Continuous-Time Distributed Algorithms for Extended Monotropic Optimization Problems

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

Abstract

This paper studies distributed algorithms for the extended monotropic optimization problem, which is a general convex optimization problem with a certain separable structure. The considered objective function is the sum of local convex functions assigned to agents in a multi-agent network, with private set constraints and affine equality constraints. Each agent only knows its local objective function, local constraint set, and neighbor information. We propose two novel continuous-time distributed subgradient-based algorithms with projected output feedback and derivative feedback, respectively, to solve the extended monotropic optimization problem. Moreover, we show that the algorithms converge to the optimal solutions under some mild conditions, by virtue of variational inequalities, Lagrangian methods, decomposition methods, and nonsmooth Lyapunov analysis. Finally, we give two examples to illustrate the applications of the proposed algorithms.

Key Words: Extended monotropic optimization, distributed algorithms, nonsmooth convex functions, decomposition methods, differential inclusions.

I. INTRODUCTION

Recently, distributed (convex) optimization problems have been studied in many fields including sensor networks, neural learning systems, and power systems [1]–[4]. Such problems are often formulated in terms of an objective 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 multi-agent networks have been proposed since these distributed algorithms have advantages over centralized ones for large-scale optimization problems, with inexpensive and low-performance computations for each agent/node. Both continuous-time and discrete-time algorithms for distributed optimization without any constraint were extensively studied recently [5]–[7].

It is known that continuous-time algorithms are becoming more and more popular, especially when the optimization may be achieved by continuous-time physical systems. In fact, the neurodynamic...
approach is one of the continuous-time optimization approaches, and has been well developed for numerous neural network models [8]–[11]; and the optimization in smart grids has also been studied based on continuous-time dynamical systems [3], [6].

In practice, many optimization problems get involved with various constraints and nonsmooth objective functions [1], [3], [6], [12]. Subgradients and projections have been found to be widely used in the study of distributed nonsmooth optimization with constraints [3], [13], [14]. In fact, different algorithms were proposed in the form of differential inclusions to solve nonsmooth optimization problems in the literature (see [2], [11], [15]).

The monotropic optimization is an optimization problem with separable 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 [16], [17]. The extended monotropic optimization (EMO) problem was then studied in [18], 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 [18]. Recently, distributed algorithms and decomposition methods for EMO problems have become more and more important, as pointed out in [19]. However, due to the complicatedness of EMO problems, very few distributed optimization designs for them have been proposed.

This paper aims to investigate EMO problems with nonsmooth objective functions via distributed approaches. Based on Lagrangian functions and decomposition methods, our distributed algorithms are given based on projections to achieve the optimal solutions of the problems. The analysis of the proposed algorithms is carried out by applying the stability theory of differential inclusions to tackle nonsmooth objective functions. The main technical contributions of the paper are three folds. Firstly, we study the distributed EMO problem, and propose two novel distributed continuous-time algorithms for the problem, by using projected output feedback and derivative feedback, respectively, to deal with the nonsmoothness and set constraints. Secondly, based on the Lagrangian function method along with projections, we show that the proposed algorithms have bounded states and solve the EMO problems with any initial condition, by virtue of the stability theory of differential inclusions and nonsmooth analysis techniques. Thirdly, we apply the given algorithms to two practical problems and illustrate their effectiveness.

The rest of the paper is organized as follows: Section II shows the preliminary knowledge related to graph theory, nonsmooth analysis, convex optimization, and projection operators. Next, Section III
formulates a class of distributed extended monotropic optimization (EMO) problems with nonsmooth objective functions, while Section IV proposes two distributed algorithms based on projected output feedback and derivative feedback. Then Section V shows the convergence of the proposed algorithms with nonsmooth analysis. Following that, Section VI gives two application examples to illustrate the theoretical results. Finally, Section VII presents some concluding remarks.

II. MATHEMATICAL PRELIMINARIES

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

A. 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 $A$, $\text{diag}\{A_1, \ldots, A_n\}$ as the block diagonal matrix of $A_1, \ldots, A_n$, $1_n (1_{n \times q})$ 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 semi-definite). Furthermore, $\| \cdot \|$ stands for the Euclidean norm; $\overline{S}$ ($S^o$) for the closure (interior) of the subset $S \subset \mathbb{R}^n$; $B_\epsilon(x), x \in \mathbb{R}^n, \epsilon > 0$, for the open ball centered at $x$ with radius $\epsilon$. $\text{dist}(p, M)$ denotes the distance from a point $p$ to the set $M$ (that is, $\text{dist}(p, M) \triangleq \inf_{x \in M} \|p - x\|$), and $x(t) \to M$ as $t \to \infty$ denotes that $x(t)$ approaches the set $M$ (that is, for each $\epsilon > 0$, there is $T > 0$ such that $\text{dist}(x(t), M) < \epsilon$ for all $t > T$).

B. Graph Theory

A weighted undirected graph is described by $G$ or $G(V, E, A)$, where $V = \{1, \ldots, n\}$ is the set of nodes, $E \subset V \times V$ is the set of edges, $A = [a_{i,j}] \in \mathbb{R}^{n \times n}$ is the weighted adjacency matrix such that $a_{i,j} = a_{j,i} > 0$ if $\{j, i\} \in E$ 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 $D_{i,i} = \sum_{j=1}^n a_{i,j}, i \in \{1, \ldots, n\}$. In this paper, we call $L_n$ the Laplacian matrix and $A$ the adjacency matrix of $G$ for convenience when there is no confusion. Specifically, if the weighted undirected graph $G$ is connected, then $L_n = L_n^T \geq 0$, $\text{rank } L_n = n - 1$ and $\ker(L_n) = \{k1_n : k \in \mathbb{R}\}$ [20].

August 4, 2016 DRAFT
C. Differential Inclusion

Following [21], a differential inclusion is given by

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

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

Let \( V : \mathbb{R}^q \to \mathbb{R} \) be a locally Lipschitz continuous function, and \( \partial V \) the Clarke generalized gradient [22] of \( V(x) \) at \( x \). The set-valued Lie derivative [22] \( \mathcal{L}_x V : \mathbb{R}^q \to \mathcal{B}(\mathbb{R}) \) of \( V \) with respect to (I) is defined as \( \mathcal{L}_x V(x) \triangleq \{a \in \mathbb{R} : \text{there exists } v \in \mathcal{F}(x) \text{ such that } p^T v = a \text{ for all } p \in \partial V(x)\} \). In the case when \( \mathcal{L}_x V(x) \) is nonempty, we use \( \max \mathcal{L}_x V(x) \) to denote the largest element of \( \mathcal{L}_x V(x) \). Recall from reference [23] that, if \( \phi(\cdot) \) is a solution of (I) and \( V : \mathbb{R}^q \to \mathbb{R} \) is locally Lipschitz and regular (see [22] p. 39)), then \( \dot{V}(\phi(t)) \) exists almost everywhere, and \( \dot{V}(\phi(t)) \in \mathcal{L}_x V(\phi(t)) \) almost everywhere.

Next, we introduce a version of the invariance principle (Theorem 2 of [24]), which is based on nonsmooth regular functions.

