KYBERNETIKA — VOLUME *60* (2024), NUMBER 4, PAGES 427–445

DISTRIBUTED DUAL AVERAGING ALGORITHM FOR MULTI-AGENT OPTIMIZATION WITH COUPLED CONSTRAINTS

Zhipeng Tu and Shu Liang

This paper investigates a distributed algorithm for the multi-agent constrained optimization problem, which is to minimize a global objective function formed by a sum of local convex (possibly nonsmooth) functions under both coupled inequality and affine equality constraints. By introducing auxiliary variables, we decouple the constraints and transform the multi-agent optimization problem into a variational inequality problem with a set-valued monotone mapping. We propose a distributed dual averaging algorithm to find the weak solutions of the variational inequality problem with an $O(1/\sqrt{k})$ convergence rate, where k is the number of iterations. Moreover, we show that weak solutions are also strong solutions that match the optimal primal-dual solutions to the considered optimization problem. A numerical example is given for illustration.

Keywords: distributed optimization, coupled constraints, dual averaging, variational inequality, multi-agent networks

Classification: 90C33, 68W15

1. INTRODUCTION

Multi-agent (distributed) optimization has attracted significant attention in recent years, which has arisen in broad application fields within the information sciences and engineering, such as multi-agent coordination, distributed localization, and packet routing [3, 24, 30, 32]. Many distributed algorithms have been investigated for smooth and nonsmooth convex optimizations [12, 18, 38].

Distributed optimization with coupled constraints is challenging, since the feasible region of one agent's variable is influenced by that of other agents. To deal with only coupled equality constraints or inequality constraints, various algorithms are applicable such as gradient-decent algorithms [31], local primal-dual perturbed subgradients algorithms [5], and operator splitting methods [36]. Moreover, there have been some works proceeding two kinds of coupled constraints together. For instance, [14] constructed a distributed continuous-time algorithm by virtue of a projected primal-dual subgradient dynamics, while [15] adopted the extragradient method with a fixed stepsize to find the exact optimal solutions with the help of two rounds of communication per iteration.

DOI: 10.14736/kyb-2024-4-0427

The existing methods cannot be extended easily to distributed nonsmooth optimization with coupled constraints.

Variational inequality provides a broad and unified setting for optimization and equilibrium problems, which serves as a promising and practical tool. For example, differentiable constrained optimization problems can be transformed into variational inequality problems with single-valued mappings [9, 13], and then many effective solvers are available [1,2,33]. Nonetheless, many difficulties occur for nonsmooth optimization problems with coupled constraints. First of all, the subdifferentials of nonsmooth functions are set-valued, which yields variational inequalities with set-valued mappings. Further investigation about the equivalence between their strong solutions and weak solutions is needed. In addition, coupled constraints have to be decoupled in the transformed variational inequality problem so that distributed algorithms can be designed over multi-agent networks.

The dual averaging method (DA), which takes averages of subgradients, was firstly presented by Nesterov in [22] along with applications in finding weak solutions to the variational inequality problems with single-valued mappings. Further, a distributed-DA algorithm for convex optimization without coupled constraints was developed and analyzed in [8]. Later, Nesterov proposed the dual subgradient method with averaging (DSMA) [23], which has a similar name, while it actually takes averages of primal variables and dual variables. And then, a distributed-DSMA algorithm for convex optimization with coupled constraints was proposed and it achieved $O(\ln k/\sqrt{k})$ convergence rate [16]. To our interests, the dual averaging method has remarkable performance of convergence in nonsmooth problems, and it leads to decentralized policies that can be used over networks.

In this paper, we propose a discrete-time distributed algorithm for the distributed nonsmooth convex optimization problem with coupled constraints by developing a variational-inequality-based approach with Nesterov's dual averaging method. The main contributions are as follows.

- By decoupling the constraints, we transform the distributed nonsmooth convex optimization problem with coupled constraints into a variational inequality problem with a set-valued monotone mapping, which may provide a new way for solving complicated distributed optimization problems.
- We extend existing results to the set-valued case, and show that the weak solutions and strong solutions to the obtained variational inequality problem are the same. We also extend the Nesterov's dual averaging method for general variational inequality problems.
- We propose a particular distributed algorithm and obtain the convergence rate of $O(1/\sqrt{k})$, which matches the optimal convergence rate of the first-order methods for nonsmooth convex optimization.

The remainder of this paper is organized as follows. Section 2 provides necessary notations, definitions, and preliminaries, while Section 3 formulates a distributed optimization problem with coupled constraints. Then Section 4 presents the main results, including three theorems. Section 5 provides a numerical example for illustration, and finally, Section 6 gives some concluding remarks.

2. MATHEMATICAL PRELIMINARIES

In this section, we introduce preliminaries about optimization, variational inequality, graph theory, and Nesterov's dual averaging scheme.

2.1. Notations

Notations $\mathbb{R}_{\leq 0}$, $\mathbb{R}_{\geq 0}$, and $\mathbb{R}$ denote the sets of nonpositive, nonnegative, and all real numbers, respectively. $\mathbb{R}^n$ denotes the set of n-dimensional real column vectors. Let $\mathbf{0}_n$ and $\mathbf{1}_n$ be the column vectors with n components being zero and one, respectively. Let $\mathrm{col}\{x_1, \ldots, x_N\}$ be the column vector stacked with $x_1, \ldots, x_N$. Furthermore, $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ denote the Euclidean inner product and Euclidean norm in $\mathbb{R}^n$. $\mathbb{R}^{p\times q}$ denotes the set of real number matrices with p rows and q columns. For $M \in \mathbb{R}^{p\times q}$, notations $\mathrm{rank}(M)$, $\ker(M)$, $M^\top$, and $M^\dagger$ denote the rank, kernel, transpose, and generalized inverse of M, respectively. Let I_n be the identity matrix in $\mathbb{R}^{n\times n}$. $\otimes$ denotes the Kronecker product for two matrices.

2.2. Some preliminaries on optimization

A nonempty set $\Omega \subset \mathbb{R}^n$ is said to be *convex* if $\lambda x' + (1-\lambda)x \in \Omega$ for any $x, x' \in \Omega$ and $\lambda \in [0,1]$. The *normal cone* to Ω at the point $x \in \Omega$ is

$$\mathcal{N}_\Omega(x) := \{v \in \mathbb{R}^n | \, \langle v, x' - x\rangle \leq 0, \forall x' \in \Omega\}.$$

On a convex set $\Omega \subset \mathbb{R}^n$, a function $f : \Omega \to \mathbb{R}$ is said to be *convex* if $f(\lambda x' + (1-\lambda)x) \leq \lambda f(x') + (1-\lambda)f(x)$ and *m-strongly convex* if $f(\lambda x' + (1-\lambda)x) \leq \lambda f(x') + (1-\lambda)f(x) - \frac{m}{2}\lambda(1-\lambda)\|x - x'\|^2$ for any $x, x' \in \Omega$ and $\lambda \in [0,1]$. f is called *proper* if $f(x) > -\infty$ for any $x \in \Omega$ and $f(x_0) < +\infty$ for some $x_0 \in \Omega$. The *subdifferential* of a (possibly nonsmooth) convex function f at $x \in \Omega$ is defined by

$$\partial f(x) := \{g \in \mathbb{R}^n \,|\, f(x') \geq f(x) + \langle g, x' - x\rangle, \, \forall x' \in \Omega\}.$$

Given a convex function $f : \Omega \to \mathbb{R}$, for any $x, x' \in \Omega$ and any $g \in \partial f(x), g' \in \partial f(x')$, the following two results hold

$$\langle x - x', g - g'\rangle \geq 0, \tag{1}$$

$$D_f^g(x', x) := f(x') - f(x) - \langle g, x' - x\rangle \geq 0. \tag{2}$$

Let $\boldsymbol{f}$ be a vector-function from a convex set $\Omega \subset \mathbb{R}^n$ to $\mathbb{R}^m$. The *subdifferential* of $\boldsymbol{f}$ at $x \in \Omega$ is defined by

$$\partial \boldsymbol{f}(x) := \{G \in \mathbb{R}^{m\times n} \,|\, \boldsymbol{f}(x') - \boldsymbol{f}(x) - G(x' - x) \succeq \mathbf{0}_m, \, \forall x' \in \Omega\},$$

where $v \succeq \mathbf{0}_m$ means that all components of the vector v are nonnegative.

The *(extended-valued) indicator function* of the convex set $\Omega \subset \mathbb{R}^n$ is defined as [4]

$$\mathcal{I}_\Omega(x) := \begin{cases} 0 & \text{if } x \in \Omega \\ +\infty & \text{if } x \notin \Omega \end{cases},$$

and its subdifferential is $\partial \mathcal{I}_\Omega(x) = \mathcal{N}_\Omega(x)$. The *projection* operator is defined as

$$\Pi_\Omega^\psi(z,\alpha) := \underset{x\in\Omega}{\operatorname{argmin}} \left\{ \langle z, x\rangle + \frac{1}{\alpha}\psi(x) \right\}, \quad \alpha > 0, \tag{3}$$

where $\psi : \Omega \to \mathbb{R}_{\geq 0}$ is an arbitrary strongly convex and nonnegative function. In this paper, we take $\psi(x) = \frac{1}{2}\|x\|^2$. The projection $\Pi_\Omega^\psi(z,\alpha)$ is $\alpha-$Lipschitz continuous [8], that is, for any pair $z, z' \in \Omega$,

$$\|\Pi_\Omega^\psi(z,\alpha) - \Pi_\Omega^\psi(z',\alpha)\| \leq \alpha\|z - z'\|. \tag{4}$$

2.3. Variational inequality problems with set-valued mappings

Let Ω be a closed convex subset in $\mathbb{R}^n$ with nonempty interior, and $F : \Omega \to 2^{\mathbb{R}^n}$ be a set-valued mapping. For any $x \in \Omega$, $F(x)$ is a set in $\mathbb{R}^n$. The graph of the operator F is defined as

$$\mathcal{G}(F) := \{(x,y) \,|\, x \in \Omega,\, y \in F(x)\}.$$

An operator F is said to be *monotone* if for any $(x,y), (x',y') \in \mathcal{G}(F)$,

$$\langle x - x', y - y'\rangle \geq 0. \tag{5}$$

Furthermore, F is called *maximal monotone* if it is impossible to find a pair (x,y) not belonging to $\mathcal{G}(F)$ such that the extended mapping $x \mapsto F(x) \cup \{y\}$ is monotone.

