Electronic Research Archive

http://www.aimspress.com/journal/era

ERA, 32(2): 857–873.
DOI: 10.3934/era.2024041
Received: 09 October 2023
Revised: 18 December 2023
Accepted: 27 December 2023
Published: 12 January 2024

Research article

Linear convergence of a primal-dual algorithm for distributed interval optimization

Yinghui Wang[1,2], **Jiuwei Wang**[3,4,*], **Xiaobo Song**[1,2] **and Yanpeng Hu**[1,2]

[1] Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

[2] Beijing Engineering Research Center of Industrial Spectrum Imaging, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

[3] University of Chinese Academy of Sciences, Beijing 100049, China

[4] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

* **Correspondence:** Email: waliuss@sina.com.

Abstract: In this paper, we investigate a distributed interval optimization problem whose local functions are interval functions rather than scalar functions. Focusing on distributed interval optimization, this paper presents a distributed primal-dual algorithm. A criterion is introduced under which linear convergence to the Pareto solution of distributed interval optimization problems can be achieved without strong convexity. Lastly, a numerical simulation is presented to illustrate the linear convergence of the algorithm that has been proposed.

Keywords: distributed interval optimization problem; Pareto solution; linear convergence rates; primal-dual algorithm

1. Introduction

Due to the theoretical significance and wide range of applications in areas such as machine learning, multi-agent system coordination, sensor networks, and smart grids, distributed optimization has received a lot of attention from researchers in recent years. Various distributed algorithms for solving distributed optimization problems have been introduced, and they involve agents collaborating with their neighboring agents in order to attain global minimization, see recent works [1–7].

The aforementioned works' objective functions are scalar functions. In practice, however, scalar functions have been frequently incapable of expressing objective functions for distributed networks explicitly or precisely (see [8–10]). On the contrary, interval functions are employed to describe problems, as exemplified in the applications of smart grids and economic systems [11, 12]. To address the challenges presented by interval functions, interval optimization problems (IOPs), have been proposed [13–19]. Initial studies on IOPs were conducted by the authors of [13], and subsequently investigated in [14, 15]. Existence conditions have been presented in [11,20] to achieve Pareto solutions of IOPs. In addition, [21–24] detail algorithms that have been designed for centralized IOPs. Without conducting a theoretical analysis, [11,12] present distributed applications of IOPs in economic systems and smart infrastructures. For centralized IOPs [21–24] have presented algorithms. These line search algorithms, nevertheless, fail in distributed environments.

Given this context, it is natural for us to consider the design of efficient algorithms to solve DIOPs over multi-agent networks. The DIOPs, nevertheless, remain a subject of ongoing research. This may be due to the ease with which line search algorithms (e.g., Wolfe or Lamke's algorithms [21–24]) can be applied in distributed settings, and very few papers [25] with related theoretical results have been published. In addition, algorithm designs are made difficult by the partial order of interval functions.

Furthermore, there is growing interest in the convergence rates of distributed algorithms for distributed optimization with scalar functions. In fact, when local objective functions were strongly convex, the algorithms of [2, 26, 27] achieved linear convergence rates for the centralized and distributed counterparts. Local scalar functions for distributed optimization are not strongly convex in a number of practical applications. Further investigation was undertaken by a group of scholars [1,28–30] regarding the substitution of strongly convex conditions that dictate linear convergence rates. For example, [1] analyzed four distinct categories of function conditions and deduced the linear convergence of numerous centralized algorithms. The authors of [28, 29] respectively demonstrated the linear rates of their distributed algorithms under metrically sub-regular and Polyak-Lojasiewicz conditions.

In this paper, we investigate the Pareto solutions of a DIOP whose local functions are interval functions rather than scalar functions. The DIOP is given as follows:

$$(DIOP) \quad \min_{s} \quad G(s), \quad G(s) = \sum_{i=1}^{n} G_i(s)$$

where $G_i = [L_i, R_i]$ is a convex interval function for each agent i. $L_i(s) \leqslant R_i(s)$ holds for every given s. Still, each agent can only get the gradient information of interval function G_i. By means of neighborhood information communication, the global Pareto solution is obtained. The contributions of this paper are summarized as follows:

(a) We investigate the Pareto solution of a DIOP whose local functions are interval functions. By incorporating convexity and well-defined partial orderings of interval functions, we convert the DIOP [11, 20, 31] into a solvable distributed optimization problem scalarization (DSIOP) with convex global constraints.

(b) In this reformulation, the optimal solutions of the DSIOP correspond to the Pareto solutions of the DIOP. With this relationship, we propose a distributed primal-dual algorithm to find a Pareto solution of the DIOP.

(c) We discuss a crucial criterion that, when applied to Pareto solutions of a DIOP, weaken the strict or strong convexity required for linear convergence. Given that this paper investigates DIOPs, the supplied criterion differ from those delineated in [1, 28, 29]. In addition, the criterion is essential for evaluating the convergence of DIOP distributed algorithms.

The rest of the paper is organized as follows. The preliminaries of this paper are given in Section 2. In Section 3, the DIOP is analyzed. The primal-dual algorithm is further given to find a Pareto solution of the DIOP in Section 4 and a numerical example is given in Section 5. Finally, the conclusion of this paper is offered in Section 6.

Notations. Denote by $\mathcal{R}$ the set of real numbers, $I_n \in \mathcal{R}^{n\times n}$ as the identity matrix, and $\mathbf{1}_n = [1, 1, \ldots, 1]^\top \in \mathcal{R}^n$, respectively. Denote $\langle\cdot, \cdot\rangle$ as the inner product and $\|\cdot\|$ as the Euclidean norm in $\mathcal{R}^n$.

2. Preliminaries

In this section, we present an introduction to convex analysis for scalar functions [32], graph theory, and interval optimization [33].

2.1. Graph theory

Define $\mathcal{N} = \{1, 2, ..., n\}$ as the agent set and $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$ as the set of edges between agents. The communication between n agents is described by an undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$. If $(i, j) \in \mathcal{E}$, then the agent i can communicate with the agent j. Therefore, each agent $i \in \mathcal{N}$ can communicate with agents in its neighborhood $N_i = \{j|(i, j) \in \mathcal{E}\} \cup \{i\}$.

Denote $\mathcal{A} \in \mathcal{R}^{n\times n}$ as the communication matrix between agents, whose elements a_{ij} satisfy the following conditions:

$$a_{ij} = \begin{cases} a_{ii}, & \text{if } i = j \\ a_{ij}, & \text{if } i \neq j \text{ and } (i, j) \in \mathcal{E} \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

Denote d_i by the degree of agent i, i.e., $|d_i| = \sum_{j=1}^n a_{ij}$. Further, denote D by the $n\times n$ diagonal degree matrix such that $D = \mathrm{diag}(\sum_{j=1}^n a_{1j}, \ldots, \sum_{j=1}^n a_{nj})$. Then, the associated Laplacian matrix $\mathcal{P} \in \mathcal{R}^{n\times n}$ is $\mathcal{P} := D - \mathcal{A}$.

The following assumption forms the basis of the communication topology $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ between agents over the network:

Assumption 1. *The undirected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ is connected.*

Assumption 1 is extensively employed in [28], this ensures the consensus of vectors for agents over the network.

2.2. Convex analysis

Prior to proceeding with the discussion of interval functions, we define convexity and the Lipschitz continuity of scalar functions.

Definition 1. *(a) A scalar function $f : \Omega \to \mathcal{R}$ is convex if for any s_1, $s_2 \in \Omega$ and $z \in [0, 1]$, $f(\lambda s_2 + (1 - \lambda)s_1) \leq \lambda f(s_2) + (1 - \lambda)f(s_1)$ holds.*

(b) A scalar function $f : \mathcal{R}^n \to \mathcal{R}$ is κ-Lipschitz continuous with respect to a constant $\kappa > 0$ if

$$\|f(s_2) - f(s_1)\| \leqslant \kappa\|s_2 - s_1\|, \; \forall\, s_1, s_2 \in \mathcal{R}^n.$$

The following lemma is crucial for the analysis of convergence in distributed optimization problems involving scalar functions and interval functions.

Lemma 1. *[32, Lemma 11, Chapter 2.2] Define* $\{v^k\}_{k\geqslant 1}$ *and* $\{w^k\}_{k\geqslant 1}$ *as two nonnegative scalar sequences. Define* $\{h^k\}_{k\geqslant 1}$ *as a scalar sequence, which is bounded from below uniformly. If there exists a nonnegative constant sequence* $\eta^k \geqslant 0$ *with* $\sum_{k=1}^{\infty} \eta^k < \infty$ *and*

$$h^{k+1} \leqslant (1+\eta^k)h^k - v^k + w^k, \quad \forall k \geqslant 1$$

then $\{h^k\}_{k\geqslant 1}$ *converges with* $\sum_{k=1}^{\infty} v^k < \infty$.

