t)$ as

$$(26) \quad \hat{x}(t) \triangleq \frac{1}{t} \int_0^t x(s) ds, \quad \hat{\lambda}(t) \triangleq \frac{1}{t} \int_0^t \lambda(s) ds, \quad \hat{z}(t) \triangleq \frac{1}{t} \int_0^t z(s) ds,$$

where $(x(t), \lambda(t), z(t))$ is a trajectory generated by algorithm (9). We show the convergence rate of $(\hat{x}(t), \hat{\lambda}(t), \hat{z}(t))$ in the following result.

THEOREM 5.5. *Consider algorithm (9). If Assumption 3.3 holds, then*

$$0 \leq \hat{L}(\hat{x}(t), \hat{z}(t), \lambda^*) - \hat{L}(x^*, z^*, \hat{\lambda}(t)) \leq \frac{1}{t} V(y_0, z_0, \lambda_0),$$

where $\hat{L}(\cdot, \cdot, \cdot)$ is defined in (10), $V(\cdot, \cdot, \cdot)$ is defined in (20), and (y^*, λ^*, z^*) is an equilibrium of (9) with $x^* = P_\Omega(y^*)$ (equivalently, (x^*, z^*, λ^*) is a saddle point of $\hat{L}$).

Proof. Suppose $a \in \mathcal{L}_{\mathcal{F}}V(y, \lambda, z)$. It follows from Lemma 5.3 that

$$a \leq -(x - x^*)^T(y - y^*) + \|x - x^*\|^2 - (x - x^*)^T(g(x) - g(x^*)) - \lambda^T L \lambda \leq 0,$$

where $x = P_{\Omega}(y)$, $x^* = P_{\Omega}(y^*)$, $g(x) \in \partial f(x)$, and $g(x^*) \in \partial f(x^*)$ is as chosen in (22). The above inequality can be rewritten as

$$a \leq -(x - x^*)^T g(x) - (x - x^*)^T(y - x) + (x - x^*)^T(g(x^*) + y^* - x^*) - \lambda^T L \lambda.$$

By letting $y = u$, $x = P_{\Omega}(y)$, and $x^* = v$ in (3), $(x - x^*)^T(y - x) \geq 0$. Due to (22), $g(x^*) + y^* - x^* = \bar{W}^T \lambda^*$. Hence, $a \leq -(x - x^*)^T g(x) + (x - x^*)^T \bar{W}^T \lambda^* - \lambda^T L \lambda$. Note that $-(x - x^*)^T g(x) \leq f(x^*) - f(x)$, $d = \bar{W}x^* + Lz^*$, and $L\lambda^* = 0_{nm}$. We have

$$\begin{aligned} a &\leq f(x^*) - f(x) + (x - x^*)^T \bar{W}^T \lambda^* - \lambda^T L \lambda \\ &= f(x^*) - f(x) - (d - \bar{W}x - Lz)^T \lambda^* - \lambda^T L \lambda \\ &= \hat{L}(x^*, z^*, \lambda) - \hat{L}(x, z, \lambda^*) \leq 0, \end{aligned}$$

since (x^*, z^*, λ^*) is a saddle point of $\hat{L}$. Integrating both sides over the interval $[0, t]$, it follows that

$$-V(y_0, \lambda_0, z_0) \leq \int_0^t \left(\hat{L}(x^*, z^*, \lambda(s)) - \hat{L}(x(s), z(s), \lambda^*) \right) ds \leq 0.$$

It follows from the Jensen's inequality for the convex-concave $\hat{L}$ that

$$\hat{L}(x^*, z^*, \hat{\lambda}(t)) \geq \frac{1}{t} \int_0^t \hat{L}(x^*, z^*, \lambda(s)) ds$$

and

$$\hat{L}(\hat{x}(t), \hat{z}(t), \lambda^*) \leq \frac{1}{t} \int_0^t \hat{L}(x(s), z(s), \lambda^*) ds.$$

The result is thus proved. $\square$

5.2. Convergence analysis of the distributed derivative feedback algorithm. Consider algorithm (14) (or (15)). $(x^*, \lambda^*, z^*) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ is an equilibrium of (14) if and only if there exists $g(x^*) \in \partial f(x^*)$ such that

$$(27a) \quad 0_{\sum_{i=1}^n q_i} = P_{\Omega}[x^* - g(x^*) + \bar{W}^T \lambda^*] - x^*,$$

$$(27b) \quad 0_{nm} = d - \bar{W}x^* - Lz^*,$$

$$(27c) \quad 0_{nm} = L\lambda^*.$$

THEOREM 5.6. *Suppose Assumption 3.3 holds. If $(x^*, \lambda^*, z^*) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ is an equilibrium of (14), then x^* is a solution to problem (4). Conversely, if $x^* \in \Omega$ is a solution to problem (4), then there exist $\lambda^* \in \mathbb{R}^{nm}$ and $z^* \in \mathbb{R}^{nm}$ such that (x^*, λ^*, z^*) is an equilibrium of (14).*

The proof is similar to that of Theorem 5.1 and hence is omitted.