Lemma 2.1. Let f be a lower semi-continuous, proper, and convex function. Then ∂f is maximal monotone [26].

A *variational inequality* problem, denoted by the pair VI(domain, operator), is the classical problem that

$$\begin{aligned} &\text{find } x^* \in \Omega \text{ and } y^* \in F(x^*), \text{ such that} \\ &\quad \langle y^*, x - x^*\rangle \geq 0,\ \forall x \in \Omega. \end{aligned} \tag{6}$$

A point x^* satisfying (6) is called a *strong solution* to the VI(Ω, F) [21]. In addition, a point x^* is called a *weak solution* if

$$\langle y, x - x^*\rangle \geq 0,\ \forall x \in \Omega,\ \forall y \in F(x). \tag{7}$$

The *merit function* that measures the distance between a point x and the weak solution set is defined as

$$\mathcal{V}(x) := \sup\{\langle y', x - x'\rangle \,|\, x' \in \Omega, y' \in F(x')\}. \tag{8}$$

Lemma 2.2. For any vector $x \in \Omega$, $\mathcal{V}(x) \geq 0$. x^* is a weak solution to the variational inequality problem if and only if $\mathcal{V}(x^*) = 0$ [22].

2.4. Some preliminaries on graph theory

An undirected graph is denoted as $\mathbb{G}(\mathbb{V}, \mathbb{E})$, where $\mathbb{V} = \{1, \ldots, N\}$ is the set of nodes, representing the set of agents, and $\mathbb{E} \subset \mathbb{V} \times \mathbb{V}$ is the set of edges. Let $A_N = [a_{ij}] \in \mathbb{R}^{N \times N}$ be the adjacency matrix of $\mathbb{G}$ such that $a_{ij} = a_{ji}$. If $(i, j) \in \mathbb{E}$, then node i and node j can exchange information, and $a_{ij} = 0$ otherwise. Denote by $\mathcal{N}_i \subset \mathbb{V}$ the neighbors of node i. We also assume that there are no self-loops, that is $a_{ii} = 0$. The graph Laplacian matrix is $L_N = D_N - A_N$, where $D_N = [d_{ij}] \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $d_{ii} = \sum_{j=1}^{N} a_{ij}$. A path between nodes i and j is defined as a sequence of edges $(i, i_1), (i_1, i_2), \ldots, (i_k, j) \in \mathbb{E}$ with distinct nodes $i_l \in \mathbb{V}$. The graph is connected if there exists a path between any pair of distinct nodes i and j. Specifically, if the undirected graph $\mathbb{G}$ is connected, then $L_N = L_N^\top \geq 0$, $\operatorname{rank}(L_N) = N - 1$, and $\ker(L_N) = \{k\mathbf{1}_N : k \in \mathbb{R}\}$.

2.5. Dual averaging method

Given $\Omega \subset \mathbb{R}^n$ and a single-valued mapping $F : \Omega \to \mathbb{R}^n$, Nesterov's dual averaging scheme [22] for computing a weak solution to the $\mathrm{VI}(\Omega, F)$ is

$$x[k+1] := \Pi_\Omega^\psi \left(\sum_{s=1}^{k} F(x[s]), \alpha_k \right), \tag{9}$$

where α_k is a stepsize. An important result of the dual averaging scheme is as follows.

Lemma 2.3. For any non-increasing sequence $\{\alpha_k\}_{k=0}^{\infty}$ of positive stepsizes, and for any $x^* \in \Omega$,

$$\sum_{k=1}^{K} \langle F(x[k]), x[k] - x^* \rangle \leq \sum_{k=1}^{K} \frac{\alpha_{k-1} \|F(x[k])\|^2}{2} + \frac{\psi(x^*)}{\alpha_K}. \tag{10}$$

3. PROBLEM FORMULATION

Consider a network of N agents described by an undirected graph $\mathbb{G}(\mathbb{V}, \mathbb{E})$. For each $i \in \mathbb{V}$, the ith agent has a local decision variable $x_i \in \mathbb{R}^{n_i}$, a local feasible set $\Omega_i \subset \mathbb{R}^{n_i}$, a local objective function $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$, and local constraint functions $\boldsymbol{g}_i : \mathbb{R}^{n_i} \to \mathbb{R}^p, \boldsymbol{h}_i : \mathbb{R}^{n_i} \to \mathbb{R}^q$. Notice that each agent can not have access to other agents' objective functions and constraints. N agents over the network cooperatively solve the following optimization problem

$$\begin{aligned} \min_{\boldsymbol{x} \in \boldsymbol{\Omega}} \quad & f(\boldsymbol{x}) := f_1(x_1) + f_2(x_2) + \cdots + f_N(x_N), \\ \text{s.t.} \quad & \boldsymbol{g}(\boldsymbol{x}) := \boldsymbol{g}_1(x_1) + \boldsymbol{g}_2(x_2) + \cdots + \boldsymbol{g}_N(x_N) \leq \mathbf{0}_p, \\ & \boldsymbol{h}(\boldsymbol{x}) := \boldsymbol{h}_1(x_1) + \boldsymbol{h}_2(x_2) + \cdots + \boldsymbol{h}_N(x_N) = \mathbf{0}_q, \end{aligned} \tag{11}$$

where $\boldsymbol{x} = \operatorname{col}\{x_1, \ldots, x_N\} \in \mathbb{R}^n, n = \sum_{i=1}^{N} n_i$ and $\boldsymbol{\Omega} = \Omega_1 \times \cdots \times \Omega_N$. Problem (11) considers local constraints, coupled inequality constraints, and coupled equality

constraints together, which is more general than those in [5, 7, 8, 17, 31, 34–36, 39, 40]. If the coupled equality constraints are specified as $\sum_{i=1}^{N}(x_i - r_i) = 0$, then problem (11) reduces to the conventional resource allocation problem. If the coupled equality constraints are specified as $x_i = x_j, \forall i, j \in \mathbb{V}$, then problem (11) reduces to the optimal consensus problem. Moreover, the following assumptions are adopted.

Assumption 3.1.

1) For each $i \in \mathbb{V}$, Ω_i is compact and convex; f_i and $\boldsymbol{g}_i$ are convex, Lipschitz continuous, and possibly nonsmooth; $\boldsymbol{h}_i(x_i) = A_i x_i + b_i$ for $A_i \in \mathbb{R}^{q\times n_i}$ and $b_i \in \mathbb{R}^q$.

2) The problem (11) has at least one finite optimal solution.

3) The strong Slater's constraint qualification satisfies, that is, there exists $\hat{\boldsymbol{x}} \in \mathrm{rint}(\boldsymbol{\Omega})$ such that $\boldsymbol{g}(\hat{\boldsymbol{x}}) < \mathbf{0}_p$ and $\boldsymbol{h}(\hat{\boldsymbol{x}}) = \mathbf{0}_q$, where $\mathrm{rint}(\boldsymbol{\Omega})$ is the relative interior of the convex set $\boldsymbol{\Omega}$.

4) The graph $\mathbb{G}$ is undirected and connected.

Assumptions 1) and 2) are common in distributed constrained optimization problems [14], while we do not assume the smoothness of the objective function as in [5] and [15]. The Slater condition 3) is sufficient for a zero duality gap as well as for the existence of a dual optimal solution [11]. 4) is widely used in distributed optimization [37], and it plays an important role in our problem transformation and analysis.

4. MAIN RESULTS

In this section, we first transform the constrained distributed optimization problem into a variational inequality problem with a set-valued monotone mapping. Then we prove that weak solutions to the obtained variational inequality problem are also strong solutions. Finally, we give a distributed algorithm based on Nesterov's dual averaging scheme with guaranteed convergence to an optimal primal-dual solution.

4.1. Transformation to variational inequality problem

For notational simplicity, define

$$\boldsymbol{\theta}_i(x_i) = \begin{bmatrix} \boldsymbol{g}_i(x_i) \\ \boldsymbol{h}_i(x_i) \end{bmatrix}, \quad \lambda = \begin{bmatrix} \lambda^g \\ \lambda^h \end{bmatrix} \in \Theta := \mathbb{R}^p_{\geq 0} \times \mathbb{R}^q.$$

Then the coupled constraints can be rewritten as

$$\boldsymbol{\theta}(\boldsymbol{x}) := \boldsymbol{\theta}_1(x_1) + \boldsymbol{\theta}_2(x_2) + \cdots + \boldsymbol{\theta}_N(x_N) \in \widetilde{\Theta} := \mathbb{R}^p_{\leq 0} \times \{\mathbf{0}_q\}.$$

The corresponding Lagrange dual problem of (11) is

$$\max_{\lambda\in\Theta} \min_{\boldsymbol{x}\in\boldsymbol{\Omega}} \mathcal{L}(\boldsymbol{x}, \lambda) := f(\boldsymbol{x}) + \lambda^\top \boldsymbol{\theta}(\boldsymbol{x}). \tag{12}$$

Assign (x_i, λ_i, w_i) to each agent i, where (x_i, λ_i) is a primal-dual variable pair and $w_i \in \mathbb{R}^{p+q}$ is an auxiliary variable. Define $\boldsymbol{\lambda} = \mathrm{col}\{\lambda_1, \ldots, \lambda_N\}, \boldsymbol{w} = \mathrm{col}\{w_1, \ldots, w_N\}, \boldsymbol{\eta} = \mathrm{col}\{\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{w}\}$. The following theorem transforms the constrained optimization problem into a variational inequality problem.