**Lemma 2.1:** [24] For the differential inclusion (I), we assume that \( \mathcal{F} \) is upper semicontinuous and locally bounded, and \( \mathcal{F}(x) \) takes nonempty, compact, and convex values. Let \( V : \mathbb{R}^q \to \mathbb{R} \) be a locally Lipschitz and regular function, \( S \subseteq \mathbb{R}^q \) be compact and strongly invariant for (I), \( \phi(\cdot) \) be a solution of (I),

\[ \mathcal{R} = \{x \in \mathbb{R}^q : 0 \in \mathcal{L}_x V(x)\}, \]
and $M$ be the largest weakly invariant subset of $\overline{R}\cap S$, where $\overline{R}$ is the closure of $R$. If $\max L_x V(x) \leq \delta$ for all $x \in S$, then $\text{dist}(\phi(t), M) \to 0$ as $t \to +\infty$.

D. Convex Analysis

A function $\psi : \mathbb{R}^q \to \mathbb{R}$ 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 $\psi : \mathbb{R}^q \to \mathbb{R}$ 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 $f : \mathbb{R}^q \to \mathbb{R}$ be a convex function. The subdifferential [25] p. 544] $\partial_{\text{sub}} \psi$ of $\psi$ at $x \in \mathbb{R}^q$ is defined by $\partial_{\text{sub}} \psi(x) \triangleq \{ p \in \mathbb{R}^q : \langle p, y - x \rangle \leq \psi(y) - \psi(x), \forall y \in \mathbb{R}^q \}$, and the elements of $\partial_{\text{sub}} \psi(x)$ are called subgradients of $\psi$ at point $x$. Recall from [25] p. 607] that continuous convex functions are locally Lipschitz continuous, 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:** Let $f : \mathbb{R}^q \to \mathbb{R}$ be a continuous strictly convex function. Then

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

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

E. 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 [26] of a projection $P_\Omega(\cdot)$ on a closed convex set $\Omega \subset \mathbb{R}^n$ is

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

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

**Lemma 2.3:** 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:** Let $\Omega \subset \mathbb{R}^n$ be closed and convex, and define $V : \mathbb{R}^n \to \mathbb{R}$ 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 $\nabla V(x) = P_\Omega(x) - P_\Omega(y)$.

III. Problem Description

In this section, we present the distributed extended monotropic optimization (EMO) problem with non-smooth objective functions, and give the optimality condition for the problem.
Consider a network of \( n \) agents interacting over a graph \( G \). There are a local objective function \( f^i : \Omega_i \to \mathbb{R} \) and a local feasible constraint set \( \Omega_i \subset \mathbb{R}^{q_i} \) for all \( i \in \{1, \ldots, n\} \). Let \( x_i \in \Omega_i \subset \mathbb{R}^{q_i} \) and denote \( x \triangleq [x_1^T, \ldots, x_n^T]^T \in \Omega \triangleq \prod_{i=1}^n \Omega_i \subset \mathbb{R}^{\sum_{i=1}^n q_i} \). 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 consider the following distributed EMO problem

\[
\min f(x), \quad f(x) = \sum_{i=1}^n f^i(x_i),
\]

\[
Wx = \sum_{i=1}^n W_i x_i = d_0, \quad x_i \in \Omega_i \subset \mathbb{R}^{q_i}, \quad i \in \{1, \ldots, n\},
\]

where \( W_i \in \mathbb{R}^{m_i \times q_i}, i \in \{1, \ldots, n\} \) and \( W = [W_1, \ldots, W_n] \in \mathbb{R}^{m \times \sum_{i=1}^n q_i} \). 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 \( W_i \in \mathbb{R}^{m_i \times q_i} \), 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 \( G \) 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 [16], [17].

**Remark 3.1:** The distributed EMO problem (4) covers many problems in the recent distributed optimization studies because of the general expression. For example, it generalizes the optimization model in resource allocation problems [3], [14] by allowing nonsmooth objective functions and a more general equality constraint. Moreover, it covers the model proposed in [8] and generalizes the model in the distributed constrained optimal consensus problem [27] by allowing heterogeneous constraints.

For illustration, we introduce two special cases of our problem:

- Consider the following optimization problem investigated in [8]

\[
\min f(x), \quad f(x) = \sum_{i=1}^n f^i(x_i), \quad x_i \in \Omega_i = \{x_i \in \mathbb{R}^q : g_i(x_i) \leq 0\},
\]

\[
A_i x_i = b_i, \quad i \in \{1, \ldots, n\},
\]

\[
L x = 0_{nq},
\]

where \( x \triangleq [x_1^T, \ldots, x_n^T]^T \in \mathbb{R}^{nq}, A_i \in \mathbb{R}^{m_i \times q}, b_i \in \mathbb{R}^{m_i}, L \in \mathbb{R}^{nq \times nq}, x_i \in \mathbb{R}^q, \) and \( f^i : \mathbb{R}^q \rightarrow \mathbb{R} \) for all \( i \in \{1, \ldots, n\} \). Equation (5b) is the local equality constraint for agent \( i \), and equation (5c)
is the global equality constraint. Let $d_0 \triangleq [b_1^T, \ldots, b_n^T, 0_{nq}^T], A = \text{diag}[A_1, \ldots, A_n] \in \mathbb{R}^{\sum_{i=1}^n m_i \times nq^2}$, and $W \triangleq \begin{bmatrix} 1 \\ L \end{bmatrix} \in \mathbb{R}^{(nq + \sum_{i=1}^n m_i) \times nq}$. Problem (5) can be written in the form of (4). Hence, the EMO problem described by (4) covers the optimization problem with local and global equality constraints given by (5).

- Consider the minimal norm problem of underdetermined linear equation [28]

$$\min f(x), \quad f(x) = \sum_{i=1}^n \|x_i\|^2, \quad \sum_{i=1}^n A_ix_i = b,$$

where $x \triangleq [x_1^T, \ldots, x_n^T]^T \in \mathbb{R}^{\sum_{i=1}^n q_i}, A_i \in \mathbb{R}^{m \times q_i}, b \in \mathbb{R}^m, x_i \in \mathbb{R}^{q_i}$, and $f_i : \mathbb{R}^{q_i} \to \mathbb{R}$ for all $i \in \{1, \ldots, n\}$. Hence, the EMO problem also covers the minimal norm problem of linear algebraic equation. In the description of our problem, each agent knows a block $A_i$, which is different from the framework in [29] solving the distributed linear equation, where each agent knows a subset of the rows of $[A_1, \ldots, A_n]$ and $b$. If the number of variables is large, our framework obviously has advantages on reducing the computation and information loads of agents.

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

**Assumption 3.1:**

1) The weighted graph $G$ is connected and undirected.

2) For all $i \in \{1, \ldots, n\}$, $f^i$ is strictly convex on an open set containing $\Omega_i$, 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 $\Omega^\circ$ is the interior of $\Omega$.

**Lemma 3.1:** Under Assumption 3.1, $x^* \in \Omega$ 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

$$x^* = P_{\Omega}(x^* - g(x^*) + W^T\lambda^*_0),$$

$$Wx^* = d_0.$$  

**Proof:** Consider problem (4). By the KKT optimality condition (Theorem 3.25 of [30]), $x^* \in \Omega$ 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 (8) holds and

$$- g(x^*) + W^T\lambda^*_0 \in N_{\Omega}(x^*),$$

where $N_{\Omega}(x^*)$ is the normal cone of $\Omega$ at an element $x^* \in \Omega$. Note that (9) holds if and only if (7) holds. Thus, the proof is completed.