Suppose $(x^*, \lambda^*, z^*) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ is an equilibrium of (14). Define the function

$$\begin{aligned} V(x, \lambda, z) &= f(x) - f(x^*) + (\lambda^*)^T(d - \bar{W}x) + \frac{1}{2}\|x - x^*\|^2 \\ &\quad + \frac{1}{2}\|\lambda - \lambda^*\|^2 + \frac{1}{2}\|z - z^*\|^2. \end{aligned} \quad (28)$$

LEMMA 5.7. Let function $V(x, \lambda, z)$ be as defined in (28) and Assumption 3.3 hold. For all $(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$, $V(x, \lambda, z)$ is positive definite, $V(x, \lambda, z) = 0$ if and only if $(x, \lambda, z) = (x^*, \lambda^*, z^*)$, and $V(x, \lambda, z) \rightarrow \infty$ as $(x, \lambda, z) \rightarrow \infty$.

Proof. By (27c), there is $\lambda_0^* \in \mathbb{R}^m$ such that $\lambda^* = 1_n \otimes \lambda_0^*$. It can be easily verified that $(\lambda^*)^T(d - \bar{W}x) = (\lambda_0^*)^T(d_0 - Wx)$. It follows from (6) that

$$(29) \quad f(x) - f(x^*) + (\lambda^*)^T(d - \bar{W}x) = f(x) - f(x^*) + (\lambda_0^*)^T W(x^* - x).$$

By (28) and (29), it is straightforward that $V(x^*, \lambda^*, z^*) = 0$.

Because $f(x)$ is convex, $f(x) - f(x^*) \geq g(x^*)^T(x - x^*)$ for all $x \in \Omega$ and $g(x^*) \in \partial f(x^*)$. According to (7),

$$(30) \quad (g(x^*) - W^T \lambda_0^*)^T(x - x^*) \geq 0$$

for all $x \in \Omega$, where $g(x^*) \in \partial f(x^*)$ is as chosen in (7). It follows from (29) and (30) that

$$(31) \quad f(x) - f(x^*) + (\lambda^*)^T(d - \bar{W}x) \geq (g(x^*) - W^T \lambda_0^*)^T(x - x^*) \geq 0$$

for all $x \in \Omega$, where $g(x^*) \in \partial f(x^*)$ is as chosen in (7).

Hence, for all $(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$, $V(x, \lambda, z) \geq \frac{1}{2}\|x - x^*\|^2 + \frac{1}{2}\|\lambda - \lambda^*\|^2 + \frac{1}{2}\|z - z^*\|^2$. Therefore, $V(x, \lambda, z)$ is positive definite, $V(x, \lambda, z) = 0$ if and only if $(x, \lambda, z) = (x^*, \lambda^*, z^*)$, and $V(x, \lambda, z) \rightarrow \infty$ as $(x, \lambda, z) \rightarrow \infty$ for all $(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$. $\square$

LEMMA 5.8. Consider algorithm (14). Under Assumption 3.3 with the function $V(x, \lambda, z)$ defined in (28). If $a \in \mathcal{L}_{\mathcal{F}}V(x, \lambda, z)$, then there exist $g(x) \in \partial f(x)$ and $g(x^*) \in \partial f(x^*)$ such that $a \leq -\|p\|^2 - (x - x^*)^T(g(x) - g(x^*)) - (g(x^*) - \bar{W}^T \lambda^*)^T(x - x^*) - \lambda^T L \lambda \leq 0$, where $p = P_{\Omega}[x - g(x) + \bar{W}^T \lambda] - x$.

Proof. The function $V(x, \lambda, z)$ along the trajectories of (14) satisfies that

$$\begin{aligned} \mathcal{L}_{\mathcal{F}}V(x, \lambda, z) = & \left\{ a \in \mathbb{R} : a = (g(x) - \bar{W}^T \lambda^* + x - x^*)^T p \right. \\ & + \nabla_{\lambda} V(x, \lambda, z)^T(d - \bar{W}x - L\lambda - Lz - \bar{W}p) + \nabla_z V(x, \lambda, z)^T L \lambda, \\ & \left. g(x) \in \partial f(x), p = P_{\Omega}[x - g(x) + \bar{W}^T \lambda] - x \right\}. \end{aligned}$$