2.3. Interval optimization problems

Let $G : \mathcal{R}^p \rightrightarrows \mathcal{R}$ be any interval map. Now, we consider the following IOP:

$$(IOP) \qquad \min_{x} \quad G(s) \quad s.\,t. \quad s \in \Omega$$

where $G(x) = [L(s), R(s)]$ is any non-empty compact interval in $\mathcal{R}$.

The Pareto optimal solution to an IOP is defined as follows:

Definition 2. *[34] A point* $s^* \in \Omega$ *is said to be a Pareto optimal solution to an IOP iff it holds that for some* $\bar{s} \in \Omega$, $L(\bar{s}) \leqslant L(s^*)$ *and* $R(\bar{s}) \leqslant R(s^*)$ *both hold implying that* $L(s^*) \leqslant L(s)$ *and* $R(s^*) \leqslant R(s)$.

The example of the DIOP is presented below. There is no solution other than the Pareto solution in the example that follows.

Example 1. *The IOP illustrated in Figure 1 does not have a solution. However, the Pareto optimal solutions to the given problem are* $[s_1, s_2]$.

(a) *For* $y \leqslant s_1$, *we have that* $R(y) \geqslant R(s_1)$ *and* $L(y) \geqslant L(s_1)$, *and* s_1 *is a Pareto solution to the IOP.*
(b) *For* $y \geqslant s_2$, *we have that* $R(y) \geqslant R(s_2)$ *and* $L(y) \geqslant L(s_2)$, *and* s_2 *is a Pareto solution to the IOP.*
(c) *For* $s_1 \leqslant y \leqslant s_2$, *we have that* $R(y) \leqslant R(s_1)$, $L(y) \geqslant L(s_1)$, $R(y) \geqslant R(s_2)$ *and* $L(y) \leqslant L(s_2)$. *For* $s_1 \leqslant y \leqslant s_2$, $\bar{s} \in \Omega$, $L(\bar{s}) \leqslant L(y)$ *and* $R(\bar{s}) \leqslant R(y)$ *could not hold concurrently.*

According to Definition 2, $[s_1, s_2]$ *are Pareto optimal solutions to this given problem.*

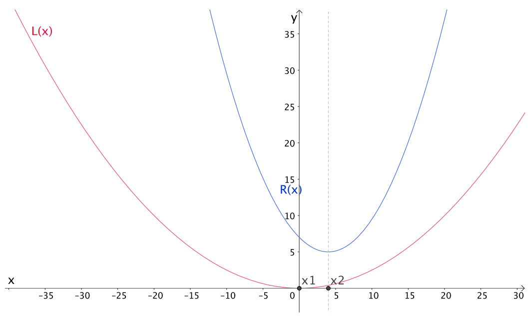

Figure 1. $L(x)$ and $R(x)$ for vector x.

To investigate the Pareto solutions of an IOP, let us consider the following IOP in conjunction with its scalarization (SIOP):

$$
\begin{aligned}
(SIOP) \quad & \min_{x} \quad \lambda L(x) + (1-\lambda) R(x) \\
& s.\,t. \quad x \in \Omega
\end{aligned}
$$

where $\lambda \in [0, 1]$. The following lemma holds for Pareto solutions of IOPs and solutions of SIOPs according to [34]. Furthermore, it remains valid in distributed settings.

Lemma 2. *[34] We assume that G is compact-valued and convex with respect to x:*

(a) If there exists a real number $\lambda \in (0, 1)$ such that $s^* \in \Omega$ is a solution to the SIOP, then $s^* \in \Omega$ is a Pareto optimization of the IOP.

(b) If a point $s^* \in \Omega$ is a Pareto optimization of the IOP, then there exists a real number $\lambda \in [0, 1]$ such that $s^* \in \Omega$ is an optimal solution of the SIOP.

3. Optimization model and algorithm

In this section, we consider a DIOP and introduce its distributed primal-dual algorithm.

3.1. Optimization model

Consider the following DIOP:

$$
\begin{aligned}
\text{(DIOP)} \qquad & \min_{s} \quad G(\boldsymbol{s}), \quad G(\boldsymbol{s}) = \sum_{i=1}^{n} G_i(s_i) \\
& \text{s. t.} \quad s_i = s_j
\end{aligned} \tag{3.1}
$$

where $\boldsymbol{s} = [s_1^\top, s_2^\top, \ldots, s_n^\top]^\top \in \mathcal{R}^{np}$, $s_i \in \mathcal{R}^p$, and $G_i = [L_i, R_i]$. $L_i, R_i : \mathcal{R}^p \to \mathcal{R}$ are convex functions. For any given s_i, $L_i(s_i) \leqslant R_i(s_i)$ holds. Each agent i knows its local interval function G_i.

Define $L(\boldsymbol{s})$ and $R(\boldsymbol{s})$ as

$$
L(\boldsymbol{s}) = \sum_{i=1}^{n} L_i(s_i), \quad R(\boldsymbol{s}) = \sum_{i=1}^{n} R_i(s_i). \tag{3.2}
$$

With (3.2), the definition of Pareto solutions is then given to the DIOP.

Definition 3. *[34] $s^* \in \Omega$ is a Pareto solution of the DIOP, iff for some $\bar{\boldsymbol{s}} \in \Omega$, $L(\bar{\boldsymbol{s}}) \leqslant L(\boldsymbol{s}^*)$ and $R(\bar{\boldsymbol{s}}) \leqslant R(\boldsymbol{s}^*)$ both hold implying that $L(\boldsymbol{s}^*) \leqslant L(\bar{\boldsymbol{s}})$ and $R(\boldsymbol{s}^*) \leqslant R(\bar{\boldsymbol{s}})$.*

The existence of Pareto solutions for the DIOP is guaranteed by Assumption 2 which is consistent with the centralized counterpart [35].

Assumption 2. *(a) $L_i(s)$ and $R_i(s)$ are strongly convex, continuous functions.*

(b) Problem (3.1) *has at least one Pareto solution.*

(c) Gradients of $L_i(s)$ and $R_i(s)$ are Lipschitz continuous.

Lemma 2 also establishes a theoretical framework for Pareto solutions for the DIOP. Consider the following scalarization of the DIOP as well. Define $f : \mathcal{R}^{np} \times \mathcal{R}^n \to \mathcal{R}$ and $f_i : \mathcal{R}^p \times [0, 1] \to \mathcal{R}$ as

$$
F(\boldsymbol{s}, \boldsymbol{z}) \triangleq \sum_{i=1}^{n} f_i(s_i, z_i) \tag{3.3}
$$

$$f_i(s_i, z_i) \triangleq z_i L_i(s) + (1 - z_i) R_i(s) \tag{3.4}$$

where $\boldsymbol{z} = [z_1, z_2, \ldots, z_n]^\top \in (0,1)^n$ and $\boldsymbol{s} = [s_1^\top, s_2^\top, \ldots, s_n^\top]^\top \in \mathcal{R}^{np}$. Let $\boldsymbol{z} = z_0 \mathbf{1}_n$ with $z_0 \in (0,1)$. The DSIOP (3.1) can be rewritten as follows:

$$\begin{aligned} \text{(DSIOP)} \qquad & \min_{s} \quad F(\boldsymbol{s}, \boldsymbol{z}), \quad F(\boldsymbol{s}, \boldsymbol{z}) = \sum_{i=1}^{n} f_i(s_i, z_i) \\ & \text{s. t.} \quad s_i = s_j, \quad z_i = z_j \end{aligned} \tag{3.5}$$

where each agent i possesses the following information: ∇f_i, s_i, $z_i \in (0,1)$ and $s_j \in N_i$. The given problem (3.5) can be modeled as a distributed optimization problem [28, 36, 37] with scalars when z represents a common vector to each agent i. Additionally, under Assumption 2, the following lemma remains valid:

Lemma 3. *[34, 35]*

(a) $f_i(s, z)$ is linear with respect to z and $f_i(s, z)$ is convex with respect to s .
(b) There are Lipschitz constants k_{i1} and K_1 such that the partial derivative $\nabla f_{i_x}(s, z)$ is Lipschitz continuous with respect to s with k_{i1} and $\nabla F_{\boldsymbol{s}}(\boldsymbol{s}, \boldsymbol{z})$ is Lipschitz continuous with respect to $\boldsymbol{s}$ with K_1.
(c) There are Lipschitz constants k_{i2} and K_2 such that $f_i(s, z)$ is Lipschitz continuous with respect to z with constant k_{i2} and $F(\boldsymbol{s}, \boldsymbol{z})$ is Lipschitz continuous with respect to $\boldsymbol{z}$ with constant K_2.
(d) There are Lipschitz constants k_{i3} and K_3 such that the partial derivative $\nabla f_{i_x}(s, z)$ is Lipschitz continuous with respect to z with constant k_{i3} and $\nabla F_{\boldsymbol{s}}(\boldsymbol{s}, \boldsymbol{z})$ is Lipschitz continuous with respect to $\boldsymbol{z}$ with constant K_3.