Theorem 4.1. Under Assumption 3.1, $\boldsymbol{x}^* \in \boldsymbol{\Omega}$ is an optimal solution to problem (11) if and only if there exist $\boldsymbol{w}^*$ and $\boldsymbol{\lambda}^*$ such that $\boldsymbol{\eta}^* = \mathrm{col}\{\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{w}^*\}$ is a strong solution to the VI$(\boldsymbol{\Lambda}, \boldsymbol{F})$:

$$\text{find } \boldsymbol{\eta}^* \in \boldsymbol{\Lambda} \text{ and } \boldsymbol{y}^* \in \boldsymbol{F}(\boldsymbol{\eta}^*), \text{ s.t. } \langle \boldsymbol{y}^*, \boldsymbol{\eta} - \boldsymbol{\eta}^* \rangle \geq 0, \ \forall \boldsymbol{\eta} \in \boldsymbol{\Lambda}, \tag{13}$$

where

$$\begin{aligned} \boldsymbol{\Theta} &= \Theta \times \cdots \times \Theta \subset \mathbb{R}^{(p+q)N}, \quad \boldsymbol{\Lambda} = \boldsymbol{\Omega} \times \boldsymbol{\Theta} \times \mathbb{R}^{(p+q)N}, \\ \boldsymbol{F}(\boldsymbol{\eta}) &= \begin{bmatrix} \mathrm{col}\left(\partial f_i\left(x_i\right) + \partial \boldsymbol{\theta}_i\left(x_i\right) \lambda_i\right)_{i=1}^N \\ -\mathrm{col}\left(\boldsymbol{\theta}_i\left(x_i\right)\right)_{i=1}^N - L_N \otimes I_{p+q} \boldsymbol{w} \\ L_N \otimes I_{p+q} \boldsymbol{\lambda} \end{bmatrix}. \end{aligned} \tag{14}$$

P r o o f. Under the Slater condition in Assumption 3.1, the optimal dual set is nonempty. According to the KKT condition [28], $\boldsymbol{x}^* \in \boldsymbol{\Omega}$ is an optimal solution to problem (11) if and only if there exists a multiplier λ^* such that

$$\mathbf{0} \in \frac{\partial \mathcal{L}}{\partial \boldsymbol{x}}(\boldsymbol{x}^*, \lambda^*) + \mathcal{N}_{\boldsymbol{\Omega}}(\boldsymbol{x}^*) \tag{15a}$$

$$\mathbf{0} \in \frac{\partial \mathcal{L}}{\partial \lambda}(\boldsymbol{x}^*, \lambda^*) - \mathcal{N}_{\Theta}(\lambda^*) \tag{15b}$$

$$\lambda^* \in \Theta, \quad \boldsymbol{\theta}\left(\boldsymbol{x}^*\right) \in \widetilde{\Theta}, \tag{15c}$$

where $\frac{\partial}{\partial \boldsymbol{x}}$ and $\frac{\partial}{\partial \lambda}$ are the partial subdifferential. Thus, we only need to show that (15) holds if and only if

$$\mathbf{0} \in \begin{bmatrix} \partial f_1\left(x_1^*\right) \\ \vdots \\ \partial f_N\left(x_N^*\right) \end{bmatrix} + \begin{bmatrix} \partial \boldsymbol{\theta}_1\left(x_1^*\right) \lambda_1^* \\ \vdots \\ \partial \boldsymbol{\theta}_N\left(x_N^*\right) \lambda_N^* \end{bmatrix} + \begin{bmatrix} \mathcal{N}_{\boldsymbol{\Omega}_1}\left(x_1^*\right) \\ \vdots \\ \mathcal{N}_{\boldsymbol{\Omega}_N}\left(x_N^*\right) \end{bmatrix}, \tag{16a}$$

$$\mathbf{0} \in \begin{bmatrix} \boldsymbol{\theta}_1\left(x_1^*\right) \\ \vdots \\ \boldsymbol{\theta}_N\left(x_N^*\right) \end{bmatrix} + \begin{bmatrix} \sum_{j \in \mathcal{N}_1} a_{1j}\left(w_1^* - w_j^*\right) \\ \vdots \\ \sum_{j \in \mathcal{N}_N} a_{Nj}\left(w_N^* - w_j^*\right) \end{bmatrix} - \begin{bmatrix} \mathcal{N}_{\Theta}\left(\lambda_1^*\right) \\ \vdots \\ \mathcal{N}_{\Theta}\left(\lambda_N^*\right) \end{bmatrix}, \tag{16b}$$

$$\mathbf{0} = \begin{bmatrix} \sum_{j \in \mathcal{N}_1} a_{1j}\left(\lambda_1^* - \lambda_j^*\right) \\ \vdots \\ \sum_{j \in \mathcal{N}_N} a_{Nj}\left(\lambda_N^* - \lambda_j^*\right) \end{bmatrix}. \tag{16c}$$

Equations in (16) can be rewritten in a compact form as

$$\mathbf{0} \in \boldsymbol{F}(\boldsymbol{\eta}^*) + \mathcal{N}_{\boldsymbol{\Lambda}}(\boldsymbol{\eta}^*), \tag{17}$$

which is equivalent to the variational inequality (13).

Suppose (16) holds. Equation (16c) is equivalent to $L_N \otimes I_{p+q} \boldsymbol{\lambda} = \mathbf{0}$, and the solution is $\boldsymbol{\lambda}^* = \mathbf{1}_N \otimes \lambda^* \in \mathbb{R}^{(p+q)N}$, where $\lambda^* \in \mathbb{R}^{p+q}$. In other words, $\lambda_1 = \cdots = \lambda_N = \lambda^*$. Then (16a) becomes $0 \in \partial f_i\left(x_i^*\right) + \partial \boldsymbol{\theta}_i\left(x_i^*\right) \lambda^* + \mathcal{N}_{\boldsymbol{\Omega}_i}\left(x_i^*\right), \forall i = 1, \ldots, N$, and its compact form is (15a). Equation (16b) is $\boldsymbol{\theta}_i\left(x_i^*\right) + \sum_{j \in \mathcal{N}_i} a_{ij}\left(w_i^* - w_j^*\right) \in \mathcal{N}_{\Theta}\left(\lambda^*\right), \forall i = 1, \ldots, N$.

Since $\mathbb{G}$ is connected, $\sum_{i=1}^{N}\sum_{j\in\mathcal{N}_i} a_{ij}\left(w_i^* - w_j^*\right) = \mathbf{0}$. Consequently, $\sum_{i=1}^{N}\boldsymbol{\theta}_i\left(x_i^*\right) \in \mathcal{N}_\Theta(\lambda^*)$, which implies (15b). Thus, (15) holds.

Conversely, suppose (15) holds. Let $\boldsymbol{\lambda}^* = \mathbf{1}_N \otimes \lambda^*$. Then (16c) holds. Also, (15a) yields $0 \in \partial f_i\left(x_i^*\right) + \partial\boldsymbol{\theta}_i\left(x_i^*\right)\lambda^* + \mathcal{N}_{\boldsymbol{\Omega}_i}\left(x_i^*\right), \forall i = 1,\ldots,N$. Thus, (16a) holds. (15b) yields $0 \in \sum_{i=1}^{N}\boldsymbol{\theta}_i\left(x_i^*\right) - \mathcal{N}_\Theta(\lambda^*)$. Let $P = I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top$, and we have $\mathbf{1}_N^\top P = 0$. Consider the following linear equation

$$L_N \otimes I_{p+q}\boldsymbol{w} = -P \otimes I_{p+q}\begin{bmatrix}\boldsymbol{\theta}_1\left(x_1^*\right)\\ \vdots \\ \boldsymbol{\theta}_N\left(x_N^*\right)\end{bmatrix}. \tag{18}$$