Suppose $a \in \mathcal{L}_{\mathcal{F}}V(x, \lambda, z)$. Then there exists $g(x) \in \partial f(x)$ such that

$$(32) \quad \begin{aligned} a = & (g(x) - \bar{W}^T \lambda^* + x - x^*)^T p \\ & + (\lambda - \lambda^*)^T(d - \bar{W}x - L\lambda - Lz - \bar{W}p) + (z - z^*)^T L \lambda, \end{aligned}$$

where

$$(33) \quad p = P_{\Omega}[x - g(x) + \bar{W}^T \lambda] - x.$$

By (3), we represent (33) in the form of a variational inequality

$$\langle p + x - (x - g(x) + \bar{W}^T \lambda), p + x - \tilde{x} \rangle \leq 0 \quad \forall \tilde{x} \in \Omega.$$

Choose $\tilde{x} = x^*$. Then,

$$(34) \quad (g(x) - \bar{W}^T \lambda + x - x^*)^T p \leq -\|p\|^2 - (g(x) - \bar{W}^T \lambda)^T(x - x^*).$$

Since (x^*, λ^*, z^*) is an equilibrium of (14), there is $g(x^*) \in \partial f(x^*)$ such that

$$(35) \quad \begin{cases} 0_{nm} &= L\lambda^*, \\ d &= \bar{W}x^* + Lz^*, \\ x^* &= P_\Omega[x^* - g(x^*) + \bar{W}^T \lambda^*]. \end{cases}$$

It follows from (32), (34), and (35) that

$$\begin{aligned} a &= (g(x) - \bar{W}^T \lambda + x - x^*)^T p + (\lambda - \lambda^*)^T \bar{W} p \\ &\quad + (\lambda - \lambda^*)^T (\bar{W} x^* + L z^* - \bar{W} x - L \lambda - L z - \bar{W} p) \\ &\quad + (z - z^*)^T L \lambda \\ &\leq -\|p\|^2 - (g(x) - \bar{W}^T \lambda)^T (x - x^*) \\ &\quad + (\lambda - \lambda^*)^T \bar{W} p - (x - x^*)^T \bar{W}^T (\lambda - \lambda^*) - \lambda^T L \lambda \\ &\quad - \lambda^T L (z - z^*) - (\lambda - \lambda^*)^T \bar{W} p + (z - z^*)^T L \lambda \\ &= -\|p\|^2 - (g(x) - \bar{W}^T \lambda)^T (x - x^*) - (x - x^*)^T \bar{W}^T (\lambda - \lambda^*) - \lambda^T L \lambda \\ (36) \quad &= -\|p\|^2 - (g(x) - g(x^*))^T (x - x^*) - (g(x^*) - \bar{W}^T \lambda^*)^T (x - x^*) - \lambda^T L \lambda. \end{aligned}$$

Because $x^* = P_\Omega[x^* - g(x^*) + \bar{W}^T \lambda^*]$, $(g(x^*) - \bar{W}^T \lambda^*)^T (x - x^*) \geq 0$ for all $x \in \Omega$ followed by (3). The convexity of f implies that $(x - x^*)^T (g(x) - g(x^*)) \geq 0$. In addition, $L = L_n \otimes I_q \geq 0$ since $L_n \geq 0$. Hence, $a \leq -\|p\|^2 - (x - x^*)^T (g(x) - g(x^*)) - (g(x^*) - \bar{W}^T \lambda^*)^T (x - x^*) - \lambda^T L \lambda \leq 0$. $\square$

Next, we give the convergence analysis of algorithm (14) (or equivalently, algorithm (15)).

THEOREM 5.9. *If Assumption 3.3 holds and algorithm (14) is a convex differential inclusion, then*

- (i) *every solution $(x(t), \lambda(t), z(t))$ is bounded;*
- (ii) *for every solution, $x(t)$ converges to the optimal solution to problem (4).*

Remark 5.10. In Theorem 5.9, algorithm (14) is assumed to be a convex differential inclusion. The convexity assumption of (14) guarantees the existence of solutions to algorithm (14) and satisfies the condition of the invariance principle Lemma 2.1. In fact, in many situations, the convexity of (14) can be satisfied. For example, it holds if $x_i \in \mathbb{R}$, or if both Ω_i and $\partial f^i(x_i)$ are “boxes” for all $i \in \{1, \dots, n\}$; moreover, it also holds if $f^i(\cdot)$ is twice differentiable for all $i \in \{1, \dots, n\}$, in which case algorithm (14) is an ordinary differential equation with Lipschitz continuous right-hand side.