August 4, 2016 DRAFT
IV. 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 continuous-time algorithms for such problems with rigorous convergence analysis.

The resource allocation problem, a special case of the EMO problem, was studied for problems with smooth objective functions in [14]. An intuitive extension of the continuous-time algorithm given in [14] to nonsmooth EMO cases may be written as

\[
\begin{align*}
\dot{x}_i(t) &\in \{ p : p = P_{\Omega_i} [x_i(t) - g_i(x_i(t)) + W_i^T \lambda_i(t)] - x_i(t), \quad g_i(x_i(t)) \in \partial f^i(x_i(t)) \}, \\
\dot{\lambda}_i(t) &= d_i - W_i x_i(t) - \sum_{j=1}^{n} a_{i,j} (\lambda_i(t) - \lambda_j(t)) - \sum_{j=1}^{n} a_{i,j} (z_i(t) - z_j(t)), \\
\dot{z}_i(t) &= \sum_{j=1}^{n} a_{i,j} (\lambda_i(t) - \lambda_j(t)), \\
x_i(0) &= x_{i0} \in \Omega_i \subset \mathbb{R}^q, \quad \lambda_i(0) = \lambda_{i0} \in \mathbb{R}^m, \quad z_i(0) = z_{i0} \in \mathbb{R}^m, \quad \sum_{i=1}^{n} d_i = d_0, \quad \text{and} \quad a_{i,j} \text{ is the } (i,j)\text{th element of the adjacency matrix of graph } G.
\end{align*}
\]

where \( t \geq 0, i \in \{1, \ldots, n\}, x_i(0) = x_{i0} \in \Omega_i \subset \mathbb{R}^q, \lambda_i(0) = \lambda_{i0} \in \mathbb{R}^m, z_i(0) = z_{i0} \in \mathbb{R}^m, \sum_{i=1}^{n} d_i = d_0, \) and \( a_{i,j} \) is the \((i,j)\)th element of the adjacency matrix of graph \( G \). However, this algorithm involves the projection of subdifferential set (from \( \partial f^i(x_i) \) to \( \Omega_i \)), which makes its convergence analysis very hard in the nonsmooth case. To overcome the technical challenges, we propose two different ideas to construct effective algorithms for the EMO problem in the following two subsections.

A. Distributed Projected Output Feedback Algorithm (DPOFA)

The first idea is to use an auxiliary variable to avoid the projection of subdifferential set in the algorithm for EMO problem [14]. 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:

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

with the auxiliary variable \( y_i(t) \) for \( t \geq 0, y_i(0) = y_{i0} \in \mathbb{R}^q, i \in \{1, \ldots, n\} \) and other notations are kept the same as those for (10). The term \( x_i(t) = P_{\Omega_i}(y_i(t)) \) is viewed as “projected output feedback”, which is inspired by [11]. In this way, we avoid the technical difficulties resulting from the projection of the subdifferential.

Let \( x \triangleq [x_1^T, \ldots, x_n^T]^T \in \Omega \subset \mathbb{R}^{\sum_{i=1}^{n} q_i}, y \triangleq [y_1^T, \ldots, y_n^T]^T \in \mathbb{R}^{\sum_{i=1}^{n} q_i}, \lambda \triangleq [\lambda_1^T, \ldots, \lambda_n^T]^T \in \mathbb{R}^{nm}, d \triangleq [d_1^T, \ldots, d_n^T]^T \in \mathbb{R}^{nm}, \) and \( z \triangleq [z_1^T, \ldots, z_n^T]^T \in \mathbb{R}^{nm}, \) where \( \Omega \triangleq \prod_{i=1}^{n} \Omega_i \). Let \( W = [W_1, \ldots, W_n] \in \mathbb{R}^{nm \times q}, \) and \( \mathbb{R}^{nm} \).
To be strict, we propose the algorithm for the EMO problem (4) as follows:

using the derivative feedback so as to eliminate the "trouble" caused by the projection term somehow.

**B. Distributed Derivative Feedback Algorithm (DDFA)**

functions and private constraints. Later, we will show that this algorithm can solve the EMO problem with nonsmooth objective

matrix of graph $G$.

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 $\Omega$ for every $t \geq 0$, though $y(t)$ may be out of $\Omega$. Later, we will show that this algorithm can solve the EMO problem with nonsmooth objective functions and private constraints.

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:

$$
\begin{aligned}
\dot{x}_i(t) &\in \{ p : p = P_\Omega[x_i(t) - g_i(x_i(t)) + W_i^T \lambda_i(t)] - x_i(t), g_i(x_i(t)) \in \partial f_i(x_i(t)) \}, \\
\dot{\lambda}_i(t) &= d_i - W_i x_i(t) - \sum_{j=1}^n a_{i,j}(\lambda_i(t) - \lambda_j(t)) - \sum_{j=1}^n a_{i,j}(z_i(t) - z_j(t)) - W_i \dot{x}_i(t), \\
\dot{z}_i(t) &= \sum_{j=1}^n a_{i,j}(\lambda_i(t) - \lambda_j(t)),
\end{aligned}
$$

(15)

where all the notations remain the same as in (10). 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 $P_\Omega[x_i(t) - g_i(x_i(t)) + W_i^T \lambda_i(t)]$ in the analysis as demonstrated later.

Denote $x$, $\lambda$, $d$, $z$, $W$, and $W$ as in (14). Then (15) can be written in a compact form

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

(16)

$$
\mathcal{F}(x, \lambda, z) \triangleq \left\{ \begin{bmatrix}
d - W x - L \lambda - Lz - W p \\
L \lambda
\end{bmatrix} : p = P_\Omega[x - g(x) + W^T \lambda] - x, g(x) \in \partial f(x) \right\},
$$

(17)
where $x(0) = x_0 \in \Omega$, $\lambda(0) = \lambda_0 \in \mathbb{R}^{nm}$, $z(0) = z_0 \in \mathbb{R}^{nm}$, $L = L_n \otimes I_m \in \mathbb{R}^{nm \times nm}$, and $L_n \in \mathbb{R}^{n \times n}$ is the Laplacian matrix of graph $G$.

**Assumption 4.1:** For algorithm (15), or equivalently, algorithm (16), the set-valued map $F(x, \lambda, z)$ is with convex values for all $(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$.

Note that Assumption 4.1 is given to guarantee the existence of solutions to algorithm (15). In fact, in many situations, it can be satisfied. For example, it holds if $x_i \in \mathbb{R}$, or if both $\Omega_i$ and $\partial f_i(x_i)$ are “boxes” for all $i \in \{1, \ldots, n\}$.

Besides the different techniques in algorithms (11) and (15), the proposed algorithms are observed to be 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 (15) 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 5.1 in Section V).

- The dynamic performances may be different: Because algorithm (15) directly changes $x(t) \in \Omega$ to estimate the optimal solution, it may show faster response speed of $x(t)$ than that of algorithm (11) (see simulation results in Section VI-A).