Let $Q = \left(L_N \otimes I_{p+q}| - P \otimes I_{p+q}\begin{bmatrix}\boldsymbol{\theta}_1\left(x_1^*\right)\\ \vdots \\ \boldsymbol{\theta}_N\left(x_N^*\right)\end{bmatrix}\right)$ be the augmented matrix obtained by appending the columns of two matrices. Since $\mathbf{1}_N^\top Q = 0$, $\operatorname{rank}(Q) \le N-1$. On the other hand, $\operatorname{rank}(Q) \ge \operatorname{rank}(L_N) = N-1$. Thus $\operatorname{rank}(Q) = N-1 = \operatorname{rank}(L_N \otimes I_{p+q})$. According to Rouché–Capelli theorem [29], the equation (18) has infinitely many solutions, and we take arbitrary one as $\boldsymbol{w}^*$. Then

$$\begin{aligned}
&\begin{bmatrix}\boldsymbol{\theta}_1\left(x_1^*\right)\\ \vdots \\ \boldsymbol{\theta}_N\left(x_N^*\right)\end{bmatrix} + L_N \otimes I_{p+q}\boldsymbol{w}^* - \begin{bmatrix}\mathcal{N}_\Theta\left(\lambda_1^*\right)\\ \vdots \\ \mathcal{N}_\Theta\left(\lambda_N^*\right)\end{bmatrix}\\
=&\begin{bmatrix}\boldsymbol{\theta}_1\left(x_1^*\right)\\ \vdots \\ \boldsymbol{\theta}_N\left(x_N^*\right)\end{bmatrix} - P \otimes I_{p+q}\begin{bmatrix}\boldsymbol{\theta}_1\left(x_1^*\right)\\ \vdots \\ \boldsymbol{\theta}_N\left(x_N^*\right)\end{bmatrix} - \begin{bmatrix}\mathcal{N}_\Theta\left(\lambda^*\right)\\ \vdots \\ \mathcal{N}_\Theta\left(\lambda^*\right)\end{bmatrix}\\
=&\frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top \otimes I_{p+q}\begin{bmatrix}\boldsymbol{\theta}_1\left(x_1^*\right)\\ \vdots \\ \boldsymbol{\theta}_N\left(x_N^*\right)\end{bmatrix} - \begin{bmatrix}\mathcal{N}_\Theta\left(\lambda^*\right)\\ \vdots \\ \mathcal{N}_\Theta\left(\lambda^*\right)\end{bmatrix}\\
=&\frac{1}{N}\begin{bmatrix}\sum_{i=1}^{N}\boldsymbol{\theta}_i\left(x_i^*\right) - \mathcal{N}_\Theta\left(\lambda^*\right)\\ \vdots \\ \sum_{i=1}^{N}\boldsymbol{\theta}_i\left(x_i^*\right) - \mathcal{N}_\Theta\left(\lambda^*\right)\end{bmatrix} \ni \mathbf{0}_N.
\end{aligned}$$

(16b) holds, and thus, (16) is proved. □

Remark 4.2. The auxiliary variable $\boldsymbol{w} = \operatorname{col}\{w_1,\ldots,w_N\}$ is introduced to decouple the constraints (15b) with the help of graph Laplacian matrix L_N. In this way, distributed algorithms can be designed over multi-agent networks.

4.2. Strong and weak solutions

In this subsection, we further show that weak solutions to the variational inequality problem (13) are also strong ones so that it is sufficient to solve the original optimization

problem by computing only weak solutions to the transformed variational inequality problem.

Theorem 4.3. Under Assumption 3.1, the following statements hold.

1) The set-valued mapping $\boldsymbol{F}:\boldsymbol{\Lambda}\to 2^{\mathbb{R}^{(1+2p+2q)N}}$ defined in (14) is a maximal monotone operator.

2) Moreover, any strong solution $\boldsymbol{\eta}^*$ to the VI$(\boldsymbol{\Lambda},\boldsymbol{F})$ (13) is the same as its weak solution $\tilde{\boldsymbol{\eta}}^*$, that is,

$$\text{Find } \tilde{\boldsymbol{\eta}}^*\in\boldsymbol{\Lambda}:\ \langle \boldsymbol{y},\boldsymbol{\eta}-\tilde{\boldsymbol{\eta}}^*\rangle\geq 0,\ \forall\boldsymbol{\eta}\in\boldsymbol{\Lambda},\ \forall\boldsymbol{y}\in\boldsymbol{F}(\boldsymbol{\eta}). \tag{19}$$

P r o o f. 1) Since f_i and g_i are convex, their subdifferentials $\partial f_i(x_i)$ and $\partial g_i(x_i)$ are nonempty sets. As a result, $\boldsymbol{F}$ is generally a set-valued mapping.

Firstly, let us prove the monotonicity. For any $(\boldsymbol{\eta},\boldsymbol{y}),(\boldsymbol{\eta}',\boldsymbol{y}')\in\mathcal{G}(\boldsymbol{F})$, any $\boldsymbol{u}\in\partial f(\boldsymbol{x}),\boldsymbol{u}'\in\partial f(\boldsymbol{x}'),\boldsymbol{v}=\mathrm{col}\{\boldsymbol{v}^{\boldsymbol{g}},\boldsymbol{v}^{\boldsymbol{h}}\}\in\partial\boldsymbol{\theta}(\boldsymbol{x}),\boldsymbol{v}'\in\partial\boldsymbol{\theta}(\boldsymbol{x}')$, we have

$$\begin{aligned}\langle\boldsymbol{\eta}'-\boldsymbol{\eta},\boldsymbol{y}'-\boldsymbol{y}\rangle&=\langle\boldsymbol{x}'-\boldsymbol{x},\boldsymbol{u}'-\boldsymbol{u}\rangle+\left\langle\boldsymbol{w}'-\boldsymbol{w},L_N\otimes I_{p+q}(\boldsymbol{\lambda}'-\boldsymbol{\lambda})\right\rangle\\&\quad-\left\langle\boldsymbol{\lambda}'-\boldsymbol{\lambda},L_N\otimes I_{p+q}(\boldsymbol{w}'-\boldsymbol{w})\right\rangle+\sum_{i=1}^{N}\left\langle\lambda_i,D_{\boldsymbol{\theta}_i}^{\boldsymbol{v}_i}(x_i',x_i)\right\rangle+\left\langle\lambda_i',D_{\boldsymbol{\theta}_i}^{\boldsymbol{v}_i'}(x_i,x_i')\right\rangle,\end{aligned} \tag{20}$$

where the lth component of the column vector $D_{\boldsymbol{\theta}_i}^{\boldsymbol{v}_i}(\cdot,\cdot)$ is $D_{\boldsymbol{\theta}_i^{(l)}}^{\boldsymbol{v}_i^{(l)}}(\cdot,\cdot)$ defined in (2). To be specific,

$$\left\langle\lambda_i,D_{\boldsymbol{\theta}_i}^{\boldsymbol{v}_i}(x_i',x_i)\right\rangle=\left\langle\lambda_i^g,D_{\boldsymbol{g}_i}^{\boldsymbol{v}_i^{\boldsymbol{g}}}(x_i',x_i)\right\rangle+\left\langle\lambda_i^h,D_{\boldsymbol{h}_i}^{\boldsymbol{v}_i^{\boldsymbol{h}}}(x_i',x_i)\right\rangle\geq 0,$$

because of $\lambda_i^g\in\mathbb{R}^p_{\geq 0}$, $D_{\boldsymbol{g}_i}^{\boldsymbol{v}_i^{\boldsymbol{g}}}(x_i',x_i)\geq\boldsymbol{0}$ and $D_{\boldsymbol{h}_i}^{\boldsymbol{v}_i^{\boldsymbol{h}}}(x_i',x_i)=\boldsymbol{0}$. Also, $\left\langle\lambda_i',D_{\boldsymbol{\theta}_i}^{\boldsymbol{v}_i'}(x_i,x_i')\right\rangle\geq 0$. Since f_i is convex, $\langle\boldsymbol{x}'-\boldsymbol{x},\boldsymbol{u}'-\boldsymbol{u}\rangle\geq 0$. Since L_N is a symmetric matrix,

$$\left\langle\boldsymbol{w}'-\boldsymbol{w},L_N\otimes I_{p+q}(\boldsymbol{\lambda}'-\boldsymbol{\lambda})\right\rangle=\left\langle\boldsymbol{\lambda}'-\boldsymbol{\lambda},L_N\otimes I_{p+q}(\boldsymbol{w}'-\boldsymbol{w})\right\rangle.$$

Then $\langle\boldsymbol{\eta}'-\boldsymbol{\eta},\boldsymbol{y}'-\boldsymbol{y}\rangle\geq 0$. Thus, $\boldsymbol{F}$ is monotone.

Next, we prove the maximality. Take another monotone operator $\widetilde{\boldsymbol{F}}$ such that $\boldsymbol{F}\subset\widetilde{\boldsymbol{F}}$. It suffices to prove that given any $\boldsymbol{\eta}\in\boldsymbol{\Lambda}$ and any $\widetilde{\boldsymbol{y}}=\mathrm{col}\{\widetilde{\boldsymbol{y}}_1,\widetilde{\boldsymbol{y}}_2,\widetilde{\boldsymbol{y}}_3\}\in\widetilde{\boldsymbol{F}}(\boldsymbol{\eta})$, there holds $\widetilde{\boldsymbol{y}}\in\boldsymbol{F}(\boldsymbol{\eta})$, that is, $\widetilde{\boldsymbol{y}}_1\in\partial_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\lambda),\widetilde{\boldsymbol{y}}_2=-\mathrm{col}\left(\boldsymbol{\theta}_i(x_i)\right)_{i=1}^N-L_N\otimes I_{p+q}\boldsymbol{w}$, and $\widetilde{\boldsymbol{y}}_3=L_N\otimes I_{p+q}\boldsymbol{\lambda}$. Define $\boldsymbol{b}:=\widetilde{\boldsymbol{y}}_2+\mathrm{col}\left(\boldsymbol{\theta}_i(x_i)\right)_{i=1}^N+L_N\otimes I_{p+q}\boldsymbol{w},\boldsymbol{c}:=\widetilde{\boldsymbol{y}}_3-L_N\otimes I_{p+q}\boldsymbol{\lambda}$. We need to prove that

$$\widetilde{\boldsymbol{y}}_1\in\partial_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\lambda),\ \boldsymbol{b}=\boldsymbol{0},\ \boldsymbol{c}=\boldsymbol{0}.$$