Proof of Theorem 5.9. (i) Suppose $(x^*, \lambda^*, z^*) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$ is an equilibrium of (14). Let function $V(x, \lambda, z)$ be as defined in (28). It follows from Lemma 5.8 that

$$\begin{aligned} \max \mathcal{L}_{\mathcal{F}} V(x, \lambda, z) &\leq \sup \{a : a = -\|p\|^2 - (x - x^*)^T (g(x) - g(x^*)) - \lambda^T L \lambda, \\ &\quad g(x) \in \partial f(x), p = P_\Omega[x - g(x) + \bar{W}^T \lambda] - x\} \leq 0. \end{aligned}$$

It follows from Lemma 5.7 that $V(x, \lambda, z)$ is positive definite for all $(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$, $V(x, \lambda, z) = 0$ if and only if $(x, \lambda, z) = (x^*, \lambda^*, z^*)$, and $V(x, \lambda, z) \rightarrow \infty$ as $(x, \lambda, z) \rightarrow \infty$. Hence, $(x(t), \lambda(t), z(t))$ is bounded for all $t \geq 0$.

(ii) Let $\mathcal{R} = \{(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : 0 \in \mathcal{L}_{\mathcal{F}} V(x, \lambda, z)\} \subset \{(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : \exists g(x) \in \partial f(x), L\lambda = 0_{nm}, (x - x^*)^T (g(x) - g(x^*)) = 0, 0_{\sum_{i=1}^n q_i} =$

$P_\Omega[x - g(x) + \bar{W}^\top \lambda] - x\}$. Let $\mathcal{M}$ be the largest weakly invariant subset of $\bar{\mathcal{R}}$. It follows from Lemma 2.1 that $(x(t), \lambda(t), z(t)) \rightarrow \mathcal{M}$ as $t \rightarrow \infty$. Note that $(x - x^*)^\top (g(x) - g(x^*)) > 0$ if $x \neq x^*$ because of the strict convexity of f and Lemma 2.2. Hence, $x(t) \rightarrow x^*$ as $t \rightarrow \infty$. $\square$

Remark 5.11. The analysis of nonconvex differential inclusions is still challenging. However, optimization algorithms described by projected nonconvex differential inclusions are investigated in [5, 6, 7] and applied to convex optimization problems with rigorous convergence analysis. Hence, the convergence analysis of algorithms in the form of nonconvex differential inclusions may be conducted by following the ideas of [5, 6, 7, 22].

The following gives a convergence result when the objective functions of problem (4) are differentiable. If the objective functions of problem (4) are differentiable, algorithm (14) becomes an ordinary differential equation and the strict convexity requirement of the objective functions can be relaxed, still with our proposed algorithm and technique.

COROLLARY 5.12. *Consider algorithm (14). With points 1 and 3 of Assumption 3.3, if f_i is twice differentiable and convex on an open set containing Ω_i for $i \in \{1, \dots, n\}$, then*

- (i) *the solution $(x(t), \lambda(t), z(t))$ is bounded;*
- (ii) *the solution $(x(t), \lambda(t), z(t))$ is convergent and $x(t)$ converges to an optimal solution to problem (4).*

Proof. (i) The proof of (i) is similar to that of Theorem 5.9(i). Hence, it is omitted.

(ii) It follows from similar arguments in the proof of Theorem 5.9(i) that

$$(37) \quad \frac{d}{dt} V(x, \lambda, z) \leq -\|\dot{x}\|^2 - (x - x^*)^\top (\nabla f(x) - \nabla f(x^*)) - \lambda^\top L \lambda \leq 0,$$

where $\dot{x} = P_\Omega[x - \nabla f(x) + \bar{W}^\top \lambda] - x$ and function $V(x, \lambda, z)$ is defined in (28).

Let $\mathcal{R} = \{(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : 0 = \frac{d}{dt} V(x, \lambda, z)\} \subset \{(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : L\lambda = 0_{nm}, (x - x^*)^\top (\nabla f(x) - \nabla f(x^*)) = 0, 0_{\sum_{i=1}^n q_i} = P_\Omega[x - \nabla f(x) + \bar{W}^\top \lambda] - x\}$. Let $\mathcal{M}$ be the largest invariant subset of $\bar{\mathcal{R}}$. It follows from the invariance principle (Theorem 2.41 of [19]) that $(x(t), \lambda(t), z(t)) \rightarrow \mathcal{M}$ as $t \rightarrow \infty$. Note that $\mathcal{M}$ is invariant. Let $(\bar{x}(t), \bar{\lambda}(t), \bar{z}(t))$ be a trajectory of algorithm (14). Then $(\bar{x}(t), \bar{\lambda}(t), \bar{z}(t)) \in \mathcal{M}$ for all $t \geq 0$ if $(\bar{x}(0), \bar{\lambda}(0), \bar{z}(0)) = (\bar{x}_0, \bar{\lambda}_0, \bar{z}_0) \in \mathcal{M}$. Assume $(\bar{x}(t), \bar{\lambda}(t), \bar{z}(t)) \in \mathcal{M}$ for all $t \geq 0$, $\dot{\bar{x}}(t) \equiv 0_{\sum_{i=1}^n q_i}$ and $\dot{\bar{z}}(t) \equiv 0_{nm}$ and, hence, $\dot{\bar{\lambda}}(t) \equiv d - \bar{W}\bar{x}_0 - L\bar{z}_0$. Suppose $\dot{\bar{\lambda}}(t) \equiv d - \bar{W}\bar{x}_0 - L\bar{z}_0 \neq 0_{nm}$, then $\bar{\lambda}(t) \rightarrow \infty$ as $t \rightarrow \infty$, which contradicts part (i). Hence, $\dot{\bar{\lambda}}(t) \equiv 0_{nm}$ and $\mathcal{M} \subset \{(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : P_\Omega[x - \nabla f(x) + \bar{W}^\top \lambda] - x = 0_{\sum_{i=1}^n q_i}, d - \bar{W}x - Lz = 0_{nm}, L\lambda = 0_{nm}\}$.