Furthermore, algorithms (11) and (15) are essentially different from existing ones. Compared with the algorithm in [31], our algorithms need not exchange information of subgradients among the agents. Unlike the algorithm given in [3], ours use different techniques (i.e., the projected output feedback in (11) or derivative feedback in (15)) 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 existing results with assuming that each agent knows a subset of rows of the equality constraints [29]. 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 [29].

- 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 $i \in \{1, \ldots, n\}$ and the information of $x_i \in \mathbb{R}^{q_i}$ is kept confidential.
V. CONVERGENCE ANALYSIS

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

A. Convergence Analysis of DPOFA

Consider algorithm (11) (or (12)). 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 (11). Then

\[0_{\sum_{i=1}^n q_i} \in \{p : p = -y^* + x^* - g(x^*) + W^T \lambda^*, x^* = P_\Omega(y^*), g(x^*) \in \partial f(x^*)\}, \quad (18a)\]

\[0_{nm} = d - Wx^* - Lz^*, \quad x^* = P_\Omega(y^*), \quad (18b)\]

\[0_{nm} = L\lambda^*. \quad (18c)\]

The following result reveals the relationship between the equilibrium points of algorithm (11) and the solutions of problem (4).

**Theorem 5.1:** 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 (11), then \(x^* = P_\Omega(y^*)\) is a solution to problem (4). Conversely, if \((y^*, \lambda^*, z^*) \in \mathbb{R}^{\sum_{i=1}^n q_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}\) such that \((y^*, \lambda^*, z^*)\) is an equilibrium of (11) 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 (11). Left-multiply both sides of (18a) by \(1_n^T \otimes I_m\), it follows that

\[1_n^T \otimes I_m (d - Wx^* - Lz^*) = \sum_{i=1}^n (d_i - W_i x_i^*) - (1_n^T \otimes I_m) Lz^*\]

\[= d_0 - Wx^* - (1_n^T \otimes I_m) Lz^* = 0_m. \quad (19)\]

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

\[(1_n^T \otimes I_m) L = (1_n^T \otimes I_m) (L_n \otimes I_m) = (1_n^T L_n) \otimes (I_m) = 0_{nm \times nm}. \quad (20)\]

In the light of (19) and (20), (8) holds.

Next, it follows from (18c) that there exists \(\lambda_0^* \in \mathbb{R}^m\) such that \(\lambda^* = 1_n \otimes \lambda_0^*\). By taking into consideration (18a) and \(\lambda^* = 1_n \otimes \lambda_0^*\), there exists \(g(x^*) \in \partial f(x^*)\) such that \(x^* - g(x^*) + W^T (1_n \otimes \lambda_0^*) = x^* - g(x^*) + W^T \lambda_0^* = y^*.\) Since \(x^* = P_\Omega(y^*)\), it follows that (7) holds.

By virtue of (7), (8) and Lemma 3.1, \(x^*\) is the solution to problem (4).

(ii) Conversely, suppose \(x^*\) is the solution to problem (4). According to Lemma 3.1, there exist \(\lambda_0^* \in \mathbb{R}^q\) and \(g(x^*) \in \partial f(x^*)\) such that (7) and (8) hold. Define \(\lambda^* = 1_n \otimes \lambda_0^*.\) As a result, (18c) holds.
Take any \( v \in \mathbb{R}^m \) and let \( v = 1_n \otimes v \). Since (8) holds, \( (d - \overline{W}x^*)^T v = \left( \sum_{i=1}^n (d_i - W_i^x) \right)^T v = 0 \). Due to the properties of Kronecker product and \( L_n 1_n = 0_n, \) \( L \mathbf{v} = (L_n \otimes I_m)(1_n \otimes \mathbf{v}) = (L_n 1_n) \otimes (I_m \mathbf{v}) = 0_n \otimes \mathbf{v} = 0_{nm} \) 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 \([32]\). It follows from \( (d - \overline{W}x^*)^T \mathbf{v} = 0 \) and \( \mathbf{v} \in \ker(L) \) that \( \mathbf{d} - \overline{W}x^* \in \text{range}(L) \). Hence, there exists \( \mathbf{z}^* \in \mathbb{R}^{nm} \) such that (18b) holds.

Because \( W^T \lambda_0^* = \overline{W}^T (1_n \otimes \lambda_0^*) = \overline{W}^T \lambda^*, \) (7) implies \( x^* = P_\Omega(x^* - g(x^*) + \overline{W}^T (1_n \otimes \lambda_0^*)) = P_\Omega(x^* - g(x^*) + \overline{W}^T \lambda^*) \) for some \( g(x^*) \in \partial f(x^*) \). Choose \( y^* = x^* - g(x^*) + \overline{W}^T \lambda^* \). (18a) 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 (18) holds and \( x^* = P_\Omega(y^*) \). Hence, \( (y^*, \lambda^*, z^*) \) is an equilibrium of (11) with \( x^* = P_\Omega(y^*) \).

Let \( x^* \) be the solution to problem (4). It follows from Theorem 5.1 that there exist \( y^*, \lambda^* \) and \( z^* \) such that \( (y^*, \lambda^*, z^*) \) is an equilibrium of (11) with \( x^* = P_\Omega(y^*) \). Define the function

\[
V(y, \lambda, z) = \frac{1}{2} \left( \|y - P_\Omega(y^*)\|^2 - \|y - P_\Omega(y)\|^2 \right) + \frac{1}{2} \|\lambda - \lambda^*\|^2 + \frac{1}{2} \|z - z^*\|^2.
\]

(21)

**Lemma 5.1:** Consider algorithm (11). Under Assumption 3.1 with \( V(y, \lambda, z) \) defined in (21), 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 \( a \leq -(x - x^*)^T (g(x) - g(x^*)) - \lambda^T L \lambda \leq 0 \).

**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 \( \lambda \) 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 (11) satisfies that

\[
\mathcal{L}_F V(y, \lambda, z) = \left\{ a \in \mathbb{R} : a = \nabla_y V(y, \lambda, z)^T (-y + x - g(x)) + \overline{W}^T \lambda \right\}.
\]

(22)

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

\[
a = (x - x^*)^T (-y + x - g(x)) + \overline{W}^T \lambda + (\lambda - \lambda^*)^T (d - \overline{W}x - L \lambda - L z) + (z - z^*)^T L \lambda,
\]

where \( x = P_\Omega(y) \).
Because \((y^*, \lambda^*, z^*)\) is an equilibrium of (11) with \(x^* = P_{\Omega}(y^*)\), there exists \(g(x^*) \in \partial f(x^*)\) such that

\[
\begin{aligned}
0_{nm} &= L\lambda^* \\
\frac{d}{dt} &= \overline{W}x^* + Lz^*, \\
0\sum_{i=1}^{n} q_i &= -y^* + x^* - g(x^*) + \bar{W}^T \lambda^*.
\end{aligned}
\]  

(23)