Since $\widetilde{\boldsymbol{y}}$ belongs to $\widetilde{\boldsymbol{F}}(\boldsymbol{\eta})$ and $\boldsymbol{F}\subset\widetilde{\boldsymbol{F}}$, it follows from the monotonicity of $\widetilde{\boldsymbol{F}}$ that

$$\langle\boldsymbol{\eta}-\boldsymbol{\eta}',\widetilde{\boldsymbol{y}}-\boldsymbol{y}'\rangle\geq 0,$$

for all $\boldsymbol{\eta}' \in \boldsymbol{\Lambda}, \boldsymbol{y}' := \mathrm{col}\{\boldsymbol{y}_1', \boldsymbol{y}_2', \boldsymbol{y}_3'\} \in \boldsymbol{F}(\boldsymbol{\eta}')$. By the definitions of $\boldsymbol{b}$ and $\boldsymbol{c}$,

$$\langle \boldsymbol{x} - \boldsymbol{x}', \widetilde{\boldsymbol{y}}_1 - \boldsymbol{y}_1' \rangle + \langle \boldsymbol{w} - \boldsymbol{w}', \boldsymbol{c} \rangle + \left\langle \boldsymbol{\lambda} - \boldsymbol{\lambda}', \boldsymbol{b} - \mathrm{col}\left(\boldsymbol{\theta}_i\left(x_i\right)\right)_{i=1}^N + \mathrm{col}\left(\boldsymbol{\theta}_i\left(x_i'\right)\right)_{i=1}^N \right\rangle \geq 0, \tag{21}$$

for all $\boldsymbol{x}' \in \boldsymbol{\Omega}, \boldsymbol{\lambda}' \in \boldsymbol{\Theta}, \boldsymbol{w}' \in W, \boldsymbol{y}_1' \in \partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}', \lambda')$. Taking $\boldsymbol{\lambda}' = \boldsymbol{\lambda}$ and $\boldsymbol{w}' = \boldsymbol{w}$ in (21) yields

$$\langle \boldsymbol{x} - \boldsymbol{x}', \widetilde{\boldsymbol{y}}_1 - \boldsymbol{y}_1' \rangle \geq 0, \quad \forall\, \boldsymbol{y}_1' \in \partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}', \lambda).$$

Since $\partial_{\boldsymbol{x}} \mathcal{L}$ is maximal monotone by Theorem 4.7.1 in [25], $\widetilde{\boldsymbol{y}}_1 \in \partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda)$. Taking $\boldsymbol{x}' = \boldsymbol{x}, \boldsymbol{\lambda}' = \boldsymbol{\lambda} + t_1 \boldsymbol{b}, \boldsymbol{w}' = \boldsymbol{w} + t_2 \boldsymbol{c}$ in (21) with $t_1, t_2 > 0$ yields

$$-t_1 \|\boldsymbol{b}\|^2 - t_2 \|\boldsymbol{c}\|^2 \geq 0,$$

which indicates $\boldsymbol{b} = \boldsymbol{0}$ and $\boldsymbol{c} = \boldsymbol{0}$. Thus, $\boldsymbol{F}$ is maximal monotone.

2) Let $\boldsymbol{\eta}^*$ be a strong solution to the VI$(\boldsymbol{\Lambda}, \boldsymbol{F})$. It holds that

$$\langle \boldsymbol{y}^*, \boldsymbol{\eta} - \boldsymbol{\eta}^* \rangle \geq 0,\ \forall \boldsymbol{\eta} \in \boldsymbol{\Lambda},\ \exists\ \boldsymbol{y}^* \in \boldsymbol{F}(\boldsymbol{\eta}^*).$$

Since $\boldsymbol{F}$ is monotone,

$$\langle \boldsymbol{y}, \boldsymbol{\eta} - \boldsymbol{\eta}^* \rangle \geq \langle \boldsymbol{y}^*, \boldsymbol{\eta} - \boldsymbol{\eta}^* \rangle \geq 0,\ \forall \boldsymbol{\eta} \in \boldsymbol{\Lambda},\ \forall \boldsymbol{y} \in \boldsymbol{F}(\boldsymbol{\eta}).$$

Thus, $\boldsymbol{\eta}^*$ is a weak solution to the VI$(\boldsymbol{\Lambda}, \boldsymbol{F})$.

Conversely, let $\tilde{\boldsymbol{\eta}}^*$ be a weak solution to the VI$(\boldsymbol{\Lambda}, \boldsymbol{F})$. Recall that the indicator function $\mathcal{I}_{\boldsymbol{\Lambda}}(\boldsymbol{\eta})$ of the convex set $\boldsymbol{\Lambda}$ is a lower semi-continuous proper convex function. Thus, $\mathcal{N}_{\boldsymbol{\Lambda}}(\boldsymbol{\eta}) = \partial \mathcal{I}_{\boldsymbol{\Lambda}}(\boldsymbol{\eta})$ is a maximal monotone operator according to Lemma 2.1. By the definition of the normal cone, $\tilde{\boldsymbol{\eta}}^*$ is also a weak solution to the VI$(\boldsymbol{\Lambda}, \mathcal{N}_{\boldsymbol{\Lambda}})$. Let $\boldsymbol{H} = \boldsymbol{F} + \mathcal{N}_{\boldsymbol{\Lambda}}$, and then $\boldsymbol{H}$ is maximal monotone according to the Rockafellar Sum Theorem [27]. Further, $\tilde{\boldsymbol{\eta}}^*$ is also a weak solution to the VI$(\boldsymbol{\Lambda}, \boldsymbol{H})$, that is,

$$\langle \boldsymbol{y} - \boldsymbol{0}, \boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}^* \rangle \geq 0,\ \forall (\boldsymbol{\eta}, \boldsymbol{y}) \in \mathcal{G}(\boldsymbol{H}).$$

Since $\boldsymbol{H}$ is maximal monotone, $\boldsymbol{0} \in \boldsymbol{H}(\tilde{\boldsymbol{\eta}}^*)$. Consequently, for any $\tilde{\boldsymbol{y}}^* \in \boldsymbol{F}(\tilde{\boldsymbol{\eta}}^*)$, we have $-\tilde{\boldsymbol{y}}^* \in \mathcal{N}_{\boldsymbol{\Lambda}}(\tilde{\boldsymbol{\eta}}^*)$, which implies

$$\langle \tilde{\boldsymbol{y}}^*, \boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}^* \rangle \geq 0,\ \forall \boldsymbol{\eta} \in \boldsymbol{\Lambda}.$$

Therefore, $\tilde{\boldsymbol{\eta}}^*$ is a strong solution to the VI$(\boldsymbol{\Lambda}, \boldsymbol{F})$. □

Remark 4.4. When f_i and $\boldsymbol{g}_i$ are smooth, $\boldsymbol{F}$ in (14) becomes a single-valued mapping. This is the case of the original Nesterov's dual averaging method for variational inequality problem in [22]. We further deal with the distributed nonsmooth optimization problem that yields a variational inequality problem with the set-valued mapping $\boldsymbol{F}$. In such case, Theorem 4.3 plays a key role in extending the Nesterov's dual averaging method for general variational inequality problems.

Algorithm 1 Dual Averaging Algorithm for Distributed Nonsmooth Convex Optimization with Coupled Constraints (DDA)

Input: The bounded sets $\boldsymbol{\Omega}, \hat{\Theta}, W$; the number of iterations K; the communication graph $\mathbb{G}$; a strongly convex and nonnegative function ψ; a constant $\gamma > 0$, the stepsize base $\hat{\alpha}_0$=1.

Initialization: All variables $x_i[1], \lambda_i[1], w_i[1], s_i^1[1], s_i^2[1], s_i^3[1]$ equal to $\mathbf{0}$, $i = 1, \ldots, N$.

for $k = 0 : K-1$ **do**
  for $i = 1 : N$ **do**
    take arbitrary $u \in \partial f_i\left(x_i[k]\right), v \in \partial \boldsymbol{\theta}_i\left(x_i[k]\right)$;
    $s_i^1[k+1] = s_i^1[k] + u + v\lambda_i[k]$;
    $s_i^2[k+1] = s_i^2[k] - \boldsymbol{\theta}_i\left(x_i[k]\right) - \sum_{j\in\mathcal{N}_i} a_{ij}\left(w_i[k] - w_j[k]\right)$;
    $s_i^3[k+1] = s_i^3[k] + \sum_{j\in\mathcal{N}_i} a_{ij}\left(\lambda_i[k] - \lambda_j[k]\right)$;
    $\alpha_k = \gamma\hat{\alpha}_k$, $\hat{\alpha}_{k+1} = \left(\hat{\alpha}_k + \hat{\alpha}_k^{-1}\right)^{-1}$;
    $x_i[k+1] = \Pi_{\Omega_i}^{\psi}(s_i^1[k+1], \alpha_k)$;
    $\lambda_i[k+1] = \Pi_{\hat{\Theta}}^{\psi}(s_i^2[k+1], \alpha_k)$;
    $w_i[k+1] = \Pi_{W}^{\psi}(s_i^3[k+1], \alpha_k)$;
  end for
end for

Output: $\bar{x}_i[K] = \frac{1}{K}\sum_{k=1}^{K} x_i[k]$, $i = 1, \ldots, N$.

4.3. Algorithm design and convergence results

In this subsection, we present Algorithm 1 by applying Nesterov's dual averaging scheme.