It should be noted that although $f_i(s_i, z_i)$ is convex with respect to s and z, $f_i(s_i, z_i)$ is not a convex function. Owing to the non-convexity of $f_i(s_i, z_i)$, the criteria for linear convergence rates of algorithms are no longer applicable to the DIOP.

3.2. Algorithm

During distributed optimization processes, $s_1, \ldots, s_n$, $z_1, \ldots, z_n$ are not necessarily equal all of the time. Therefore, it is natural to treat those variables separately and impose the soft constraints $s_1 = \ldots = s_n$, $z_1 = \ldots = z_n$. By using the Laplacian matrix $\mathcal{P}$, these constraints are equivalent to $\mathbf{P}\boldsymbol{s} = 0$ and $\mathcal{P}\boldsymbol{z} = 0$, where $\boldsymbol{z} = [z_1, z_2, \ldots, z_n]^\top \in (0,1)^n$, $\boldsymbol{s} = [s_1^\top, s_2^\top, \ldots, s_n^\top]^\top \in \mathcal{R}^{np}$, and $\mathbf{P} = \mathcal{P} \otimes I_p$. Consequently, problem (3.5) is reformulated as follows:

$$\begin{aligned} & \min_{s} \quad F(\boldsymbol{s}, \boldsymbol{z}), \quad F(\boldsymbol{s}, \boldsymbol{z}) = \sum_{i=1}^{n} f_i(s_i, z_i) \\ & \text{s. t.} \quad \mathbf{P}\boldsymbol{s} = 0, \quad \mathcal{P}\boldsymbol{z} = 0, \quad \boldsymbol{z} \in (0,1)^n. \end{aligned} \tag{3.6}$$

Let $\boldsymbol{t} = [t_1, t_2, \ldots, t_n]^\top$. Recall that the dual problem of (3.6) is

$$\begin{aligned} & \min_{\boldsymbol{s} \in \mathcal{R}^{nm}} \left[F(\boldsymbol{s}, \boldsymbol{z}) + \max_{\boldsymbol{t} \in \mathcal{R}^{np}} \langle \boldsymbol{t}, \boldsymbol{P}\boldsymbol{s} \rangle \right] \\ & \text{s. t.} \quad \mathcal{P}\boldsymbol{z} = 0, \quad \boldsymbol{z} \in (0,1)^n. \end{aligned} \tag{3.7}$$

and the augmented Lagrangian function of (3.7) with respect to $\boldsymbol{s}$ is

$$\tilde{\mathcal{L}}(\boldsymbol{s}, z, \boldsymbol{t}) = F(\boldsymbol{s}, z) + \langle \boldsymbol{t}, \boldsymbol{Ps}\rangle + \frac{1}{2}\langle \boldsymbol{s}, \boldsymbol{Ps}\rangle. \tag{3.8}$$

Define by $\bar{z}^0 = \frac{1}{n}\sum_{i=1}^{n} z_i^0$, $\bar{\boldsymbol{z}}^{\boldsymbol{0}} = [(\bar{z}^0)^\top, (\bar{z}^0)^\top, \ldots, (\bar{z}^0)^\top] \in \mathcal{R}^{np}$, where $z_i^0 \in (0,1)$ is an initial value for any agent i. For the vector $\bar{\boldsymbol{z}}^{\boldsymbol{0}}$, denote S^* as the optimal solution set of problem (3.6) and T^* as the saddle point set of problem (3.7), respectively. According to Assumption 2, for a proper given $\boldsymbol{z}^0$, there exists $\boldsymbol{t}^*$ such that $(\boldsymbol{s}^*, \boldsymbol{t}^*) \in S^* \times T^*$. $(\boldsymbol{s}^*, \boldsymbol{t}^*) \in S^* \times T^*$ also satisfies the following lemma, which is also a basis for the analysis of convergence:

Lemma 4. *(Karush-Kuhn-Tucker condition, [38, Theorems 3.25–3.27]) With Assumption 2, for a particular given* $\bar{\boldsymbol{z}}^{\boldsymbol{0}} = \bar{z}^0 \otimes \mathbf{1}_n \in (0,1)^n$*,* $(\boldsymbol{s}^*, \boldsymbol{t}^*)$ *is a solution to* (3.7) *if*

$$\begin{cases} 0 = -\nabla_{\boldsymbol{s}}\tilde{\mathcal{L}}(\boldsymbol{s}^*, \boldsymbol{t}^*) = -\nabla F_{\boldsymbol{s}^*}(\boldsymbol{s}^*, \bar{\boldsymbol{z}}^{\boldsymbol{0}}) - \boldsymbol{Ps}^* - \boldsymbol{Pt}^*, \\ 0 = \quad \nabla_{\boldsymbol{t}}\tilde{\mathcal{L}}(\boldsymbol{s}^*, \boldsymbol{t}^*) = \boldsymbol{Ps}^*. \end{cases}$$

With Lemma 4, we introduce a distributed primal-dual algorithm as follows:

$$s_i^{k+1} = s_i^k - h\Big(\nabla f_{i_{s_i^k}}(s_i^k, z_i^k) + \sum_{j=1}^{m} a_{ij}(s_i^k - s_j^k) + \sum_{j=1}^{m} a_{ij}(t_i^k - t_j^k)\Big) \tag{3.9a}$$

$$z_i^{k+1} = \sum_{j=1}^{m} a_{ij} z_j^k \tag{3.9b}$$

$$t_i^{k+1} = t_i^k + h\Big(\sum_{j=1}^{m} a_{ij}(s_i^k - s_j^k)\Big) \tag{3.9c}$$

where the step-size h satisfies that $0 < h < \frac{2}{L+4\sigma}$, σ is the largest eigenvalue of the Laplacian matrix $\mathcal{P}$. At the k-th iteration, for all $i \in \mathcal{V} = \{1, 2, \ldots, n\}$, each agent i only obtains a partial gradient in the form of $\nabla f_{i_{s_i^k}}(s_i^k, z_i^k)$ for its local function $f_i(s_i^k, z_i^k)$, and it is cooperative with neighbors to achieve a Pareto solution of problem (3.1).

The constraint $\mathcal{P}\lim_{k\to\infty} \boldsymbol{z(k)} = 0$, $z_i \in (0,1)$ in (3.6), is satisfied through (3.9b) and the initialization of $z_i(0) \in (0,1)$ in (3.9), while the constraint $\mathcal{P}\lim_{k\to\infty} \boldsymbol{x(k)} = 0$ and the minimization of $F(\boldsymbol{x}, z)$ are satisfied through (3.9a) and (3.9c) in (3.9). Define $\boldsymbol{s^k} = col\{s_1^k, \ldots, s_n^k\}$, $\boldsymbol{t^k} = col\{t_1^k, \ldots, t_n^k\}$ and $\boldsymbol{z^k} = col\{z_1^k, \ldots, z_n^k\}$. Then, with $\boldsymbol{w} \triangleq col\{\boldsymbol{s}, \boldsymbol{t}\} \in \mathcal{R}^{2qn}$, $\boldsymbol{w^*} \triangleq col\{\boldsymbol{s^*}, \boldsymbol{t^*}\} \in W^* \subset S^* \times T^*$ for a proper given $\bar{\boldsymbol{z}}^{\boldsymbol{0}}$, where W^* is the primal-dual solution set of problem (3.7). Algorithm (3.9) can be rewritten in a compact form in terms of $\{\boldsymbol{w}, \boldsymbol{z}\}$:

$$\begin{cases} \boldsymbol{w(k+1)} &= \boldsymbol{w(k)} - h\boldsymbol{I}(\boldsymbol{w(k)}, \boldsymbol{z(k)}) \\ \boldsymbol{z(k+1)} &= \mathcal{A}\boldsymbol{z(k)} \end{cases} \tag{3.10}$$

where

$$\boldsymbol{I}(\boldsymbol{w}, \boldsymbol{z}) \triangleq \begin{bmatrix} \boldsymbol{I_1}(\boldsymbol{w}, \boldsymbol{z}) \\ \boldsymbol{I_2}(\boldsymbol{w}, \boldsymbol{z}) \end{bmatrix} = \begin{bmatrix} \nabla F_{\boldsymbol{s}}(\boldsymbol{s}, \boldsymbol{z}) + \boldsymbol{Ps} + \boldsymbol{Pt} \\ -\boldsymbol{Ps} \end{bmatrix}. \tag{3.11}$$

We have the following basic result, whose proof is in the Appendix.