Take any $(\bar{x}, \bar{\lambda}, \bar{z}) \in \mathcal{M}$. Obviously, $(\bar{x}, \bar{\lambda}, \bar{z})$ is an equilibrium point of algorithm (14). Define a new function $\bar{V}(x, \lambda, z)$ by replacing (x^*, λ^*, z^*) with $(\bar{x}, \bar{\lambda}, \bar{z})$ in $V(x, \lambda, z)$. It follows from similar arguments in the proof of Lemma 5.8 that $\frac{d}{dt} \bar{V}(x, \lambda, z) \leq 0$. Hence, $(\bar{x}, \bar{\lambda}, \bar{z})$ is Lyapunov stable. By Proposition 4.7 of [19], there exists $(\tilde{x}, \tilde{\lambda}, \tilde{z}) \in \mathcal{M}$ such that $(x(t), \lambda(t), z(t)) \rightarrow (\tilde{x}, \tilde{\lambda}, \tilde{z})$ as $t \rightarrow \infty$. Since $(\tilde{x}, \tilde{\lambda}, \tilde{z}) \in \mathcal{M}$ is an equilibrium point of algorithm (14), $\tilde{x}$ is an optimal solution to problem (4) by Theorem 5.6. $\square$

Define $\hat{x}(t)$, $\hat{\lambda}(t)$, and $\hat{z}(t)$ as in (26), where $(x(t), \lambda(t), z(t))$ is a trajectory gen-

erated by algorithm (14). We show the convergence rate of $(\hat{x}(t), \hat{\lambda}(t), \hat{z}(t))$ in the following result.

THEOREM 5.13. *Consider algorithm (14). If Assumption 3.3 holds, then*

$$0 \leq \hat{L}(\hat{x}(t), \hat{z}(t), \lambda^*) - \hat{L}(x^*, z^*, \hat{\lambda}(t)) \leq \frac{1}{t} V(x_0, \lambda_0, z_0),$$

where $\hat{L}(\cdot, \cdot, \cdot)$ is defined in (10), $V(\cdot, \cdot, \cdot)$ is defined in (28), and (x^*, λ^*, z^*) is an equilibrium of (14) (equivalently, a saddle point of $\hat{L}$).

Proof. Suppose $a \in \mathcal{L}_{\mathcal{F}} V(x, \lambda, z)$. It follows from Lemma 5.8 that there exists $g(x) \in \partial f(x)$ such that

$$a \leq -\|p\|^2 - (x - x^*)^T g(x) + (\bar{W}^T \lambda^*)^T (x - x^*) - \lambda^T L \lambda \leq 0,$$

where $p = P_{\Omega}[x - g(x) + \bar{W}^T \lambda] - x$. Note that $-(x - x^*)^T g(x) \leq f(x^*) - f(x)$, $d = \bar{W}x^* + Lz^*$, and $L\lambda^* = 0_{nm}$. We have

$$\begin{aligned} a &\leq -\|p\|^2 + f(x^*) - f(x) - (d - \bar{W}x - Lz)^T \lambda^* - \lambda^T L \lambda \\ &\leq -\|p\|^2 + \hat{L}(x^*, z^*, \lambda) - \hat{L}(x, z, \lambda^*) \leq 0, \end{aligned}$$

since (x^*, z^*, λ^*) is a saddle point of $\hat{L}$. Integrating both sides over the interval $[0, t]$, it follows that

$$-V(x_0, \lambda_0, z_0) \leq \int_0^t \left(\hat{L}(x^*, z^*, \lambda(s)) - \hat{L}(x(s), z(s), \lambda^*) \right) ds \leq 0.$$

It follows from the Jensen's inequality for the convex-concave $\hat{L}$ that

$$\hat{L}(x^*, z^*, \hat{\lambda}(t)) \geq \frac{1}{t} \int_0^t \hat{L}(x^*, z^*, \lambda(s)) ds$$

and

$$\hat{L}(\hat{x}(t), \hat{z}(t), \lambda^*) \leq \frac{1}{t} \int_0^t \hat{L}(x(s), z(s), \lambda^*) ds.$$

Then the result is proved. $\square$

Remark 5.14. The results of this paper are related to that in [40] but different in a few ways. First, the problem formulation is more general and considers nonsmooth objective functions with potential applications in power dispatch [11, 29, 37], compress sensing [15, 44], and LASSO problems [8]. Second, the proposed algorithms are designed using new ideas as discussed in section 4.3. Third, nonsmooth analysis is used to prove the convergence property and convergence rates of the proposed algorithms.