In our algorithm, the restricted set of λ_i is

$$\hat{\Theta} = \left\{\lambda \in \Theta | \|\lambda\| \leq \frac{f(\hat{\boldsymbol{x}}) - \tilde{q}}{\zeta} + \delta\right\},$$

where $\hat{\boldsymbol{x}}$ is a *Slater vector* of (11); $\tilde{q} = \min_{\boldsymbol{x}\in\boldsymbol{\Omega}} \mathcal{L}(\boldsymbol{x}, \tilde{\lambda})$ is the dual function value for arbitrary $\tilde{\lambda} \in \Theta$; $\delta > 0$ is arbitrary; $\zeta = \min_{i=1,\ldots,p+q}\{-\boldsymbol{\theta}^{(i)}(\hat{\boldsymbol{x}})\}$, where (i) denotes the ith component. The restricted set of w_i is

$$W = \{\boldsymbol{w} \subset \mathbb{R}^{p+q} | \|\boldsymbol{w}\| \leq \frac{N+1}{\sqrt{N}} \frac{D_{\boldsymbol{\theta}}}{\rho_2}\},$$

where ρ_2 is the second smallest eigenvalue of the Laplacian matrix L_N and

$$D_{\boldsymbol{\theta}} = \max_i\{\max_{x_i\in\Omega_i}\{\|\boldsymbol{\theta}_i(x_i)\|\}\}.$$

The update rules in **Algorithm 1** can be rewritten in a compact form as

$$\boldsymbol{s}[k+1] = \boldsymbol{s}[k] + \boldsymbol{y}[k],\ \boldsymbol{y}[k] \in \boldsymbol{F}(\boldsymbol{\eta}[\boldsymbol{k}]), \tag{22a}$$

$$\boldsymbol{\eta}[k+1] = \Pi^{\psi}_{\hat{\boldsymbol{\Lambda}}}(\boldsymbol{s}[k+1], \alpha_k), \tag{22b}$$

$$\bar{\boldsymbol{\eta}}[K] = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\eta}[k], \tag{22c}$$

where

$$\hat{\boldsymbol{\Theta}} = \hat{\Theta} \times \cdots \times \hat{\Theta} \subset \boldsymbol{\Theta}, \quad \hat{\boldsymbol{\Lambda}} = \boldsymbol{\Omega} \times \hat{\boldsymbol{\Theta}} \times W \subset \boldsymbol{\Lambda}.$$

Theorem 4.5. Under Assumption 3.1, the following statements hold.

1) The solutions to the VI($\hat{\boldsymbol{\Lambda}}, \boldsymbol{F}$) are the same as the solutions to the VI($\boldsymbol{\Lambda}, \boldsymbol{F}$).

2) The limit of the running average sequence $\{\bar{\boldsymbol{\eta}}[K]\}$ generated by (22) is the solution to $\mathcal{V}(\boldsymbol{\eta}) = 0$ with the convergence rate

$$\mathcal{V}(\bar{\boldsymbol{\eta}}[K]) \leq \frac{1}{K\hat{\alpha}_K}\left(\frac{D}{\gamma} + \frac{\gamma\kappa^2}{2}\right), \tag{23}$$

where the merit function $\mathcal{V}$ was defined in (8), D and κ are constants satisfying

$$D \geq \max\{\psi(\boldsymbol{\eta}) | \boldsymbol{\eta} \in \hat{\boldsymbol{\Lambda}}\}, \quad \kappa \geq \max\{\|\boldsymbol{y}\| | \boldsymbol{\eta} \in \hat{\boldsymbol{\Lambda}}, \boldsymbol{y} \in \boldsymbol{F}(\boldsymbol{\eta})\}.$$

3) Let $\bar{\boldsymbol{\eta}}^* = \mathrm{col}\{\bar{\boldsymbol{x}}^*, \bar{\boldsymbol{\lambda}}^*, \bar{\boldsymbol{w}}^*\}$ be the limit of the sequence $\{\bar{\boldsymbol{\eta}}[K]\}$ generated by (22). Then $\bar{\boldsymbol{x}}^*$ is an optimal solution to problem (11).

P r o o f. 1) It has been shown in [19] that the optimal dual solution λ^* of (12) satisfies

$$\|\lambda^*\| \leq \frac{f(\hat{\boldsymbol{x}}) - \tilde{q}}{\zeta}, \tag{24}$$

and thus, λ^* lies in $\hat{\Theta}$. Consider the following saddle-point problem

$$\max_{\lambda \in \hat{\Theta}} \min_{\boldsymbol{x} \in \boldsymbol{\Omega}} \mathcal{L}(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda^T \boldsymbol{\theta}(\boldsymbol{x}). \tag{25}$$

Both (12) and (25) have the same optimal dual solution λ^* and attain the same optimal objective value. Then we consider the problem (25), instead of the original Lagrange dual problem (12).

In the proof of Theorem 4.1, we just mention the existence of $\boldsymbol{w}^*$. Actually, we can give the general solution to the equation (18) as

$$\boldsymbol{w}^* = (L_N \otimes I_{p+q})^{\dagger}\left(-P \otimes I_{p+q}\,\mathrm{col}\left(\boldsymbol{\theta}_i(x_i)\right)_{i=1}^{N}\right) + \mathbf{1}_N \otimes \boldsymbol{w}_0, \tag{26}$$

where $\boldsymbol{w}_0 \in \mathbb{R}^{p+q}$. The eigenvalues of the Laplacian matrix L_N are labeled so that $\rho_N \geq \cdots \geq \rho_2 \geq \rho_1 = 0$, and the corresponding eigenvectors are denoted by $u_s, s = 1, \ldots, N$. $U = (u_1, \ldots, u_N)$ is an orthogonal matrix. From [10], the generalized inverse of L_N can be expressed as

$$L_N^{\dagger} = U\mathrm{diag}(1/\rho_N, 1/\rho_{N-1}, \ldots, 1/\rho_2, 0)U^{\top}.$$

Since x_i lies in the compact set Ω_i and the function $\boldsymbol{\theta}_i$ is continuous,

$$D_{\boldsymbol{\theta}} = \max_i \{ \max_{x_i \in \Omega_i} \{ \|\boldsymbol{\theta}_i(x_i)\| \} \}$$

is finite and can be evaluated. Then

$$\|\boldsymbol{w}^*|_{\boldsymbol{w}_0=\mathbf{0}}\| \leq \| \left(L_N \otimes I_{p+q}\right)^\dagger \|_2 \| P \otimes I_{p+q} \operatorname{col}\left(\boldsymbol{\theta}_i\left(x_i\right)\right)_{i=1}^N \| \leq \frac{N+1}{\sqrt{N}} \frac{D_{\boldsymbol{\theta}}}{\rho_2},$$

where $\|\cdot\|_2$ denotes the l_2 norm of the matrix. Without loss of generality, we can restrict $\boldsymbol{w}$ to the bounded set $W = \{\boldsymbol{w} \subset \mathbb{R}^{p+q} | \|\boldsymbol{w}\| \leq \frac{N+1}{\sqrt{N}} \frac{D_{\boldsymbol{\theta}}}{\rho_2}\}$ such that there still exists $\boldsymbol{w} \in W$ satisfying (18).

2) Since $\boldsymbol{\Omega}_i, \hat{\boldsymbol{\Theta}}$ and W are compact and convex, $\hat{\boldsymbol{\Lambda}}$ is also compact and convex. Then the function $\mathcal{V}(\boldsymbol{\eta})$ is well-defined. It follows from Lemma 2.2 that $\mathcal{V}(\boldsymbol{\eta}) \geq 0$ for any $\boldsymbol{\eta} \in \hat{\boldsymbol{\Lambda}}$. Note that the average $\bar{\boldsymbol{\eta}}[K]$ lies in the set $\hat{\boldsymbol{\Lambda}}$, since $\hat{\boldsymbol{\Lambda}}$ is convex and $\boldsymbol{\eta}[k] \in \hat{\boldsymbol{\Lambda}}$ for all k. By calculations,

$$\begin{aligned}
\mathcal{V}(\bar{\boldsymbol{\eta}}[K]) &= \sup \left\{ \langle \boldsymbol{y}, \bar{\boldsymbol{\eta}}[K] - \boldsymbol{\eta} \rangle \, | \boldsymbol{\eta} \in \hat{\boldsymbol{\Lambda}}, \boldsymbol{y} \in \boldsymbol{F}(\boldsymbol{\eta}) \right\} \\
&= \sup \left\{ \left\langle \boldsymbol{y}, \frac{1}{K} \sum_{k=1}^K \boldsymbol{\eta}[k] - \boldsymbol{\eta} \right\rangle | \boldsymbol{\eta} \in \hat{\boldsymbol{\Lambda}}, \boldsymbol{y} \in \boldsymbol{F}(\boldsymbol{\eta}) \right\} \\
&= \frac{1}{K} \sup \left\{ \sum_{k=1}^K \langle \boldsymbol{y}, \boldsymbol{\eta}[k] - \boldsymbol{\eta} \rangle \, | \boldsymbol{\eta} \in \hat{\boldsymbol{\Lambda}}, \boldsymbol{y} \in \boldsymbol{F}(\boldsymbol{\eta}) \right\} \\
&\overset{(a)}{\leq} \frac{1}{K} \sup \left\{ \sum_{k=1}^K \langle \boldsymbol{y}[k], \boldsymbol{\eta}[k] - \boldsymbol{\eta} \rangle \, | \boldsymbol{\eta} \in \hat{\boldsymbol{\Lambda}}, \boldsymbol{y}[k] \in \boldsymbol{F}(\boldsymbol{\eta}[k]) \right\} \qquad (27) \\
&\overset{(b)}{\leq} \frac{1}{K} \left(\sum_{k=1}^K \frac{\alpha_{k-1} \|\boldsymbol{y}[k]\|^2}{2} + \max \left\{ \frac{\psi(\boldsymbol{\eta})}{\alpha_K} | \boldsymbol{\eta} \in \hat{\boldsymbol{\Lambda}} \right\} \right) \\
&\leq \frac{1}{K} \left(\sum_{k=1}^K \frac{\alpha_{k-1} \kappa^2}{2} + \frac{D}{\alpha_K} \right) \\
&\overset{(c)}{=} \frac{1}{K \hat{\alpha}_K} \left(\frac{\gamma \kappa^2}{2} + \frac{D}{\gamma} \right) \overset{K \to \infty}{\longrightarrow} 0,
\end{aligned}$$

where (a) is due to the monotonicity of $\boldsymbol{F}$, (b) follows from (10) in Lemma 2.3, and (c) holds because of the update rule of $\hat{\alpha}_k$. Since $\hat{\boldsymbol{\Lambda}}$ is a compact set, as well as f_i and g_i are Lipschitz continuous, D and κ are finite.