Theorem 1. *Under Assumptions 1 and 2, $\{s^k, t^k\}$ converges to the Pareto solution set W^*.*

Consider a Lyapunov function

$$V(w, z) = V_a(w, z) + V_b(w, z) + V_c(w, z) \tag{3.12}$$

where $V_a(w, z) = \sigma d^2(w, W^*)$, $V_b(w, z) = F(s, z) - F(s^*, \bar{z}^0) + \frac{1}{2}\langle s, Ps\rangle + \langle s, Pt\rangle$, $V_c(w, z) = K_2\|z - \bar{z}^0\|$ and K_2 is a Lipschitz constant given in Lemma 3. Theorem 1 is based on Lemmas 5 and 6, whose proof are also given in Appendix.

Lemma 5. *With Assumption 1, $\{z^k\}$ converges to $\bar{z}^0$ with a linear convergence rate γ_1 whose elements belong to $(0, 1)$: $\lim_{k\to\infty} z^k = \bar{z}^0$, $\left\|z^k - \bar{z}^0\right\| \leqslant \gamma_1\left\|z^{k-1} - \bar{z}^0\right\|$. $\left\|z^k - \bar{z}^0\right\|$ is also summable with respect to k: $\sum_{k=1}^{\infty}\left\|z^k - \bar{z}^0\right\| < \infty$.*

Lemma 6 is additionally presented to illustrate the minimum and maximum values of $V(w, z)$.

Lemma 6. *With Assumptions 1 and 2, the following inequality holds for the Lyapunov function $V(w, z)$:*

$$\frac{\sigma}{2}\left[\|s - s^*\|^2 + \|t - t^*\|^2\right] \leqslant V(w, z) \leqslant \frac{K_1 + 4\sigma}{2}\left[\|s - s^*\|^2 + \|t - t^*\|^2\right] + 2K_2\left\|z - \bar{z}^0\right\|$$

where K_1, K_2 are Lipschitz constants given in Lemma 1, and σ is the largest eigenvalue of the Laplacian matrix $\mathcal{P}$.

The asymptotic convergence of (3.9) is demonstrated by Theorem 1, which is consistent with that of [28] for distributed optimization. It should be noted that the inclusion of the partial gradient term $\nabla F_s(s, z)$ renders inapplicable the contraction mapping principle. In contrast to numerous distributed algorithms that rely on the contraction mapping principle for their proofs [26, 28, 37, 39], this awork involves employing the martingale convergence theorem (Lemma 1) in Theorem 1.

4. Main results

In this section, we present our main results. A criterion without strong convexity is first introduced for the DIOP, which, together with (3.9) will imply linear convergence. Our criterion for (3.9) to achieve exponential convergence is as follows.

Criterion. The continuously differentiable function $\tilde{\mathcal{L}} > 0$ has a *restricted quadratic gradient growth*. That is, if there exists a constant κ_L such that for any w, $w^* = P_{W^*}(w)$, we have

$$\left\langle \mathcal{I}(w, \bar{z}^0) - \mathcal{I}(w^*, \bar{z}^0), w - w^*\right\rangle \geqslant \kappa_L\|w - w^*\|^2 \tag{4.1}$$

where $\tilde{\mathcal{L}}$ is the augmented Lagrangian function defined in (3.8).

The criterion given in this paper differs from the quadratic convex condition given in [1] and the metrically irregular condition discussed in [28] for distributed optimization problems with scalar functions. This criterion is given for DIOPs. On the other hand, regarding the dynamics given by (3.9), we will show that (4.1) is sufficient to achieve linear convergence.

Theorem 2. *Under Assumptions 1 and 2 and* (4.1), $\{s^k, \boldsymbol{t^k}\}$ *converges linearly to the optimal set* W^*.

Proof. If $\boldsymbol{w} = \boldsymbol{w}^*$, we have that $\left\|\boldsymbol{I}(\boldsymbol{w}, z)\right\| \geqslant 0$. Further, consider the case when $\boldsymbol{w} \neq \boldsymbol{w}^*$. With Lemma 5, we obtain

$$\begin{aligned}\left\langle \boldsymbol{I}(\boldsymbol{w}, z), \boldsymbol{w} - \boldsymbol{w}^* \right\rangle =& \left\langle \boldsymbol{I}(\boldsymbol{w}, z) - \boldsymbol{I}(\boldsymbol{w}^*, z), \boldsymbol{w} - \boldsymbol{w}^* \right\rangle + \left\langle \boldsymbol{I}(\boldsymbol{w}^*, z) - \boldsymbol{I}(\boldsymbol{w}^*, \bar{z}^{\boldsymbol{0}}), \boldsymbol{w} - \boldsymbol{w}^* \right\rangle \\ \geqslant& \kappa_L \left\|\boldsymbol{w} - \boldsymbol{w}^*\right\|^2 - K_3 \left\|z - z^{\boldsymbol{0}}\right\| \cdot \left\|\boldsymbol{w} - \boldsymbol{w}^*\right\| \end{aligned} \tag{4.2}$$

where the last inequality holds by $\langle \boldsymbol{a}, \boldsymbol{b} \rangle \leqslant \frac{\|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2}{2}$. Still,

$$\left\langle \boldsymbol{I}(\boldsymbol{w}, z), \boldsymbol{w} - \boldsymbol{w}^* \right\rangle \leqslant \left\|\boldsymbol{I}(\boldsymbol{w}, z)\right\| \cdot \left\|\boldsymbol{w} - \boldsymbol{w}^*\right\|. \tag{4.3}$$

Equations (4.2) and (4.3) indicate that $\left\|\boldsymbol{I}(\boldsymbol{w}, z)\right\| \geqslant \kappa_L \left\|\boldsymbol{w} - \boldsymbol{w}^*\right\| - K_3 \left\|z - \bar{z}^{\boldsymbol{0}}\right\|$. Therefore, if Assumption 2 holds, $\left\|\boldsymbol{I}(\boldsymbol{w}, z)\right\| \geqslant \kappa_L \left\|\boldsymbol{w} - \boldsymbol{w}^*\right\| - K_3 \left\|z - z^{\boldsymbol{0}}\right\|$. By Lemma 6,

$$\begin{aligned}\left\|\boldsymbol{I}(\boldsymbol{w}^k, z^k)\right\|^2 \geqslant& \kappa_L^2 \left\|\boldsymbol{w^k} - \boldsymbol{w}^*\right\|^2 + K_3^2 \left\|z^k - z^{\boldsymbol{0}}\right\|^2 - 2\kappa_L K_3 \left\|\boldsymbol{w^k} - \boldsymbol{w}^*\right\| \cdot \left\|z^{\boldsymbol{k}} - z^{\boldsymbol{0}}\right\| \\ \geqslant& \frac{2\kappa_L^2}{K_1 + 4\sigma}\Big[V(\boldsymbol{w}^k, z^k) - 2K_2 \|z^{\boldsymbol{k}} - \bar{z}^{\boldsymbol{0}}\|\Big] - 2\kappa_L K_3 \left\|\boldsymbol{w^k} - \boldsymbol{w}^*\right\| \cdot \left\|z^{\boldsymbol{k}} - z^{\boldsymbol{0}}\right\| + K_3^2 \left\|z^{\boldsymbol{k}} - z^{\boldsymbol{0}}\right\|^2. \end{aligned} \tag{4.4}$$

Substituting (4.4) into (A12) yields

$$V(\boldsymbol{w^{k+1}}, z^{\boldsymbol{k+1}}) \leqslant T_1^k + T_2^k + T_3^k + T_4^k - T_5^k$$

where $T_1^k = \left(1 - \frac{h(2-\nu_0 h)\kappa_L^2}{K_1 + 4\sigma}\right) V(\boldsymbol{w}^k, z^k)$, $T_2^k = 2hK_3\sigma \left\|z^{\boldsymbol{k}} - \bar{z}^{\boldsymbol{0}}\right\| \cdot \left\|\boldsymbol{s^k} - \boldsymbol{s}^*\right.$, $T_3^k = K_2\left(\left\|z^{\boldsymbol{k+1}} - z^{\boldsymbol{k}}\right\| + (1-\gamma_1)\left\|z^{\boldsymbol{k}} - \bar{z}^{\boldsymbol{0}}\right\|\right)$, $T_4^k = \frac{2hK_2(2-\nu_0 h)\kappa_L^2}{K_1 + 4\sigma}\left\|z^{\boldsymbol{k}} - \bar{z}^{\boldsymbol{0}}\right\|$ and $T_5^k = -\dfrac{h(2 - \nu_o h)K_3^2}{2}\left\|z^{\boldsymbol{k}} - \bar{z}^{\boldsymbol{0}}\right\|^2$.