It follows from (22) and (23) that

\[
a = (x - x^*)^T \left[ (-y + x - g(x) + \overline{W}^T \lambda) - (-y^* + x^* - g(x^*) + \overline{W}^T \lambda^*) \right]
\]

\[
+ (\lambda - \lambda^*)^T (\overline{W}x^* + Lz^* - \overline{W}x - L\lambda - Lz) + (z - z^*)^T L\lambda
\]

\[
= -(x - x^*)^T (y - y^*) + \|x - x^*\|^2 - (x - x^*)^T (g(x) - g(x^*)) + (x - x^*)^T \overline{W}^T (\lambda - \lambda^*)
\]

\[
- (x - x^*)^T \overline{W}^T (\lambda - \lambda^*) - \lambda^T L\lambda - \lambda^T L(z - z^*) + (z - z^*)^T L\lambda
\]

\[
= -(x - x^*)^T (y - y^*) + \|x - x^*\|^2 - (x - x^*)^T (g(x) - g(x^*)) - \lambda^T L\lambda.
\]  

(24)

Since \(x = P_{\Omega}(y)\) and \(x^* = P_{\Omega}(y^*)\), we obtain from Lemma 2.3 that \(-(x - x^*)^T (y - y^*) + \|x - x^*\|^2 \leq 0\). Hence,

\[
a \leq -(x - x^*)^T (g(x) - g(x^*)) - \lambda^T L\lambda.
\]

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, \(a \leq -(x - x^*)^T (g(x) - g(x^*)) - \lambda^T L\lambda \leq 0\).

The following result shows the correctness of the proposed algorithm.

**Theorem 5.2**: Consider algorithm (11). If Assumption 3.1 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 (21). It follows from Lemma 5.1 that

\[
\max L_x V(y, \lambda, z) \leq \max \{ -(x - x^*)^T (g(x) - g(x^*)) - \lambda^T L\lambda : g(x) \in \partial f(x) \} \leq 0.
\]  

(25)

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 (25) 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

\[
\|x(t) - g(x(t)) + \overline{W}^T \lambda(t)\| < m,
\]  

(26)

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 (11) satisfies that
\[
\mathcal{L}_F X(y) = \{ y^T(-y + x - g(x) + \nabla^T \lambda) : g(x) \in \partial f(x) \}.
\]
Note that \( y^T(-y(t) + x(t) - g(x(t)) + \lambda(t)) \leq -\| y(t) \|^2 + m\| y(t) \|, \) where \( t \geq 0, \ m \) is defined by (26) and \( g(x(t)) \in \partial f(x(t)) \). Hence,
\[
\max \mathcal{L}_F X(y(t)) \leq -\| y(t) \|^2 + m\| y(t) \| = -2X(y(t)) + m\sqrt{2X(y(t))}.
\]
It can be easily verified that \( X(y(t)), t \geq 0, \) is bounded, so is \( y(t), t \geq 0 \).

Part (i) is thus proved.

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

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, \( \mathcal{R} \subset \{ (y, \lambda, z) \in \mathbb{R}^{\sum_{i=1}^{n} q_i} \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : x = P_\Omega(y) = x^*, L\lambda = 0 \} \).

Let \( \mathcal{M} \) be the largest weakly invariant subset of \( \overline{\mathcal{R}} \). 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 \( t \rightarrow \infty \).

Part (ii) is thus proved.

B. Convergence Analysis of DDFA

Consider algorithm (15) (or (16)). \( (x^*, \lambda^*, z^*) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \) is an equilibrium of (15) if and only if there exists \( g(x^*) \in \partial f(x^*) \) such that
\[
0_{\sum_{i=1}^{n} q_i} = P_\Omega[x^* - g(x^*) + \nabla^T \lambda^*] - x^*, \tag{27a}
\]
\[
0_{nm} = d - \nabla x^* - Lz^*, \tag{27b}
\]
\[
0_{nm} = L\lambda^*. \tag{27c}
\]

Theorem 5.3: If \( (x^*, \lambda^*, z^*) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} \) is an equilibrium of (15), then \( x^* \) is a solution to problem (4). Conversely, if \( x^* \in \Omega \) is a solution to problem (4), then there exists \( \lambda^* \in \mathbb{R}^{nm} \) and \( z^* \in \mathbb{R}^{nm} \) such that \( (x^*, \lambda^*, z^*) \) is an equilibrium of (15).

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 (15). Define the function
\[
V(x, \lambda, z) = f(x) - f(x^*) + (\lambda^*)^T(d - \nabla x) + \frac{1}{2}\| x - x^* \|^2 + \frac{1}{2}\| \lambda - \lambda^* \|^2 + \frac{1}{2}\| z - z^* \|^2. \tag{28}
\]
Lemma 5.2: Let function $V(x, \lambda, z)$ be as defined in (28) and Assumption 5.1 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 $V(x, \lambda, z) \to \infty$ as $(x, \lambda, z) \to \infty$.

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 $(\lambda^*)^T (d - Wx) = (\lambda^0)^T (d_0 - Wx)$. It follows from (8) that

$$f(x) - f(x^*) + (\lambda^*)^T (d - Wx) = f(x) - f(x^*) + (\lambda^0)^T W (x^* - x).$$

(29)

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 (9),

$$(g(x^*) - W^T \lambda^0)^T (x - x^*) \geq 0$$

(30)

for all $x \in \Omega$, where $g(x^*) \in \partial f(x^*)$ is as chosen in (9). It follows from (29) and (30) that

$$f(x) - f(x^*) + (\lambda^*)^T (d - Wx) \geq (g(x^*) - W^T \lambda^0)^T (x - x^*) \geq 0$$

(31)

for all $x \in \Omega$, where $g(x^*) \in \partial f(x^*)$ is as chosen in (9).

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 $V(x, \lambda, z) \to \infty$ as $(x, \lambda, z) \to \infty$ for all $(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}$.

Lemma 5.3: Consider algorithm (15). Under Assumptions 5.1 and 4.1 with $V(x, \lambda, z)$ defined in (28), if $a \in \mathcal{L}_F V(x, \lambda, z)$, then there exist $g(x) \in \partial f(x)$ and $g(x^*) \in \partial f(x^*)$ such that $a \leq -\|p\|^2 - (x - x^*)^T (g(x) - g(x^*)) - \lambda^T L \lambda \leq 0$, where $p = P_{\Omega} [x - g(x) + W^T \lambda] - x$.

Proof: The function $V(x, \lambda, z)$ along the trajectories of (15) satisfies that

$$\mathcal{L}_F V(x, \lambda, z) = \left\{ a \in \mathbb{R} : a = (g(x) - W^T \lambda^* + x - x^*)^T p + \nabla \lambda V(x, \lambda, z)^T (d - Wx - L\lambda - Lz - Wp) + \nabla z V(x, \lambda, z)^T L \lambda, \right.$$ 

$$g(x) \in \partial f(x), p = P_{\Omega} [x - g(x) + W^T \lambda] - x \right\}.$$ 

Suppose $a \in \mathcal{L}_F V(x, \lambda, z)$, then there exists $g(x) \in \partial f(x)$ such that

$$a = (g(x) - W^T \lambda^* + x - x^*)^T p + (\lambda - \lambda^*)^T (d - Wx - L\lambda - Lz - Wp) + (z - z^*)^T L \lambda,$$

(32)

where

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

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

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

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

\[ (g(x) - W^T \lambda + x - x^*)^T p \leq -\|p\|^2 - (g(x) - W^T \lambda)^T (x - x^*). \quad (34) \]