3) By the second part of this theorem, $\mathcal{V}(\bar{\boldsymbol{\eta}}^*) = 0$. According to Lemma 2.2, the solutions to $\mathcal{V}(\boldsymbol{\eta}) = 0$ are the same as the weak solutions to the VI($\hat{\boldsymbol{\Lambda}}, \boldsymbol{F}$), which are also the solutions to the VI($\boldsymbol{\Lambda}, \boldsymbol{F}$). By Theorem 4.3, the weak solutions to the variational inequality problem are the same as its strong solutions, and then $\bar{\boldsymbol{\eta}}^*$ is also a strong solution to the VI($\boldsymbol{\Lambda}, \boldsymbol{F}$) (13). Finally, as Theorem 4.1 states, the strong solutions to the variational inequality problem are corresponding to the optimal solutions to the optimization problem. Thus, $\bar{\boldsymbol{x}}^*$ is an optimal solution to the optimization problem (11). □

Remark 4.6.

1) Our Algorithm 1 is distributed since the ith agent only needs local data x_i, λ_i, $\partial f_i(x_i)$, $\partial \boldsymbol{\theta}_i(x_i)$, and neighbors' information $\lambda_j, w_j, j \in \mathcal{N}_i$.

2) By the upper bound of $\mathcal{V}(\bar{\boldsymbol{\eta}}[K])$ in (23), the optimal choice of γ is $\sqrt{2D}/\kappa$.

3) As in calculating the projection operator Π, we take $\psi(x) = \frac{1}{2}\|x\|^2$, and thus, $x_i[k+1] = \Pi^{\psi}_{\Omega_i}(s_i^1[k+1], \alpha_k)$ turns out to be $x_i[k+1] = P_{\Omega_i}(-\alpha_k s_i^1[k+1])$, where P_{Ω_i} denotes the Euclidean projection of a vector onto the set Ω_i. In the same way, $\lambda_i[k+1] = \Pi^{\psi}_{\hat{\Theta}}(s_i^2[k+1], \alpha_k)$ turns out to be $\lambda_i[k+1] = P_{\hat{\Theta}}(-\alpha_k s_i^2[k+1])$, and $w_i[k+1] = \Pi^{\psi}_{W}(s_i^3[k+1], \alpha_k)$ turns out to be $w_i[k+1] = P_W(-\alpha_k s_i^3[k+1])$.

4) The choice of $\hat{\alpha}_k$ is inspired by [22], which yields the bounds:

$$\sqrt{2k-1} \le \frac{1}{\hat{\alpha}_k} \le \frac{1}{1+\sqrt{3}} + \sqrt{2k-1}, \quad k \ge 1.$$

Since $\hat{\alpha}_K \propto O(1/\sqrt{K})$, $\mathcal{V}(\bar{\boldsymbol{\eta}}[K]) \propto O(1/\sqrt{K})$. Thus, Algorithm 1 achieves a convergence rate of $O(1/\sqrt{K})$. Note that this convergence rate is optimal for general nonsmooth convex optimization and cannot be further improved [20].

5. SIMULATION

This section gives a numerical example to show the convergence of our DDA algorithm and show the advantage of our algorithm over the state-of-the-art MDBD algorithm [6].

Consider the optimization problem (11) with

$$\begin{aligned} f_i(x_i) &= a_i x_i^2 + b_i|x_i - c_i|, \quad \Omega_i = [-0.6, 0.6], \\ g_i(x_i) &= d_i x_i^2 + e_i, \\ h_i(x_i) &= u_i x_i + v_i, \end{aligned} \tag{28}$$

over a 6-agent network. The objective functions are nonsmooth. The inequality constraints describe an elliptic region, and the equality constraints characterize an affine subspace. The coefficients of this problem, as shown in Table 1, are chosen so that there exists at least one finite optimal solution. Note that the vector $\boldsymbol{s} = [0, 0, 0, -0.5, 0, 0]$ satisfies the Slater condition in Assumption 3.1. An undirected circle graph for the network topology is given, whose adjacency matrix is assume to be

$$A = \begin{bmatrix} 0 & 0.5 & 0 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0.5 & 0 \end{bmatrix}.$$

We take $\gamma = 20$. With 100000 iterations in 0.1084 seconds, our algorithm obtains an approximate solution

$$\bar{\boldsymbol{x}}[100000] = [0.444760, 0.270539, 0.100033, -0.129843, -0.159883, -0.426349]$$

i	a_i	b_i	c_i	d_i	e_i	u_i	v_i
1	0.4	0.5	0.3	0.2	-0.3	0.6	0.5
2	0.5	0.3	0.2	0.3	-0.6	0.4	-0.3
3	0.7	0.2	0.1	0.8	-0.2	0.1	-0.2
4	0.6	0.7	0	0.5	-0.5	-0.6	-0.3
5	0.8	0.6	-0.1	0.8	-0.3	-0.6	-0.4
6	0.4	0.8	-0.2	0.4	-0.4	-0.8	-0.2

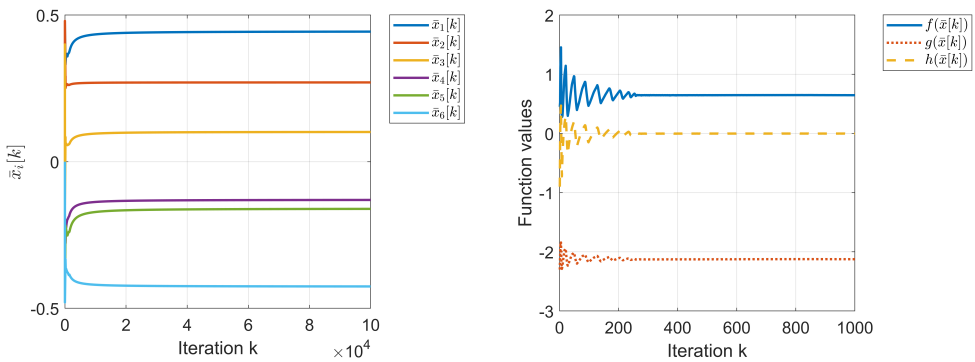
Tab. 1: Coefficients in the example.

with $f(\bar{\boldsymbol{x}}[100000]) = 0.627477$, $\boldsymbol{g}(\bar{\boldsymbol{x}}[100000]) = -2.12889$, and $\boldsymbol{h}(\bar{\boldsymbol{x}}[100000]) = -1.06167 \times 10^{-5}$. The trajectories of outputs and function values are shown in Figure 1a and 1b, which confirms that our DDA algorithm converges. Moreover, $\lim_{k\to\infty} \boldsymbol{g}(\bar{\boldsymbol{x}}[k]) < \boldsymbol{0}$ and $\lim_{k\to\infty} \boldsymbol{h}(\bar{\boldsymbol{x}}[k]) = \boldsymbol{0}$ indicate that our solution is feasible. MATLAB CVX is a traditional toolbox for solving convex optimization problems numerically, and we use it to get the exact solution $\boldsymbol{x}^*$, whose six-significant-figure form is

$$\boldsymbol{x}^*_{6SF} = [0.444767, 0.270543, 0.100000, -0.129845, -0.159884, -0.426357].$$

$f(\boldsymbol{x}^*_{6SF}) = 0.627455$, $\boldsymbol{g}(\boldsymbol{x}^*_{6SF}) = -2.12889$, and $\boldsymbol{h}(\boldsymbol{x}^*_{6SF}) = -4.1411 \times 10^{-10}$. CVX spent 2.2420 seconds on solving the problem and is time-consuming. $\|\bar{\boldsymbol{x}}[100000] - \boldsymbol{x}^*\| = 3.516 \times 10^{-5}$ and the trajectory of the optimality gap $\|\bar{\boldsymbol{x}}[k] - \boldsymbol{x}^*\|$ is shown in Figure 2, which indicates that $\bar{\boldsymbol{x}}[k] \to \boldsymbol{x}^*$, as $k \to \infty$.

In addition, we compare our DDA algorithm with the Distributed Mirror Descent algorithm with Bregman Damping (MDBD) [6] which is the state-of-the-art algorithm to solve distributed nonsmooth constrained optimization problems. As shown in Figure 2, the optimality gap of the DDA algorithm is less than the MDBD algorithm, which indicates that our DDA algorithm is better.



(a) Trajectories of $\bar{x}_i$ (b) Trajectories of function values

Fig. 1: Trajectories of outputs and function values of the DDA algorithm

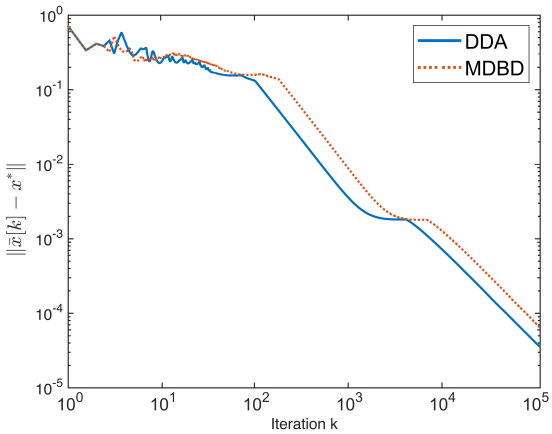


Fig. 2: The comparison between our DDA algorithm and the MDBD algorithm.