Still, according to Lemma 5, $\|z^{\boldsymbol{k}} - \bar{z}^{\boldsymbol{0}}\Big\|$ converges linearly at a rate γ_1. Therefore, residue terms T_2^k, T_3^k, T_4^k and T_5^k diminish with linear rates. Since $\nu_0 \leqslant K_1 + 4\sigma$, the main term T_1^k converges with a linear rate, which is no less than $\left(1 - \frac{h\left(2-(K_1+4\sigma)h\right)\kappa_L^2}{2(K_1+4\sigma)}\right)$. With Lemma 6, we obtain that $\left[\|\boldsymbol{s^{k+1}} - \boldsymbol{s}^*\|^2 + \|\boldsymbol{y^{k+1}} - \boldsymbol{t}^*\|^2\right] \leqslant \dfrac{2}{\sigma} V(\boldsymbol{w^{k+1}}, z^{\boldsymbol{k+1}})$, which completes the proof.

As shown in Theorem 1, (4.1) plays an important role in achieving linear convergence even in the absence of strong convexity of $f_i(s_i, z_i)$. In this paper, we extend the quadratic convex condition given in [1] to (4.1) for interval functions. Criterion (4.1) also describes another linear growth condition of gradients for distributed optimization problems.

5. Simulation

In this section, we demonstrate the following simulation:

$$\min \quad G(\boldsymbol{s}) = \sum_{i=1}^{9} [\upsilon_{1i}, \upsilon_{2i}] \|s - \rho_i\|^2$$

where υ_{1i}, $\upsilon_{1i} \in \mathcal{R}$ and $\rho_i \in \mathcal{R}^p$. The problem is motivated from both a centralized IOP [35] and the distributed optimization [40]. The communication topology between agents is described by Figure 2.

Define $[\upsilon_{1i}, \upsilon_{2i}] = [0.5, 2]$. Take $\rho_1 = 5$, $\rho_2 = 4$, $\rho_3 = 3$, $\rho_4 = 2$, $\rho_5 = 1$, $\rho_6 = 0$, $\rho_7 = -1$, $\rho_8 = -2$, and $\rho_9 = -3$. Next, initialize (3.9) by setting the step-size $h = 0.1$, z_i^0 as random numbers in [0, 1], and $s_i^0 = 0$. Then we investigate the convergence of (3.9). Also, Figures 3 and 4 show the consensus of z_i^k and convergence of s_i^k for the proposed algorithm. We get a Pareto solution as (0.4695; 1.002) for 1000 iterations. Figure 5 shows the convergence of s_i^k for a centralized primal-dual algorithm (an algorithm generated according to the properties of solutions in [35]) for each agent i, where z_i denotes random numbers in [0, 1]. In addition, we take a performance index R as $R^k = \log \|\boldsymbol{s}^k - \boldsymbol{s}^*\|^2$. The performance of R is shown in Figure 6, which implies the linear convergence of (3.9).

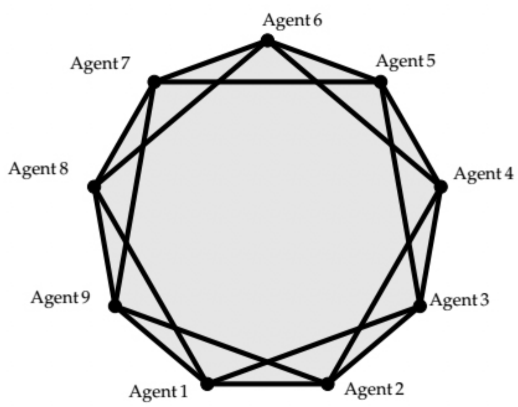

Figure 2. Communication topology between agents.

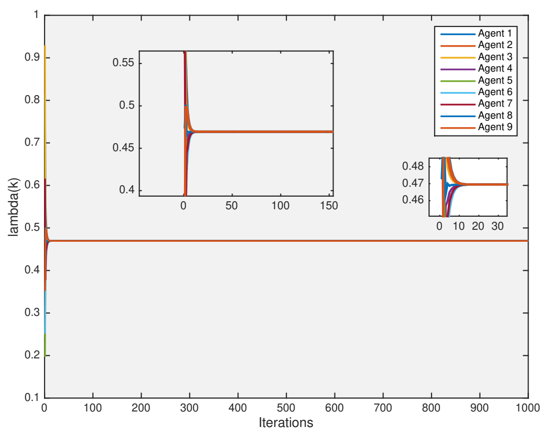

Figure 3. z_i^k for agent i of (3.9).

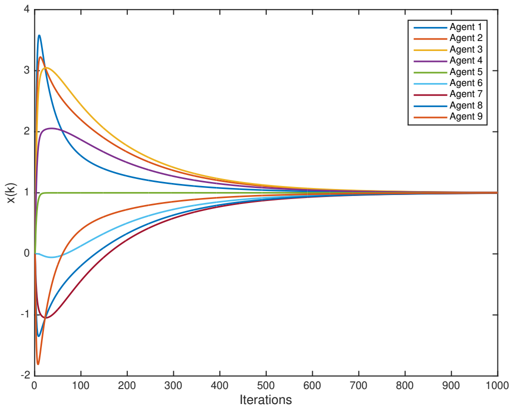

Figure 4. s_i^k for agent i of (3.9).

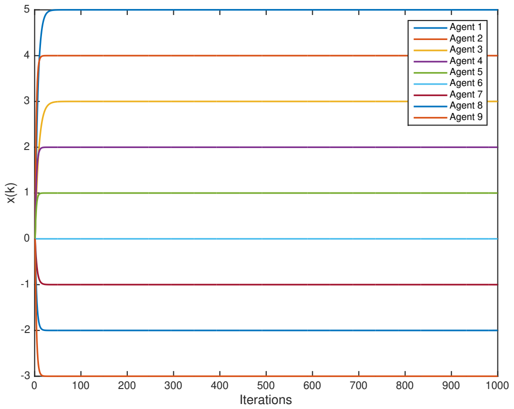

Figure 5. s_i^k for agent i of centralized primal-dual algorithm.

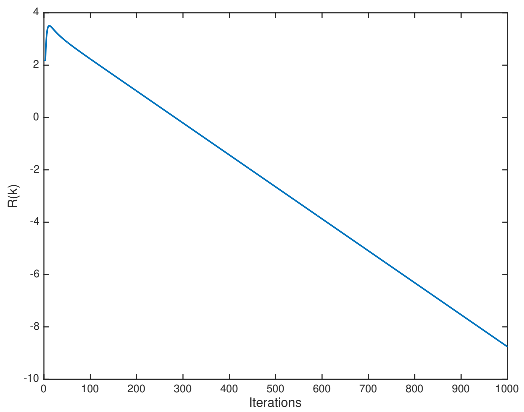

Figure 6. Convergence Rate of (3.9).

6. Conclusions

We have investigated a DIOP in which the local functions are interval functions in this paper. With distributed interval optimization as its primary focus, this article introduces a distributed primal-dual algorithm. A criterion has been proposed that allows the linear convergence to the Pareto solution of a DIOP without strong convexity. Finally, a numerical simulation has been executed to demonstrate the linear convergence of the proposed algorithm. Given that the existing research on DIOPs primarily focuses on objective interval functions, the investigation of distributed problems involving interval constraints should be expanded in the future.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was supported by the NSFC (72101026, 62203045) and Operation Expenses for Universities' Basic Scientific Research of Central Authorities (FRF-TP-22-141A1).