Since \((x^*, \lambda^*, z^*)\) is an equilibrium of (15), there is \( g(x^*) \in \partial f(x^*) \) such that

\[
\begin{cases}
0_{nm} = LA^* \\
d = \tilde{W} x^* + L z^*, \\
x^* = P_{\Omega}[x^* - g(x^*) + W^T \lambda^*].
\end{cases}
\]

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

\[
a = (g(x) - W^T \lambda + x - x^*)^T p + (\lambda - \lambda^*)^T W p + (\lambda - \lambda^*)^T (W x + L z - W p) + (z - z^*)^T L \lambda
\]

\[
\leq -\|p\|^2 - (g(x) - W^T \lambda)^T (x - x^*) + (\lambda - \lambda^*)^T W p - (x - x^*)^T W^T (\lambda - \lambda^*) - \lambda^T L \lambda
\]

\[
-\lambda^T L (z - z^*) - (\lambda - \lambda^*)^T W p + (z - z^*)^T L \lambda
\]

\[
= -\|p\|^2 - (g(x) - g(x^*))^T (x - x^*) - (x - x^*)^T W^T (\lambda - \lambda^*) - \lambda^T L \lambda
\]

\[
= -\|p\|^2 - (g(x) - g(x^*))^T (x - x^*) - (g(x^*) - W^T \lambda^*)^T (x - x^*) - \lambda^T L \lambda. \quad (36)
\]

Because \( x^* = P_{\Omega}[x^* - g(x^*) + W^T \lambda^*], (g(x^*) - W^T \lambda^*)^T (x - x^*) \geq 0 \) 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^*)) - \lambda^T L \lambda \leq 0. \]

**Theorem 5.4:** Consider algorithm (15). If Assumptions 3.1 and 4.1 hold, 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).

**Proof:** (i) Suppose \((x^*, \lambda^*, z^*) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm}\) is an equilibrium of (15). Let function \( V(x, \lambda, z) \) be as defined in (28). It follows from Lemma 5.3 that

\[
\max L_x V(x, \lambda, z) \leq \sup \{ a : a = -\|p\|^2 - (x - x^*)^T (g(x) - g(x^*)) - \lambda^T L \lambda, \quad g(x) \in \partial f(x), \quad p = P_{\Omega}[x - g(x) + W^T \lambda] - x \} \leq 0.
\]

It follows from Lemma 5.2 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 \( V(x, \lambda, z) \to \infty \) as \((x, \lambda, z) \to \infty\). Hence, \((x(t), \lambda(t), z(t))\) is bounded for all \( t \geq 0 \).
(ii) Let \( \mathcal{R} = \{(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : 0 \in \mathcal{L}_x V(y, \lambda, z)\} \subset \{(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : \exists g(x) \in \partial f(x), L\lambda = 0_{nq}, (x-x^*)^T(g(x) - g(x^*)) = 0, 0_{\sum_{i=1}^n q_i} = P_{\Omega}[x - g(x) + \overline{W}^T \lambda] - x\}. \) Let \( \mathcal{M} \) be the largest weakly invariant subset of \( \mathcal{R} \). It follows from Lemma 2.1 that \( (x(t), \lambda(t), z(t)) \to \mathcal{M} \) as \( t \to \infty \). 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, \( x(t) \to x^* \) as \( t \to \infty \).

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 (15) 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 5.1:** Consider algorithm (15). With 1) and 3) of Assumption 3.1, if \( f_i \) is differentiable and convex on an open set containing \( \Omega_i \), and \( \nabla f_i \) is Lipschitz continuous on \( \Omega_i \) for \( i \in \{1, \ldots, n\} \), 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 5.4 (i). Hence, it is omitted.

(ii) It follows from similar arguments in the proof of Theorem 5.4 (ii) that
\[
\frac{d}{dt} V(x, \lambda, z) \leq -\|\dot{x}\|^2 - (x-x^*)^T(\nabla f(x) - \nabla f(x^*)) - \lambda^T L \lambda \leq 0,
\]
where \( \dot{x} = P_{\Omega}[x - \nabla f(x) + \overline{W}^T \lambda] - x \).

Let \( \mathcal{R} = \{(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : 0 = \frac{d}{dt} V(y, \lambda, z)\} \subset \{(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : L\lambda = 0_{nm}, (x-x^*)^T(\nabla f(x) - \nabla f(x^*)) = 0, 0_{\sum_{i=1}^n q_i} = P_{\Omega}[x - \nabla f(x) + \overline{W}^T \lambda] - x\} \).

Let \( \mathcal{M} \) be the largest invariant subset of \( \mathcal{R} \). It follows from the invariance principle (Theorem 2.41 of [33]) that \( (x(t), \lambda(t), z(t)) \to \mathcal{M} \) as \( t \to \infty \). Note that \( \mathcal{M} \) is invariant. The trajectory \( (\bar{x}(t), \bar{\lambda}(t), \bar{z}(t)) \in \mathcal{M} \) 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, \( \bar{\lambda}(t) \equiv \frac{d}{dt} \bar{z}(t) \equiv 0_{nm} \). Suppose \( \bar{\lambda}(t) \equiv \frac{d}{dt} \bar{z}(t) \equiv 0_{nm} \), then \( \bar{\lambda}(t) \to \infty \) as \( t \to \infty \), which contradicts part (i). Hence, \( \bar{\lambda}(t) \equiv 0_{nm} \) and \( \mathcal{M} \subset \{(x, \lambda, z) \in \Omega \times \mathbb{R}^{nm} \times \mathbb{R}^{nm} : P_{\Omega}[x - \nabla f(x) + \overline{W}^T \lambda] - x = 0_{\sum_{i=1}^n q_i}, d - \overline{W} x - Lz = 0_{nm}, L\lambda = 0_{nm}\} \).

Take any \((\bar{x}, \bar{\lambda}, \bar{z}) \in \mathcal{M} \). Obviously, \((\bar{x}, \bar{\lambda}, \bar{z}) \) is an equilibrium point of algorithm (15). Define a new function \( \tilde{V}(x, \lambda, z) \) by replacing \((x^*, \lambda^*, z^*) \) with \((\bar{x}, \bar{\lambda}, \bar{z}) \) in \( V(x, \lambda, z) \). It follows from similar arguments in the proof of Lemma 5.3 that \( \frac{d}{dt} \tilde{V}(x, \lambda, z) \leq 0 \). Hence, \((\bar{x}, \bar{\lambda}, \bar{z}) \) is Lyapunov stable. By Proposition 4.7 of [33], there exists \((\bar{x}, \bar{\lambda}, \bar{z}) \in \mathcal{M} \) such that \( (x(t), \lambda(t), z(t)) \to (\bar{x}, \bar{\lambda}, \bar{z}) \) as \( t \to \infty \).
Since $(\bar{x}, \bar{\lambda}, \bar{z}) \in \mathcal{M}$ is an equilibrium point of algorithm (15), $\bar{x}$ is an optimal solution to problem (4) by Theorem 5.3.

VI. NUMERICAL EXPERIMENTS

In this section, we give two numerical examples to illustrate the effectiveness of the proposed algorithms.

