Received 15 September 2023, accepted 16 October 2023, date of publication 23 October 2023, date of current version 7 November 2023.
Digital Object Identifier 10.1109/ACCESS.2023.3327199

RESEARCH ARTICLE

Distributed Interval Optimization Over Time-Varying Networks: A Numerical Programming Perspective

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]Academy of Mathematics and Systems Science, University of Chinese Academy of Sciences, Beijing 100049, China
[4]Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
Corresponding author: Jiuwei Wang (waliuss@sina.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 72101026, Grant 62203045, and Grant 62173030; and in part by the Fundamental Research Fund for the Central Universities under Grant FRF-TP-22-141A1.

ABSTRACT In this study, we investigate a distributed interval optimization problem involving agents linked by a time-varying network, optimizing interval objective functions under global convex constraints. Through scalarization, we first reformulate the distributed interval optimization problem as a distributed constrained optimization problem. The optimal Pareto solutions to the reformulated problem are then illustrated. We establish a distributed subgradient-free algorithm for the distributed constrained optimization problems by generating random differences of reformulated optimal objective functions, and the optimal solutions of the distributed constrained optimization problem are equivalent to Pareto optimal solutions of the distributed interval optimization problem. Moreover, we demonstrate that a Pareto optimal solution can be reached over the time-varying network using the proposed algorithm almost surely. FInally, we conclude with a numerical simulation to demonstrate the effectiveness of the proposed algorithm.

INDEX TERMS Distributed interval optimization, time-varying network, Pareto optimal solution, subgradient-free algorithm.

I. INTRODUCTION

Recently, distributed optimization and control in a network environment have attracted a growing amount of interest, as they are more effective than centralized designs for many large-scale problems when agents only have access to local information and exchange data with their neighbours over the network. In fact, distributed algorithms for a variety of (constrained) optimization problems have been extensively studied, with potential applications to sensor networks, smart grids, and equation solutions (see [1], [2], [3], [4], [5], [6], [7], [8]). Note that connectivity is an important aspect of distributed design. Although fixed topologies are still required for distributed optimization designs in certain circumstances, time-varying jointly connected networks have been considered in a number of algorithms, including [1], [3], [7], and [9].

Nonetheless, objective functions and constraints of some practical optimization problems may not be described precisely or explicitly. For instance, some conditions in power systems may be time-varying or uncertain, and data mining can produce inaccurate results (see [10], [11], and [12]). Motivated by the present setting, interval optimization is investigated in [13], [14], and [15], which provides a framework for capturing the uncertainty in optimization. In fact, interval optimization problems (IOP), which were first proposed by [13] and further studied in [14] and [15] and references thereto, have been extensively studied in numerous research fields, including economic systems [16] and smart grids [17]. Objective

The associate editor coordinating the review of this manuscript and approving it for publication was Xiwang Dong.

functions in an interval optimization problem are interval-valued, i.e., they are described by intervals rather than real integers, via interval-valued maps. The well-defined partial orderings and convexity of interval-valued maps [16], [18], [19] guarantee the existence of solutions for maximization and minimization of interval optimization problems. Literature (referring to [20], [21], [22], [23]) has provided various programming methods, particularly based on Wolfe's method or Lamke's algorithm (Lingo software provides different algorithm boxes to solve linear programming problems), to address centralized interval optimization problems.

With this background, it is nature for us to consider the construction of effective algorithms for distributed interval optimization problems over multi-agent networks. However, the distributed interval optimization problems are still under investigation. It may be because it is so simple to distribute the ideas of Wolfe or Lamke's algorithms, and very few papers with related theoretical results have been published on the subject of [24]. A further reason is that the partial order resulting from the interval makes gradient-based methods challenging, particularly when only local information is available in a distributed design.

In systems and control, randomization and stochastic methods have proven to be effective instruments. In contrast to traditional robustness methods [25], stochastic gradients are advantageous for controlling uncertain systems, for instance. When developing distributed algorithms, for instance [26], randomization and stochastic methods can enhance the system's overall performance as they are quite natural for the study of network dynamics and have a close relationship with real-world systems.