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. I. Necoara, Y. Nesterov, F. Glineur, Linear convergence of first order methods for non-strongly convex optimization, *Math. Program.*, **175** (2019), 69–107. https://doi.org/10.1007/s10107-018-1232-1
2. A. Makhdoumi, A. Ozdaglar, Convergence rate of distributed admm over networks, *IEEE Trans. Autom. Control*, **62** (2017), 5082–5095. https://doi.org/10.1109/TAC.2017.2677879
3. X. He, T. Huang, J. Yu, C. Li, Y. Zhang, A continuous-time algorithm for distributed optimization based on multiagent networks, *IEEE Trans. Syst. Man Cybern.: Syst.*, **49** (2017), 2700–2709. https://doi.org/10.1109/TSMC.2017.2780194
4. H. Li, Q. Lü, X. Liao, T. Huang, Accelerated convergence algorithm for distributed constrained optimization under time-varying general directed graphs, *IEEE Trans. Syst. Man Cybern.: Syst.*, **50** (2018), 2612–2622. https://doi.org/10.1109/TSMC.2018.2823901
5. Q. Wang, J. Chen, B. Xin, X. Zeng, Distributed optimal consensus for euler-lagrange systems based on event-triggered control, *IEEE Trans. Syst. Man Cybern.: Syst.*, **51** (2021), 4588–4598. https://doi.org/10.1109/TSMC.2019.2944857
6. J. Guo, R. Jia, R. Su, Y. Zhao, Identification of fir systems with binary-valued observations against data tampering attacks, *IEEE Trans. Syst. Man Cybern.: Syst.*, **53** (2023), 5861–5873. https://doi.org/10.1109/TSMC.2023.3276352

7. J. Guo, X. Wang, W. Xue, Y. Zhao, System identification with binary-valued observations under data tampering attacks, *IEEE Trans. Autom. Control*, **66** (2020), 3825–3832. https://doi.org/10.1109/TAC.2020.3029325
8. X. Zeng, Y. Peng, Y. Hong, Distributed algorithm for robust resource allocation with polyhedral uncertain allocation parameters, *J. Syst. Sci. Complexity*, **31** (2018), 103–119. https://doi.org/10.1007/s11424-018-7145-5
9. V. Kekatos, G. B. Giannakis, Distributed robust power system state estimation, *IEEE Trans. Power Syst.*, **28** (2013), 1617–1626. https://doi.org/10.1109/TPWRS.2012.2219629
10. S. Sra, S. Nowozin, S. J. Wright, *Optimization for Machine Learning*, Mit Press, 2012.
11. B. Q. Hu, S. Wang, A novel approach in uncertain programming part I: new arithmetic and order relation for interval numbers, *J. Ind. Manage. Optim.*, **2** (2006), 351–371. https://doi.org/10.3934/jimo.2006.2.351
12. L. Wu, M. Shahidehpour, Z. Li, Comparison of scenario-based and interval optimization approaches to stochastic SCUC, *IEEE Trans. Power Syst.*, **27** (2012), 913–921. https://doi.org/10.1109/TPWRS.2011.2164947
13. A. Neumaier, *Interval Methods for Systems of Equations*, Cambridge University Press, 1990.
14. J. Rohn, Positive definiteness and stability of interval matrices, *SIAM J. Matrix Anal. Appl.*, **15** (1994), 175–184. https://doi.org/10.1137/S0895479891219216
15. V. I. Levin, Nonlinear optimization under interval uncertainty, *Cybern. Syst. Anal.*, **35** (1999), 297–306. https://doi.org/10.1007/BF02733477
16. T. Saeed, S. Treanţă, New classes of interval-valued variational problems and inequalities, *Results Control Optim.*, **13** (2023), 100324. https://doi.org/10.1016/j.rico.2023.100324
17. M. Ciontescu, S. Treanţă, On some connections between interval-valued variational control problems and the associated inequalities, *Results Control Optim.*, **12** (2023), 100300. https://doi.org/10.1016/j.rico.2023.100300
18. Y. Guo, G. Ye, W. Liu, D. Zhao, S. Treanţă, Solving nonsmooth interval optimization problems based on interval-valued symmetric invexity, *Chaos, Solitons Fractals*, **174** (2023), 113834. https://doi.org/10.1016/j.chaos.2023.113834
19. S. Treanţă, T. Saeed, On weak variational control inequalities via interval analysis, *Mathematics*, **11** (2023), 2177. https://doi.org/10.3390/math11092177
20. I. Hisao, T. Hideo, Multiobjective programming in optimization of the interval objective function, *Eur. J. Oper. Res.*, **48** (1990), 219–225. https://doi.org/10.1016/0377-2217(90)90375-L
21. S. T. Liu, R. T. Wang, A numerical solution method to interval quadratic programming, *Appl. Math. Comput.*, **189** (2007), 1274–1281. https://doi.org/10.1016/j.amc.2006.12.007
22. C. Jiang, X. Han, G. Liu, G. Liu, A nonlinear interval number programming method for uncertain optimization problems, *Eur. J. Oper. Res.*, **188** (2008), 1–13. https://doi.org/10.1016/j.ejor.2007.03.031
23. A. Jayswal, I. Stancu-Minasian, I. Ahmad, On sufficiency and duality for a class of interval-valued programming problems, *Appl. Math. Comput.*, **218** (2011), 4119–4127. https://doi.org/10.1016/j.amc.2011.09.041

24. M. Hladık, Interval linear programming: a survey, in *Linear Programming-New Frontiers in Theory and Applications*, (2012), 85–120.

25. A. Bellet, Y. Liang, A. B. Garakani, M. F. Balcan, F. Sha, A distributed Frank-Wolfe algorithm for communication-efficient sparse learning, in *Proceedings of the 2015 SIAM International Conference on Data Mining (SDM)*, (2015), 478–486. https://doi.org/10.1137/1.9781611974010.54

26. G. Qu, N. Li, Accelerated distributed Nesterov gradient descent, *IEEE Trans. Autom. Control*, **65** (2020), 2566–2581. https://doi.org/10.1109/TAC.2019.2937496

27. A. Nedic, A. Olshevsky, W. Shi, Achieving geometric convergence for distributed optimization over time-varying graphs, *SIAM J. Optim.*, **27** (2017), 2597–2633. https://doi.org/10.1137/16M1084316

28. S. Liang, L. Y. Wang, G. Yin, Exponential convergence of distributed primal–dual convex optimization algorithm without strong convexity, *Automatica*, **105** (2019), 298–306. https://doi.org/10.1016/j.automatica.2019.04.004

29. X. Yi, S. Zhang, T. Yang, K. H. Johansson, T. Chai, Exponential convergence for distributed optimization under the restricted secant inequality condition, *IFAC-PapersOnLine*, **53** (2020), 2672–2677. https://doi.org/10.1016/j.ifacol.2020.12.383

30. S. Treanţă, Lu-optimality conditions in optimization problems with mechanical work objective functionals, *IEEE Trans. Neural Networks Learn. Syst.*, **33** (2021), 4971–4978. https://doi.org/10.1109/TNNLS.2021.3066196

31. H. C. Wu, On interval-valued nonlinear programming problems, *J. Math. Anal. Appl.*, **338** (2008), 299–316. https://doi.org/10.1016/j.jmaa.2007.05.023

32. B. T. Polyak, *Introduction to Optimization*, Chapman and Hall, 1987.

33. R. Durrett, *Probability: Theory and Examples*, Cambridge University Press, 2010. https://doi.org/10.1017/CBO9780511779398

34. T. Maeda, On optimization problems with set-valued objective maps: existence and optimality, *J. Optim. Theory Appl.*, **153** (2012), 263–279. https://doi.org/10.1007/s10957-011-9952-x

35. A. K. Bhurjee, G. Panda, Efficient solution of interval optimization problem, *Math. Methods Oper. Res.*, **76** (2012), 273–288. https://doi.org/10.1007/s00186-012-0399-0

36. S. S. Ram, A. Nedić, V. V. Veeravalli, Distributed stochastic subgradient projection algorithms for convex optimization, *J. Optim. Theory Appl.*, **147** (2010), 516–545. https://doi.org/10.1007/s10957-010-9737-7

37. A. Nedic, A. Ozdaglar, Distributed subgradient methods for multi-agent optimization, *IEEE Trans. Autom. Control*, **54** (2009), 48–61. https://doi.org/10.1109/TAC.2008.2009515

38. A. P. Ruszczyński, A. Ruszczynski, *Nonlinear Optimization*, Princeton University Press, 2006. https://doi.org/10.1515/9781400841059