6. Numerical experiments. In this section, we give a numerical example to illustrate the effectiveness of the proposed algorithms.

Consider the nonsmooth optimization problem with a six-agent undirected and connected network,

$$(38) \quad \min_x f(x), \quad f(x) = \sum_{i=1}^6 \frac{1}{2} \|x_i\|^2 + \|x_i\|_1, \quad \sum_{i=1}^6 A_i x_i = \sum_{i=1}^6 d_i = d_0, \quad \|x_i\|_{\infty} \leq 1,$$

where $i \in \{1, \dots, 6\}$, $W = [W_1, \dots, W_6] \in \mathbb{R}^{3 \times 24}$, $x_i \in \mathbb{R}^4$, and $x = [x_1^T, \dots, x_6^T]^T \in \mathbb{R}^{24}$. Each agent i knows A_i and d_i with

$$\begin{aligned} W_1 &= \begin{bmatrix} 0.63 & 0.58 & 0.65 & 0.33 \\ 0.04 & 0.6 & 0.5 & 0.81 \\ 0.8 & 0.25 & 0.53 & 0.79 \end{bmatrix}, & W_2 &= \begin{bmatrix} 0.68 & 0.22 & 0.49 & 0.21 \\ 0.01 & 0.51 & 0.23 & 0.29 \\ 0.13 & 0.79 & 0.34 & 0.45 \end{bmatrix}, \\ W_3 &= \begin{bmatrix} 0.62 & 0.57 & 0.71 & 0.28 \\ 0.25 & 0.21 & 0.66 & 0.9 \\ 0.1 & 0.94 & 0.78 & 0.7 \end{bmatrix}, & W_4 &= \begin{bmatrix} 0.44 & 0.06 & 0.77 & 0.16 \\ 0.34 & 0.94 & 0.28 & 0.41 \\ 0.99 & 0.65 & 0.38 & 0.12 \end{bmatrix}, \\ W_5 &= \begin{bmatrix} 0.84 & 0.62 & 0.74 & 0.26 \\ 0.75 & 0.56 & 0.41 & 0.89 \\ 0.76 & 0.52 & 0.55 & 0.24 \end{bmatrix}, & W_6 &= \begin{bmatrix} 0.44 & 0.28 & 0.50 & 0.38 \\ 0.69 & 0.23 & 0.88 & 0.63 \\ 0.55 & 0.51 & 0.58 & 0.85 \end{bmatrix}, \end{aligned}$$

$$d_1 = d_2 = d_3 = [0.47, 0.52, 0.77]^T, \quad d_4 = d_5 = d_6 = [0.63, 0.33, 0.34]^T.$$

The adjacency matrix of graph $\mathcal{G}$ is given by

$$\mathcal{A} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Some simulation results using the distributed projected output feedback algorithm (DPOFA) (9) proposed in section 4.1 and the distributed derivative feedback algorithm (DDFA) algorithm (14) proposed in section 4.2 are shown in Figures 1–5.

Figure 1 shows the trajectories of estimates for x versus time under DPOFA algorithm (9) proposed in section 4.1 and Figure 2 depicts the trajectories of the estimates for x versus time under DDFA algorithm (14) proposed in section 4.2. Algorithm (9) proposed in section 4.1 uses an auxiliary variable y and estimates the optimal solution using $x = P_\Omega(y)$, while algorithm (14) proposed in section 4.2 directly uses x to estimate the solution. Both algorithms are able to find the optimal solution of the optimization problem. Figures 3 and 4 verify the boundedness of trajectories of DPOFA algorithm (9) and DDFA algorithm (14). Figure 5 gives the trajectories of the objective function $f(x)$ and constraint $\|Wx - d_0\|$ versus time under DPOFA algorithm (9) and DDFA algorithm (14) and demonstrates that the trajectories of x converge to the equality constraint.

In Figure 5, the trajectories of $f(x)$ and $\|Wx - d_0\|$ versus time under the DPOFA algorithm show slow response speed at the beginning of the simulation. This is because the change of y in the algorithm may not generate the changing behavior of $x = P_\Omega(y)$ when $y \notin \Omega$. Due to the indirect feedback effect on x (changing x by controlling y) in the DPOFA algorithm, the trajectory of variable x may show slow changing behaviors in applications.