A. Nonsmooth Optimization Problem

Consider the following nonsmooth optimization problem

$$
\min f(x), \quad f(x) = \sum_{i=1}^{n} (|x_i|^2 + |x_i|), \quad \sum_{i=1}^{n} A_i x_i = d_0, \quad |x_i| \leq 1, \quad i \in \{1, \ldots, n\},
$$

(38)

where $x = [x_1, \ldots, x_n]^T \in \mathbb{R}^n$ and $x_i \in \mathbb{R}$. Let $n = 10$, $d_0 = [3, 2]^T$, $d_i = [0.3, 0.2]^T$, $i \in \{1, \ldots, 10\}$, and

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

Take a network of 10 agents interacting over a graph $G$ to solve this problem. The information sharing graph $G$ of the optimization algorithms is given in Fig. 1. Some simulation results using DPOFA algorithm (11) proposed in Section IV-A and DDFA algorithm (15) proposed in Section IV-B are shown in Figs. 2-6.

Figs. 2 and 3 show the trajectories of estimates for $x$ and the auxiliary variable $y$ versus time under DPOFA algorithm (11) proposed in Section IV-A, respectively, and Fig. 4 depicts the trajectories of the estimates for $x$ versus time under DDFA algorithm (15) proposed in Section IV-B.
(11) proposed in Section IV-A uses an auxiliary variable $y$ and estimates the optimal solution using $x = P_\Omega(y)$, while algorithm (15) proposed in Section IV-B directly uses $x$ to estimate the solution. Both algorithms are able to find the optimal solution of the optimization problem. Figs. 2-4 indicate that the trajectories of $y$ may be out of the constraint set $\Omega$, but the trajectories of $x$ stays in the constraint set $\Omega$.

Fig. 5 gives the trajectories of the objective function $f(x)$ and constraint $\|Wx - d_0\|^2$ versus time under DPOFA algorithm (11) and DDFA algorithm (15), and demonstrates that the trajectories of $x$ converge to the equality constraint. Furthermore, Fig. 6 verifies the boundedness of the trajectories of auxiliary variables $\lambda$ and $z$.

In Fig. 5, the trajectories of $f(x)$ and $\|Wx - d_0\|^2$ versus time under 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 $y \notin \Omega$ (see the trajectories at the beginning of the simulation in Figs. 2 and 3). Due to the indirect feedback effect on $x$ (changing $x$ by controlling $y$) in DPOFA algorithm, the trajectory of variable $x$ may show slow changing behaviors in applications.

B. Multi-commodity Network Flow Problem

1) Problem Description: Consider a directed network consisting of a set $\mathcal{N} = \{1, \ldots, m\}$ of nodes and a set $\mathcal{E}$ of directed arcs. The flows on the arcs are of $S$ different types (commodities). The flows of the $s$th type on the arc $(i, j)$ is denoted by $t_{i,j}(s) \in [l_{i,j}(s), u_{i,j}(s)]$, where $[l_{i,j}(s), u_{i,j}(s)]$ is the capacity
constraint for $t_{i,j}(s)$. The flows must satisfy the conservation of flow and supply/demand constraints of the form

$$\sum_{j \mid (i,j) \in E} t_{i,j}(s) - \sum_{j \mid (j,i) \in E} t_{j,i}(s) = b_i(s),$$

(39)

where $i \in \mathcal{N}$, $s \in \{1, \ldots, S\}$, $b_i(s)$ is the amount of flow of the $s$th type entering the network at node $i$ ($b_i(s) > 0$ indicates supply, and $b_i(s) < 0$ indicates demand). The supplies/demands $b_i(s)$ are given.
and satisfy $\sum_{i \in \mathbb{N}} b_i(s) = 0$ for the feasibility of the problem, which have been studied in the literature (see [18]).

The problem is to minimize $\sum_{(i,j) \in \mathcal{E}} f_{i,j}(t_{i,j}(1), \ldots, t_{i,j}(S))$ subject to the constraints (39), where $f_{i,j} : \mathbb{R} \rightarrow \mathbb{R}$ are continuous, strictly convex functions.
2) Reformulation of Problem: Let \( n \) be the number of arcs in \( \mathcal{E} \). We assign an index \( k = k(i, j) \in \{1, \ldots, n\} \) to every arc \((i, j) \in \mathcal{E} \). We use \( x_k = [t_{i,j}(1), \ldots, t_{i,j}(S)]^T \in \mathbb{R}^S \) to denote the flow vector on arc \( k \). Then constraint (39) can be rewritten as

\[
A \otimes I_S x = \sum_{k=1}^{n} A_k \otimes I_S x_k = b,
\]

with \( A \) as the vertex-edge incidence matrix of the graph, \( A_k \) as the \( k \)th column of \( A \), \( x = [x_1^T, \ldots, x_n^T]^T \), and \( b = [b_1^T, \ldots, b_m^T]^T \).

Let \( f^k(x_k) \equiv f_{i,j}(t_{i,j}(1), \ldots, t_{i,j}(S)) \), \( \Omega_k = \prod_{s=1}^{S} [l_{i,j}(s), u_{i,j}(s)] \), where \( k \) is the index of arc \((i, j) \in \mathcal{E} \). The optimization problem can be reformulated as

\[
\min f(x) = \sum_{k=1}^{n} f^k(x_k), \quad \sum_{k=1}^{n} A_k \otimes I_S x_k = b, \quad x_k \in \Omega_k, \quad k \in \{1, \ldots, n\}.
\]

3) Numerical Simulation: Consider a network of 6 nodes and 12 arcs as shown in Fig. 7 with \( S = 1 \). Let \( f^k(x_k) = \|x_k\|^2 \), \( x_k \in \mathbb{R} \) be the flows on arc \( k \), and \( \Omega_k = [0, 10] \) for \( k \in \{1, \ldots, 12\} \). Problem (41) can be formulated as

\[
\min \sum_{k=1}^{12} \|x_k\|^2, \quad Ax = \sum_{k=1}^{12} A_k x_k = b, \quad x_k \in \Omega_k, \quad k \in \{1, \ldots, 12\},
\]

where \( b = [6, -7.2, -4.8, -9.6, 8.4, 7.2]^T \) and \( A_k \) is the \( k \)th column of the vertex-edge incidence matrix \( A \) of the network in Fig. 7.

Simulation results using DPOFA algorithm (11) proposed in Section [IV-A] and DDFA algorithm (15) proposed in Section [IV-B] are shown in Figs. 8-12.

Fig. 8 shows the trajectories of estimates for \( x \) versus time of DPOFA algorithm (11) proposed in Section [IV-A] while Fig. 9 shows those of the auxiliary variables \( y \) versus time of DPOFA algorithm (11) proposed in Section [IV-A] Fig. 10 exhibits the trajectories of estimates for \( x \) versus time of DDFA algorithm (15) proposed in Section [IV-B] Algorithm (11) proposed in Section [IV-A] uses an auxiliary variable \( y \) and estimates the optimal solution using \( x = P_{\Omega}(y) \), while algorithm (15) proposed in Section [IV-B] directly uses \( x \) to estimate the solution. Both algorithms are able to find the optimal solution of the optimization problem. Figs. 8-10 indicate that the trajectories of \( y \) may be out of the constraint set \( \Omega \), but the trajectories of \( x \) stay in the constraint set \( \Omega \).