39. A. Nedic, A. Ozdaglar, P. A. Parrilo, Constrained consensus and optimization in multi-agent networks, *IEEE Trans. Autom. Control*, **55** (2010), 922–938. https://doi.org/10.1109/TAC.2010.2041686

40. A. Nedić, A. Olshevsky, Stochastic gradient-push for strongly convex functions on time-varying directed graphs, *IEEE Trans. Autom. Control*, **61** (2016), 3936–3947. https://doi.org/10.1109/TAC.2016.2529285

Appendix

Proof of Lemma 5

Proof. According to Assumption 1(b), the adjacency matrix $\mathcal{A}$ is irreducible and aperiodic. With [33, Theorem 6.64], $\lim_{k\to\infty}\mathcal{A}^k = \mathcal{B}$ with a linear convergence rate $\gamma_1 \in (0,1)$, where $\mathcal{B} = \frac{1}{n}\mathbf{1}_n^\top\mathbf{1}_n$. With (3.9b), we have

$$\lim_{k\to\infty} \boldsymbol{z(k)} = \lim_{k\to\infty}\mathcal{A}^k\boldsymbol{z(0)} = \mathcal{B}\boldsymbol{z(0)} = \bar{\boldsymbol{z}}\boldsymbol{(0)}. \tag{A1}$$

According to [37, Lemma 3], $\sum_{k=1}^{\infty}\left\|A^k - B\right\| < \infty$ holds, which completes the proof.

Proof of Lemma 6

Proof. (a) Lower bound of the Lyapunov function $V(\boldsymbol{w}, z)$: Let $\boldsymbol{w}^* = col\{\boldsymbol{s}^*, \boldsymbol{t}^*\}$ be the projection of $\boldsymbol{w}^k$ onto the optimal set W^*. Since the symmetry of $\boldsymbol{P}$ holds, given Lemma 4, $\nabla F_{s^*}(\boldsymbol{x^*}, \bar{z}^{\mathbf{0}}) = -\boldsymbol{P}\boldsymbol{t}^*$ and $\langle \boldsymbol{s^*}, \boldsymbol{P}\rangle = 0$. We further obtain that $\langle \boldsymbol{s}-\boldsymbol{s}^*, \nabla F_{s^*}(\boldsymbol{x^*}, \bar{z}^{\mathbf{0}})\rangle = -\langle \boldsymbol{s}-\boldsymbol{s}^*, \boldsymbol{P}\boldsymbol{t}^*\rangle$, and $\langle \boldsymbol{s}, \boldsymbol{P}\boldsymbol{t}\rangle = \langle \boldsymbol{s}-\boldsymbol{s}^*, \boldsymbol{P}\boldsymbol{t}\rangle$. $V_b(\boldsymbol{w}, z)$ can be further written as

$$\begin{aligned}
V_b(\boldsymbol{w}, z) =& F(\boldsymbol{s}, z) - F(\boldsymbol{s^*}, \bar{z}^{\mathbf{0}}) + \frac{1}{2}\langle \boldsymbol{s}, \boldsymbol{P}\boldsymbol{s}\rangle + \langle \boldsymbol{s}, \boldsymbol{P}\boldsymbol{t}\rangle \\
=& F(\boldsymbol{s}, z) - F(\boldsymbol{s^*}, \bar{z}^{\mathbf{0}}) + \langle \boldsymbol{s} - \boldsymbol{s}^*, \boldsymbol{P}(\boldsymbol{t} - \boldsymbol{t}^*)\rangle + \langle \boldsymbol{s} - \boldsymbol{s}^*, \boldsymbol{P}\boldsymbol{t}^*\rangle + \frac{1}{2}\langle \boldsymbol{s} - \boldsymbol{s}^*, \boldsymbol{P}(\boldsymbol{s} - \boldsymbol{s}^*)\rangle \\
=& F(\boldsymbol{s}, z) - F(\boldsymbol{s^*}, \bar{z}^{\mathbf{0}}) - \langle \boldsymbol{s} - \boldsymbol{s}^*, \nabla F_{s^*}(\boldsymbol{x^*}, \bar{z}^{\mathbf{0}})\rangle + \langle \boldsymbol{s} - \boldsymbol{s}^*, \boldsymbol{P}(\boldsymbol{t} - \boldsymbol{t}^*)\rangle + \frac{1}{2}\langle \boldsymbol{s} - \boldsymbol{s}^*, \boldsymbol{P}(\boldsymbol{s} - \boldsymbol{s}^*)\rangle.
\end{aligned} \tag{A2}$$

According to Lemma 3, $F(\boldsymbol{s}, z)$ is convex with respect to $\boldsymbol{s}$ and Lipschitz continuous with respect to z. Therefore,

$$\begin{aligned}
& F(\boldsymbol{s}, z) - F(\boldsymbol{s^*}, \bar{z}^{\mathbf{0}}) - \langle \boldsymbol{s} - \boldsymbol{s}^*, \nabla F_{s^*}(\boldsymbol{x^*}, \bar{z}^{\mathbf{0}})\rangle \\
=& F(\boldsymbol{x^*}, \bar{z}^{\mathbf{0}}) - F(\boldsymbol{x}, \bar{z}^{\mathbf{0}}) - \langle \boldsymbol{s} - \boldsymbol{s}^*, \nabla F_{s^*}(\boldsymbol{x^*}, \bar{z}^{\mathbf{0}})\rangle + F(\boldsymbol{s}, z) - F(\boldsymbol{x}, \bar{z}^{\mathbf{0}}) \geqslant -K_2\|z - \bar{z}^{\mathbf{0}}\| = -K_2\|z - \bar{z}^{\mathbf{0}}\|.
\end{aligned}$$

Since $\boldsymbol{P}$ is positive semidefinite,

$$\frac{1}{2}\langle \boldsymbol{s} - \boldsymbol{s}^*, \boldsymbol{P}(\boldsymbol{s} - \boldsymbol{s}^*)\rangle \geqslant 0.$$

Therefore, $V_b(\boldsymbol{w}, z) \geqslant \langle \boldsymbol{s} - \boldsymbol{s}^*, \boldsymbol{P}(\boldsymbol{t} - \boldsymbol{t}^*)\rangle \geqslant -\frac{\sigma}{2}\Big[\|\boldsymbol{s} - \boldsymbol{s}^*\|^2 + \|\boldsymbol{t} - \boldsymbol{t}^*\|^2\Big] - K_2\|z - \bar{z}^{\mathbf{0}}\|$, which implies the lower bound of the Lyapunov function $V(\boldsymbol{w}, z) \geqslant \frac{\sigma}{2}\Big[\|\boldsymbol{s} - \boldsymbol{s}^*\|^2 + \|\boldsymbol{t} - \boldsymbol{t}^*\|^2\Big]$.

(b) Upper bound of the Lyapunov function $V(\boldsymbol{w}, z)$: According to Lemma 3 $\Big(L$-Lipschitz continuity of $\nabla F_s(s, z)$ with respect to $\boldsymbol{s}\Big)$ and Assumption 2, $F(\boldsymbol{s}, \bar{z}^{\mathbf{0}}) - F(\boldsymbol{s^*}, \bar{z}^{\mathbf{0}}) - \langle \boldsymbol{s} - \boldsymbol{s}^*, \nabla F_{s^*}(\boldsymbol{x^*}, \bar{z}^{\mathbf{0}})\rangle \leqslant$

$\frac{L}{2}\|s - s^*\|^2$. According to Lemma 3, $F(s, z)$ is Lipschitz continuous with respect to z, we have that $F(s, z) - F(x, \bar{z}^0) \leqslant K_2\|z - \bar{z}^0\|$. Note that

$$\frac{1}{2}\langle s - s^*, P(s - s^*)\rangle \leqslant \frac{\sigma}{2}\|s - s^*\|^2.$$

Moreover, $\langle s - s^*, P(t - t^*)\rangle \leqslant \sigma\|s - s^*\| \cdot \|t - t^*\| \leqslant \frac{\sigma}{2}\Big[\varepsilon\|s - s^*\|^2 + \frac{1}{\varepsilon}\|t - t^*\|^2\Big]$ for any $\varepsilon > 0$. Through choosing $\varepsilon = \frac{\sigma}{K_1+\sigma}$, we get

$$V_b(w, z) \leqslant \frac{L+\sigma}{2}\Big[(\|s - s^*\|^2 + \|z - \bar{z}^0\|^2)\Big] + \frac{\sigma^2}{2(K_1+\sigma)}\|s - s^*\|^2 \leqslant \frac{K_1 + 2\sigma}{2}\Big[\|s - s^*\|^2 + \|t - t^*\|^2\Big],$$

which implies that $V(w, z) \leqslant \frac{K_1+4\sigma}{2}\Big[\|s - s^*\|^2 + \|t - t^*\|^2\Big] + 2K_2\|z - \bar{z}^0\|$.