7. Conclusions. In this paper, distributed designs for the extended monotropic optimization problems have been addressed, which are related to various applications in large-scale optimization and evolutionary computation. In this paper, two novel distributed continuous-time algorithms using projected output feedback design and derivative feedback design have been proposed to solve this problem in multiagent networks. The design of the algorithms used the decomposition of problem constraints

FIG. 1. Trajectories of estimates for x versus time under algorithm (9) for problem (38). FIG. 2. Trajectories of estimates for x versus time under algorithm (14) for problem (38).

and distributed techniques. Based on stability theory and the invariance principle for differential inclusions, the convergence properties and convergence rates of the proposed algorithms have been established. The trajectories of all the agents have been proved to be bounded and convergent to the optimal solution with any initial condition in mathematical and numerical ways.

FIG. 3. $\|y\|$, $\|\lambda\|$, and $\|z\|$ versus time under algorithm (9) for problem (38).

FIG. 4. $\|x\|$, $\|\lambda\|$, and $\|z\|$ versus time under algorithm (14) for problem (38).

FIG. 5. Objective functions $f(x)$ and constraints $\|Wx - d_0\|$ versus time under algorithms (9) and (14) for problem (38).

REFERENCES

- [1] B. ANDERSON, S. MOU, A. MORSE, AND U. HELMKE, *Decentralized gradient algorithm for solution of a linear equation*, Numer. Algebra Control Optim., 6 (2016), pp. 319–328.
- [2] J. P. AUBIN AND A. CELLINA, *Differential Inclusions*, Springer-Verlag, Berlin, 1984.
- [3] A. BACCIOTTI AND F. CERAGIOLI, *Stability and stabilization of discontinuous systems and nonsmooth Lyapunov functions*, ESAIM Control Optim. Calc. Var., 4 (1999), pp. 361–376.
- [4] D. P. BERTSEKAS, *Extended monotropic programming and duality*, J. Optim. Theory Appl., 139 (2008), pp. 209–225.
- [5] W. BIAN AND X. XUE, *Evolution differential inclusion with projection for solving constrained nonsmooth convex optimization in Hilbert space*, Set-Valued Var. Anal., 20 (2012), pp. 203–227.
- [6] W. BIAN AND X. XUE, *Asymptotic behavior analysis on multivalued evolution inclusion with projection in Hilbert space*, Optimization, 64 (2013), pp. 853–875.
- [7] J. BOLTE, *Continuous gradient projection method in Hilbert spaces*, J. Optim. Theory Appl., 119 (2003), pp. 235–259.
- [8] S. P. BOYD, N. PARIKH, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Machine Learning, 3 (2011), pp. 1–122.
- [9] N. CHATZIPANAGIOTIS, D. DENTCHEVA, AND M. M. ZAVLANOS, *An augmented lagrangian method for distributed optimization*, Math. Program. Ser. A, 152 (2015), pp. 405–434.
- [10] A. CHERUKURI AND J. CORTÉS, *Distributed generator coordination for initialization and any-time optimization in economic dispatch*, IEEE Trans. Control Network Systems, 2 (2015), pp. 226–237.
- [11] C. L. CHIANG, *Improved genetic algorithm for power economic dispatch of units with valve-point effects and multiple fuels*, IEEE Trans. Power Systems, 20 (2005), pp. 1690–1699.
- [12] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
- [13] J. CORTÉS, *Discontinuous dynamical systems*, IEEE Control Systems Magazine, 44 (1999), pp. 1995–2006.
- [14] Z. DENKOWSKI, S. MIGÓRSKI, AND N. S. PAPAGEORGIOU, *An Introduction to Nonlinear Analysis: Theory*, Springer-Verlag, New York, 2003.
- [15] D. L. DONOHO AND J. TANNER, *Sparse nonnegative solutions of underdetermined linear equations by linear programming*, Proc. Nat. Acad. Sci., 102 (2005), pp. 9446–9451.