6. CONCLUSION

This paper investigated a distributed optimization problem with nonsmooth convex functions and coupled constraints. Firstly, we transformed the problem into a variational inequality problem with a set-valued mapping. Then we proposed a distributed algorithm based on Nesterov's dual averaging method and obtained the sublinear convergence rate of $O(1/\sqrt{k})$. With the help of a large number of developed methods in variational inequality problems, we will study more general distributed optimization problems. Future study includes extending the method to directed communication graphs.

ACKNOWLEDGEMENT

This work is supported in part by Shanghai Municipal Science and Technology Major Project under grant 2021SHZDZX0100 and in part by the National Natural Science Foundation of China under grant 61903027.

(Received June 24, 2022)

REFERENCES

[1] A. Auslender and R. Correa: Primal and dual stability results for variational inequalities. Comput. Optim. Appl. *17* (2000), 117–130. DOI:10.1023/A:1026594114013

[2] A. Auslender and M. Teboulle: Projected subgradient methods with non-Euclidean distances for non-differentiable convex minimization and variational inequalities. Math. Program. *120* (2009), 27–48. DOI:10.1007/s10107-007-0147-z

[3] D. P. Bertsekas and J. N. Tsitsiklis: Parallel and Distributed Computation: Numerical Methods. Prentice hall Englewood Cliffs, NJ 1989.

[4] J. M. Borwein and Q. J. Zhu: Techniques of Variational Analysis. Springer Science and Business Media, New York 2004.

[5] T. H. Chang, A. Nedić, and A. Scaglione: Distributed constrained optimization by consensus-based primal-dual perturbation method. IEEE Trans. Automat. Control *59* (2014), 1524–1538. DOI:10.1109/TAC.2014.2308612

[6] G. Chen, G. Xu, W. Li, and Y. Hong: Distributed mirror descent algorithm with Bregman damping for nonsmooth constrained optimization. IEEE Trans. Automat. Control (2023), 1–8. DOI:10.1109/tac.2023.3244995

[7] A. Cherukuri and J. Cortés: Distributed generator coordination for initialization and anytime optimization in economic dispatch. IEEE Trans. Control Network Syst. *2* (2015), 226–237. DOI:10.1109/TCNS.2015.2399191

[8] J. C. Duchi, A. Agarwal, and M. J. Wainwright: Dual averaging for distributed optimization: Convergence analysis and network scaling. IEEE Trans. Automat. Control *57* (2011), 592–606. DOI:10.1109/TAC.2011.2161027

[9] F. Facchinei and J. S. Pang: Finite-dimensional Variational Inequalities and Complementarity Problems. Springer Science and Business Media, 2007.

[10] I. Gutman and W. Xiao: Generalized inverse of the Laplacian matrix and some applications. Bulletin (Académie serbe des sciences et des arts. Classe des sciences mathématiques et naturelles. Sciences mathématiques) (2004), 15–23.

[11] J. B. Hiriart-Urruty and C. Lemaréchal: Convex Analysis and Minimization Algorithms I: Fundamentals. Springer Science and Business Media, 2013.

[12] B. Johansson, T. Keviczky, M. Johansson, and K. H. Johansson: Subgradient methods and consensus algorithms for solving convex optimization problems. In: 47th IEEE Conference on Decision and Control, IEEE 2008, pp. 4185–4190. DOI:10.1109/cdc.2008.4739339

[13] J. Koshal, A. Nedić, and U. V. Shanbhag: Multiuser optimization: Distributed algorithms and error analysis. SIAM J. Optim. *21* (2011), 1046–1081. DOI:10.1137/090770102

[14] S. Liang, X. Zeng, and Y. Hong: Distributed nonsmooth optimization with coupled inequality constraints via modified Lagrangian function. IEEE Trans. Automat. Control *63* (2017), 1753–1759. DOI:10.1109/TAC.2017.2752001

[15] S. Liang, L. Wang, and G. Yin: Distributed smooth convex optimization with coupled constraints. IEEE Trans. Automat. Control *65* (2019), 347–353. DOI:10.1109/TAC.2019.2912494

[16] S. Liang, L. Wang, and G. Yin: Distributed dual subgradient algorithms with iterate-averaging feedback for convex optimization with coupled constraint. IEEE Trans. Cybernetics *51* (2019), 2529-2539. DOI:10.1109/TCYB.2019.2933003

[17] Q. Liu and J. Wang: A second-order multi-agent network for bound-constrained distributed optimization. IEEE Trans. Automat. Control *60* (2015), 3310–3315. DOI:10.1109/TAC.2015.2416927

[18] Y. Lou, Y. Hong, and S. Wang: Distributed continuous-time approximate projection protocols for shortest distance optimization problems. Automatica *69* (2016), 289–297. DOI:10.1016/j.automatica.2016.02.019

[19] A. Nedić and A. Ozdaglar: Approximate primal solutions and rate analysis for dual subgradient methods. SIAM J. Optim. *19* (2009), 1757–1780. DOI:10.1137/070708111

[20] Y. Nesterov: Introductory Lectures on Convex Optimization: A Basic Course. Springer Science and Business Media, 2003.

[21] Y. Nesterov: Dual extrapolation and its applications to solving variational inequalities and related problems. Math. Program. *109.2-3* (2007), 319-344. DOI:10.1007/s10107-006-0034-z

[22] Y. Nesterov: Primal-dual subgradient methods for convex problems. Math. Program. *120* (2009), 221–259. DOI:10.1007/s10107-007-0149-x

[23] Y. Nesterov and V. Shikhman: Dual subgradient method with averaging for optimal resource allocation. Europ. J. Oper. Res. *270* (2018), 907–916. DOI:10.1016/j.ejor.2017.09.043

[24] M. Rabbat and R. Nowak: Distributed optimization in sensor networks. In: Proc. 3rd International Symposium on Information Processing in Sensor Networks, 2004, pp. 20–27.

[25] S. B. Regina and A. Iusem: Set-Valued Mappings and Enlargements of Monotone Operators. Springer, New York 2003.

[26] R. T. Rockafellar: Characterization of the subdifferentials of convex functions. Pacific J. Math. *17* (1966), 497–510. DOI:10.2140/pjm.1966.17.497

[27] R. T. Rockafellar: On the maximality of sums of nonlinear monotone operators. Trans. Amer. Math. Soc. *149* (1970), 75–88. DOI:10.1090/S0002-9947-1970-0282272-5

[28] A. Ruszczynski: Nonlinear Optimization. Princeton University Press, 2011.

[29] I. R. Shafarevich and A. O. Remizov: Linear Algebra and Geometry. Springer Science and Business Media, 2012.

[30] Z. Tu and W. Li: Multi-agent solver for non-negative matrix factorization based on optimization. Kybernetika *57* (2021), 60–77. DOI:10.14736/kyb-2021-1-0060

[31] L. Xiao and S. Boyd: Optimal scaling of a gradient method for distributed resource allocation. J. Optim. Theory Appl. *129* (2006), 469–488. DOI:10.1007/s10957-006-9080-1

[32] L. Xiao, S. Boyd, and S. J. Kim: Distributed average consensus with least-mean-square deviation. J. Parallel Distributed Comput. *67* (2007), 33–46. DOI:10.1016/j.jpdc.2006.08.010

[33] J. C. Yao: Variational inequalities with generalized monotone operators. Math. Oper. Res. *19* (1994), 691–705. DOI:10.1287/moor.19.3.691

[34] 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), 45–52. DOI:10.1016/j.sysconle.2015.06.006

[35] P. Yi, Y. Hong, and F. Liu: Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems. Automatica *74* (2016), 259–269. DOI:10.1016/j.automatica.2016.08.007

[36] P. Yi and L. Pavel: A distributed primal-dual algorithm for computation of generalized Nash equilibria via operator splitting methods. In: 2017 IEEE 56th Annual Conference on Decision and Control, IEEE 2017, pp. 3841–3846. DOI:10.1109/cdc.2017.8264224

[37] X. Zeng, S. Liang, Y. Hong, and J. Chen: Distributed computation of linear matrix equations: An optimization perspective. IEEE Trans. Automat. Control *64* (2018), 1858–1873. DOI:10.1109/TAC.2018.2847603

[38] X. Zeng, P. Yi, and Y. Hong: Distributed continuous-time algorithm for constrained convex optimizations via nonsmooth analysis approach. IEEE Trans. Automat. Control *62* (2016), 5227–5233. DOI:10.1109/TAC.2016.2628807

[39] Y. Zhang and M. Zavlanos: A consensus-based distributed augmented Lagrangian method. In: 2018 Conference on Decision and Control, IEEE 2018, pp. 1763-1768. DOI:10.1109/cdc.2018.8619512

[40] M. Zhu and S. Martínez: On distributed convex optimization under inequality and equality constraints. IEEE Trans. Automat. Control *57* (2011), 151–164. DOI:10.1109/TAC.2011.2167817

Zhipeng Tu, Key Laboratory of Systems and Control, Institute of Systems Science, Chinese Academy of Sciences, Beijing, 100190. P. R. China.
e-mail: tuzhipeng@amss.ac.cn

Shu Liang, Department of Control and Engineering, Tongji University, Shanghai 201804. P. R. China.
e-mail: sliang@tongji.edu.cn