Proof. Proof of Theorem 1

It follows from the K_1-Lipschitz of $\nabla F(s, z)$ in Lemma 3 that

$$\begin{aligned} F(s^{k+1}, z^{k+1}) - F(s^k, z^k) =& F(s^{k+1}, z^{k+1}) - F(s^{k+1}, z^k) + F(s^{k+1}, z^k) - F(s^k, z^k) \\ \leqslant& \Big\langle \nabla F_{s^k}(s^k, z^k), s^{k+1} - s^k\Big\rangle + \frac{K_1}{2}\Big\|s^{k+1} - s^k\Big\|^2 + K_2\Big\|z^{k+1} - z^k\Big\| \\ \leqslant& -h\Big\langle \nabla F_{s^k}(s^k, z^k), I_1(w^k, z^k)\Big\rangle + \frac{h^2K_1}{2}\Big\|I_1(w^k, z^k)\Big\|^2 + K_2\Big\|z^{k+1} - z^k\Big\|, \end{aligned} \tag{A3}$$

where the second inequality builds on the definition of $I_1(w)$. Since $\|P\| \leqslant \sigma$, we have

$$\Big\langle s^{k+1}, Ps^{k+1}\Big\rangle - \Big\langle s^k, Ps^k\Big\rangle \leqslant -2h\Big\langle I_1(w^k, z^k), Ps^k\Big\rangle + h^2\sigma\Big\|I_1(w^k, z^k)\Big\|^2 \tag{A4}$$

and

$$\begin{aligned} &\Big\langle y^{k+1}, Ps^{k+1}\Big\rangle - \Big\langle t^k, Ps^k\Big\rangle \\ &\leqslant -h\Big\langle I_2(w(k), z^k), Ps^k\Big\rangle + \frac{h^2\sigma}{2}\Big\|I_2(w^k, z^k)\Big\|^2 - h\Big\langle I_1(w^k, z^k), Pt^k\Big\rangle + \frac{h^2\sigma}{2}\Big\|I_1(w^k, z^k)\Big\|^2. \end{aligned} \tag{A5}$$

Combine (A3)–(A5). Given the definition of $V_b(w, z)$, we get

$$\begin{aligned} &V_b(w^{k+1}, z^{k+1}) - V_b(w^k, z^k) \\ &\leqslant -h\Big\|I_1(w^k, z^k)\Big\|^2 + h\Big\|I_2(w^k, z^k)\Big\|^2 + \frac{h^2(K_1+2\sigma)}{2}\Big\|I_1(w^k, z^k)\Big\|^2 + \frac{h^2\sigma}{2}\Big\|I_2(w^k, z^k)\Big\|^2 + K_2\Big\|z^{k+1} - z^k\Big\|, \end{aligned} \tag{A6}$$

which is based on $\langle Px, t\rangle = \langle Py, s\rangle$.

With Lemma 5 and $\|P\| \leqslant \sigma$, we obtain

$$\begin{aligned} &\langle -I(w^k, z^k), w^k - w^*\rangle \\ =& -\Big\langle s^k - s^*, \nabla F_{s^k}(s^k, z^k) + Pt^k + Ps^k\Big\rangle + \Big\langle t^k - t^*, Ps^k\Big\rangle \end{aligned}$$

$$= -\left\langle s^k - s^*, \nabla F_{s^k}(s^k, z^k)\right\rangle - \left\langle s^k, Pt^*\right\rangle - \left\langle s^k - s^*, Ps\right\rangle$$
$$= -\left\langle s^k - s^*, \nabla F_{s^k}(s^k, z^k) - \nabla F_{s^*}(s^*, \bar{z}^0)\right\rangle - \left\langle s^k, Ps^k\right\rangle$$
$$= -\left\langle s^k - s^*, \nabla F_{s^k}(s^k, z^k) - \nabla F_{s^*}(s^*, z^k)\right\rangle - \left\langle s^k - s^*, \nabla F_{s^*}(s^*, z^k) - \nabla F_{s^*}(s^*, \bar{z}^0)\right\rangle - \left\langle s^k, Ps^k\right\rangle. \quad \text{(A7)}$$

Since $F(\cdot)$ is a convex function with respect to s,

$$\left\langle s^k - s^*, \nabla F_{s^k}(s^k, z^k) - \nabla F_{s^*}(s^*, z^k)\right\rangle \geqslant 0. \quad \text{(A8)}$$

According to the K_3-Lipschitz continuity of $\nabla F_s(s, z)$ in Lemma 3, we have

$$\left\langle s^k - s^*, \nabla F_{s^*}(s^*, z^k) - \nabla F_{s^*}(s^*, \bar{z}^0)\right\rangle \geqslant -K_3\left\|z^k - \bar{z}^0\right\| \cdot \left\|s^k - s^*\right\|. \quad \text{(A9)}$$

Combining (A8) and (A9) with (A7) yields

$$\langle -I(w^k, z^k), w^k - w^*\rangle \leqslant -\frac{1}{\sigma}\left\|Ps^k\right\|^2 + K_3\left\|z^k - \bar{z}^0\right\| \cdot \left\|s^k - s^*\right\|. \quad \text{(A10)}$$

According to the σ-Lipschitz continuity of $V_a(w, z)$with respect to z in (3.12), we have

$$\begin{aligned} V_a(w^{k+1}, z^{k+1}) - V_a(w^k, z^k) &\leqslant \langle \nabla V_{a_w}(w^k, z^k), w^{k+1} - w^k\rangle + \frac{\sigma}{2}\left\|w^{k+1} - w^k\right\|^2 \\ &\leqslant -2h\sigma\langle I(w^k, z^k), w^k - w^*\rangle + \frac{\sigma}{2}\left\|w^{k+1} - w^k\right\|^2 \\ &\leqslant -2h\left\|I_2(w^k, z^k)\right\|^2 + \frac{\sigma h^2}{2}\left\|I(w^k, z^k)\right\|^2 + 2hK_3\sigma\left\|z^k - \bar{z}^0\right\| \cdot \left\|s^k - s^*\right\|. \end{aligned} \quad \text{(A11)}$$

Therefore, by using (A6)m (A11) and the definition of $V(w, z)$ in (3.12), we have

$$V(w^{k+1}, z^{k+1}) - V(w^k, z^k) \leqslant -\frac{h(2 - \nu_0 h)}{2}\left\|I(w^k, z^k)\right\|^2 + 2hK_3\sigma\left\|z^k - \bar{z}^0\right\| \cdot \left\|s^k - s^*\right\| + K_2 M^k \quad \text{(A12)}$$

where $\nu_0 = 4\sigma + K_1$. and $M^k = \left\|z^{k+1} - z^k\right\| + \left\|z^{k+1} - \bar{z}^0\right\| - \left\|z^k - \bar{z}^0\right\|$.

According to Lemma 5 and Assumption 2,

$$\sum_{k=1}^{\infty} 2hK_3\sigma\left\|z^k - \bar{z}^0\right\| \cdot \left\|s^k - s^*\right\| = \sum_{k=1}^{\infty} 2hK_3\sigma\left\|(\mathcal{A}^k - \mathcal{B})z^0\right\|\left\|s^k - s^*\right\| < \infty, \quad \text{(A13)}$$

and

$$\sum_{k=1}^{\infty} K_2 M^k = \sum_{k=1}^{\infty} K_2\left(\left\|(\mathcal{A} - I_n)\mathcal{A}^k z^0\right\| + \left\|(\mathcal{A} - I_n)(\mathcal{A}^k - \mathcal{B})z^0\right\|\right) < \infty. \quad \text{(A14)}$$

Consequently, with Lemma 1, $V(w^k, z^k)$ converges with $\sum_{k=0}^{\infty}\left\|I(w^k, z^k)\right\|^2 < +\infty$, which implies that $\lim_{k\to\infty} I(w^k, z^k) = 0$. By Lemma 4 and the continuity of I, $\lim_{k\to\infty}\left\|s^k - s^*\right\| = 0$ and $\lim_{k\to\infty}\left\|t^k - t^*\right\| = 0$.