- [16] F. FACCINEI AND J. PANG, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, 2003.
- [17] B. GHARESIFARD AND J. CORTÉS, *Distributed continuous-time convex optimization on weight-balanced digraphs*, IEEE Trans. Automat. Control, 59 (2014), pp. 781–786.
- [18] C. GODSIL AND G. F. ROYLE, *Algebraic Graph Theory*, Springer-Verlag, New York, 2001.
- [19] W. M. HADDAD AND V. CHELLABOINA, *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*, Princeton University Press, Princeton, NJ, 2008.
- [20] D. KINDERLEHRER AND G. STAMPACCHIA, *An Introduction to Variational Inequalities and Their Applications*, Academic Press, New York, 1982.
- [21] J. KOSHAL, A. NEDIC, AND U. SHANBHAG, *Multiuser optimization: distributed algorithms and error analysis*, SIAM J. Optim., 21 (2011), pp. 1046–1081.
- [22] S. LIANG, X. ZENG, AND Y. HONG, *Lyapunov stability and generalized invariance principle for nonconvex differential inclusions*, Control Theory Technol., 14 (2016), pp. 140–150.
- [23] Q. LIU AND J. WANG, *A one-layer projection neural network for nonsmooth optimization subject to linear equalities and bound constraints*, IEEE Trans. Neural Networks Learning Systems, 24 (2013), pp. 812–824.
- [24] Q. LIU AND J. WANG, *A second-order multi-agent network for bound-constrained distributed optimization*, IEEE Trans. Automat. Control, 60 (2015), pp. 3310–3315.
- [25] Q. LIU, S. YANG, AND J. WANG, *A collective neurodynamic approach to distributed constrained optimization*, IEEE Trans. Neural Networks Learning Systems, 28 (2017), pp. 1747–1758.
- [26] D. MATEOS-NUNEZ AND J. CORTES, *Noise-to-state exponentially stable distributed convex optimization on weight-balanced digraphs*, SIAM J. Control Optim., 54 (2016), pp. 266–290.
- [27] S. MOU, J. LIU, AND A. S. MORSE, *A distributed algorithm for solving a linear algebraic equation*, IEEE Trans. Automat. Control, 60 (2015), pp. 2863–2878.
- [28] A. NEDIC, A. OZDAGLAR, AND P. A. PARRILO, *Constrained consensus and optimization in multi-agent networks*, IEEE Trans. Automat. Control, 55 (2010), pp. 922–938.
- [29] J. PARK, K. LEE, J. SHIN, AND K. Y. LEE, *Economic load dispatch for non-smooth cost functions using particle swarm optimization*, in Proceedings of the IEEE Power Engineering Society General Meeting, Toronto, Canada, 2003, pp. 938–943.
- [30] Z. QIU, S. LIU, AND L. XIE, *Distributed constrained optimal consensus of multi-agent systems*, Automatica, 68 (2016), pp. 209–215.
- [31] R. ROCKAFELLAR, *Network Flows and Monotropic Optimization*, Wiley, New York, 1984.
- [32] R. ROCKAFELLAR, *Monotropic Programming: A Generalization of Linear Programming and Network Programming. in Convexity and Duality in Optimization*, Springer-Verlag, Berlin, 1985.
- [33] A. RUSZCZYNSKI, *Nonlinear Optimization*, Princeton University Press, Princeton, NJ, 2006.
- [34] G. SHI, B. D. O. ANDERSON, AND U. HELMKE, *Network flows that solve linear equations*, IEEE Trans. Automat. Control, 62 (2017), pp. 2659–2674.
- [35] G. SHI, K. JOHANSSON, AND Y. HONG, *Reaching an optimal consensus: Dynamical systems that compute intersections of convex sets*, IEEE Trans. Automat. Control, 55 (2013), pp. 610–622.
- [36] G. STRANG, *The fundamental theorem of linear algebra*, Amer. Math. Monthly, 100 (1993), pp. 848–855.
- [37] D. C. WALTERS AND G. B. SHEBLE, *Genetic algorithm solution of economic dispatch with valve point loading*, IEEE Trans. Power Systems, 8 (1993), pp. 1325–1332.
- [38] S. YANG, Q. LIU, AND J. WANG, *A multi-agent system with a proportional-integral protocol for distributed constrained optimization*, IEEE Trans. Automat. Control, 62 (2017), pp. 3461–3467.
- [39] P. YI, Y. HONG, AND F. LIU, *Distributed gradient algorithm for constrained optimization with application to load sharing in power systems*, Systems Control Lett., 83 (2015), pp. 45–52.
- [40] P. YI, Y. HONG, AND F. LIU, *Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and its application to economic dispatch of power systems*, Automatica, 74 (2016), pp. 259–269.
- [41] D. YUAN, D. W. C. HO, AND S. XU, *Zeroth-order method for distributed optimization with approximate projections*, IEEE Trans. Neural Networks Learning Systems, 27 (2016), pp. 284–294.
- [42] X. ZENG, P. YI, AND Y. HONG, *Distributed continuous-time algorithm for constrained convex optimizations via nonsmooth analysis approach*, IEEE Trans. Automat. Control, 62 (2017), pp. 5227–5233.
- [43] Y. ZHANG, Z. DENG, AND Y. HONG, *Distributed optimal coordination for multiple heterogeneous Euler-Lagrangian systems*, Automatica, 79 (2017), pp. 207–213.
- [44] J. ZHOU, W. XUAN, S. MOU, AND B. D. O. ANDERSON, *Finite-Time Distributed Linear Equation Solver for Minimum l_1 Norm Solutions*, arXiv:1709.10154v1, 2017.