Fig. 11 depicts the trajectories of the objective function \( f(x) \) and constraint \( \|Ax - b\|^2 \) versus time under DPOFA algorithm (11) and DDFA algorithm (15), where the trajectories of \( x \) converge to the equality constraint. Fig. 12 illustrates the boundedness of the trajectories of auxiliary variables \( \lambda \) and \( z \). Clearly, both algorithms can solve the problem.
VII. CONCLUSIONS

In this paper, the distributed design for the extended monotropic optimization (EMO) problem has been addressed, which is 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 multi-
agent networks. The design of the algorithms has used the decomposition of problem constraints and distributed techniques. Based on stability theory and invariance principle for differential inclusions, the convergence properties 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.
The distributed EMO problem definitely deserves more efforts because of its broad range of applications. In the future, distributed EMO problems with more complicated situations such as parameter uncertainties and online concerns will be further investigated.

REFERENCES

[1] Q. Liu and J. Wang, “A second-order multi-agent network for bound-constrained distributed optimization,” *IEEE Transaction on Automatic Control*, vol. 60, no. 12, pp. 3310–3315, 2015.
[2] W. Bian and X. Xue, “Subgradient-based neural networks for nonsmooth nonconvex optimization problems,” *IEEE Transactions on Neural Networks*, vol. 20, no. 6, pp. 1024–1038, 2009.

[3] P. Yi, Y. Hong, and F. Liu, “Distributed gradient algorithm for constrained optimization with application to load sharing in power systems,” *Systems & Control Letters*, vol. 83, pp. 45–52, 2015.

[4] D. Yuan, W. C. Ho, and S. Xu, “Zeroth-order method for distributed optimization with approximate projections,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 2, pp. 284–294, 2016.

[5] A. Nedic, A. Ozdaglar, and P. A. Parrilo, “Constrained consensus and optimization in multi-agent networks,” *IEEE Transactions on Automatic Control*, vol. 55, pp. 922–938, 2010.

[6] B. Gharesifard and J. Cortés, “Distributed continuous-time convex optimization on weight-balanced digraphs,” *IEEE Transactions on Automatic Control*, vol. 59, pp. 781–786, 2014.

[7] G. Shi, K. Johansson, and Y. Hong, “Reaching an optimal consensus: Dynamical systems that compute intersections of convex sets,” *IEEE Transactions on Automatic Control*, vol. 55, no. 3, pp. 610–622, 2013.

[8] Q. Liu, S. Yang, and J. Wang, “A collective neurodynamic approach to distributed constrained optimization,” *IEEE Transactions on Neural Networks and Learning Systems*, to appear.

[9] D. Tank and J. Hopfield, “Simple neural optimization networks: An a/d converter, signal decision circuit, and a linear programming circuit,” *IEEE Transactions on Circuits and Systems*, vol. 33, no. 5, pp. 533–541, 2013.

[10] E. Chong, S. Hui, and S. Zak, “An analysis of a class of neural networks for solving linear programming problems,” *IEEE Transactions on Automatic Control*, vol. 28, no. 3, pp. 36–73, 2008.

[11] Q. Liu and J. Wang, “A one-layer projection neural network for nonsmooth optimization subject to linear equalities and bound constraints,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 5, pp. 812–824, 2013.

[12] J. Park, K. Lee, J. Shin, and K. Y. Lee, “Economic load dispatch for non-smooth cost functions using particle swarm optimization,” in *IEEE Power Engineering Society 2003 General Meeting*, Toronto, Canada, 2003, pp. 938–943.

[13] E. Ramírez-Llanos and S. Martínez, “Distributed and robust resource allocation algorithms for multi-agent systems via discrete-time iterations,” in *Proc. IEEE Conf. Decision Control*, Osaka, Japan, 2015, pp. 1390–1396.

[14] 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,” *arXiv:1510.08579*, 2015.

[15] P. N. M. Forti and M. Quincampoix, “Generalized neural network for nonsmooth nonlinear programming problems,” *IEEE Transactions on Circuits and Systems-I*, vol. 51, no. 9, pp. 1741–1754, 2004.

[16] R. Rockafellar, *Network Flows and Monotropic Optimization*. New York: Wiley, 1984.

[17] ———, “Monotonic programming: A generalization of linear programming and network programming,” in *Convexity and duality in optimization*, *Lecture Notes in Economics and Mathematical Systems*, vol. 256, pp. 226–237, 1985.

[18] D. P. Bertsekas, “Extended monotropic programming and duality,” *Journal of Optimization Theory and Applications*, vol. 139, no. 2, pp. 209–225, 2008.

[19] N. Chatzipanagiotis, D. Dentcheva, and M. M. Zavlanos, “An augmented lagrangian method for distributed optimization,” *Mathematical Programming Series A*, vol. 152, no. 1, pp. 405–434, 2015.

[20] C. Godsil and G. F. Royle, *Algebraic Graph Theory*. New York: Springer-Verlag, 2001.

[21] J. P. Aubin and A. Cellina, *Differential Inclusions*. Berlin, Germany: Springer-Verlag, 1984.

[22] F. H. Clarke, *Optimization and Nonsmooth Analysis*. New York: Wiley, 1983.

[23] A. Bacciotti and F. Ceragioli, “Stability and stabilization of discontinuous systems and nonsmooth Lyapunov functions,” *ESAIM: Control, Optimisation and Calculus of Variations*, vol. 4, pp. 361–376, 1999.

[24] J. Cortés, “Discontinuous dynamical systems,” *IEEE Control Systems Magazine*, vol. 44, no. 11, pp. 1995–2006, 1999.

[25] Z. Denkowski, S. Migórski, and N. S. Papageorgiou, *An Introduction to Nonlinear Analysis: Theory*. New York, NY: Springer-Verlag New York Inc., 2003.

[26] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*. New York: Academic, 1982.

[27] Z. Qiu, S. Liu, and L. Xie, “Distributed constrained optimal consensus of multi-agent systems,” *Automatica*, vol. 68, pp. 209–215, 2016.

[28] M. Tygert, “A fast algorithm for computing minimal-norm solutions to underdetermined systems of linear equations,” *arXiv:0905.4745*, 2009.

[29] S. Mou, J. Liu, and A. S. Morse, “A distributed algorithm for solving a linear algebraic equation,” *IEEE Transactions on Automatic Control*, vol. 60, no. 11, pp. 2863–2878, 2015.

[30] A. Ruszczyński, *Nonlinear Optimization*. Princeton, New Jersey: Princeton University Press, 2006.

[31] A. Cherukuri and J. Cortés, “Distributed generator coordination for initialization and anytime optimization in economic dispatch,” *IEEE Transactions on Control of Network Systems*, vol. 2, no. 3, pp. 226–237, 2015.

[32] G. Strang, “The fundamental theorem of linear algebra,” *American Mathematical Monthly*, vol. 100, no. 9, pp. 848–855, 1993.

[33] W. M. Haddad and V. Chellaboina, *Nonlinear Dynamical Systems and Control: A Lyapunov-Based Approach*. Princeton, NJ: Princeton Univ. Press, 2008.