The purpose of this paper is to propose a distributed algorithm for interval optimization problems, based on recent results on distributed subgradient-free algorithms that overcome the challenge of obtaining subgradient information of local interval-valued functions. Zeroth-order/subgradient-free algorithms have been extensively researched in [27], [28], [29], [30], and [31] and references listed therein due to their applications in situations where obtaining gradient/subgradient information is computationally expensive or even impractical. Therefore, we propose a subgradient-free stochastic algorithm for a class of interval optimization problems based on numerical programming. The contributions of this paper are summarized as follows:

(a) In keeping with the rapid evolution of data science and engineering systems, we extend the centralized interval optimization problem [14], [15] to a distributed setting. Through well-defined partial orderings and convexity [16], [18], [19] of the distributed interval optimization problem, we transform it into a solvable distributed optimization problem with convex global constraints. The optimal solutions of the distributed constrained optimization problem are equivalent to Pareto optimal solutions of the distributed interval optimization problem in this reformulation. Wolfe's and Lamke's methods cannot be easily extended to distributed versions; therefore, we employ zeroth-order or subgradient-free ideas in the distributed design.

(b) We develop a new distributed subgradient-free algorithm with random differences for solving the reformulated distributed constrained non-smooth optimization problem because it is challenging to acquire the subgradient of the interval optimization problem. The algorithm employs random differences to approximate subgradients of locally reformulated objective functions, which is distinct from many existing distributed gradient/subgradient-free algorithms see, [32], [33], [34], and [35] though consistent with them when the local objective function is smooth.

(c) The proposed algorithm, which is related to the distributed stochastic optimization algorithm, is subjected to both theoretical and numerical analysis. We establish the mean-square convergence rate of $O(\frac{1}{\sqrt{k}})$. after establishing the consensus of estimates and accomplishment of global minimization with probability one with the proposed algorithm. With decreasing step-size, the convergence results match the best of the first-order stochastic algorithms [4], [36], [37]

The main contributions of this article could also be outlined in fig. 1. The remaining sections are organized as follows. In Section II, preliminary information regarding the analysis and design of distributed interval optimization is provided. The distributed interval optimization problem is then formulated and the corresponding distributed algorithm is presented in Section III, while Section IV analyses the proposed algorithm. The following numerical example is provided in Section V. The section VI concludes with some concluding remarks.

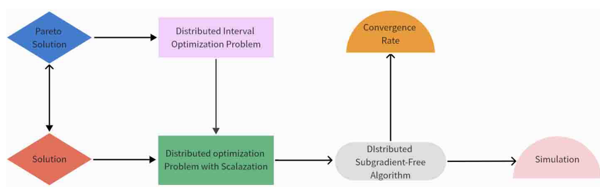

FIGURE 1. Outline.

Notations: Let $\mathcal{R}^p$ be the p-dimensional Euclidean space. Denote $\mathcal{R}^p_+$ as its non-negative orthants. $\|\cdot\|$ denotes the Euclidean norm. Denote the sets of all non-empty compact intervals of $\mathcal{R}$ by $\mathcal{C}(\mathcal{R})$.

II. MATHEMATICAL PRELIMINARIES

This section introduces mathematical prerequisites for convex analysis [3], [38], [39], probability theory [40], [41] and interval optimization.

A. CONVEX ANALYSIS

Here are some concepts about convex analysis [38], [39].

Definition 1 ([38] Sub-Gradient): Let $f(x) : \mathcal{R}^p \to \mathcal{R}$ be a non-smooth convex function. Vector-valued function $\nabla f(x) \in \partial f(x) \subset \mathcal{R}^p$ is called the subgradient of $f(x)$ if for any $x, y \in dom(f)$, the following inequality holds:

$$f(x) - f(y) - \langle \nabla f(y), x - y \rangle \geqslant 0.$$

Lemma 1 ([39] Lebourg's Mean Value Theorem): Let x, $y \in X$. Suppose $f(x) : \mathcal{R}^m \to \mathcal{R}$ is Lipschitz on an open set containing line segment $[x, y]$. Then, there exists a point $u \in (x, y)$ such that $f(x) - f(y) \in \langle \partial f(u), x - y \rangle$.

Then, we summarize the Euclidean norm inequalities [3], [39] that will be used in this paper.

Lemma 2 [4]: Let $x_1, x_2, \ldots, x_n$ be vectors in $\mathcal{R}^p$. Then

$$\sum_{i=1}^{n} \left\| x_i - \frac{1}{n}\sum_{i=1}^{n} x_j \right\|^2 \leqslant \sum_{i=1}^{n} \left\| x_i - x \right\|^2, \quad \forall x \in \mathcal{R}^p.$$

Denote the projection of x onto set X by $P_X(x)$, i.e., $P_X(x) = \arg\min_{y \in X} \|x - y\|$, where X is a closed bounded convex set in $\mathcal{R}^p$. The following results are on the projection operators in Euclidean norm:

Lemma 3 [3], [38]: Let X be a closed convex set in $\mathcal{R}^p$. Then for any $x \in \mathcal{R}^p$, it holds that

(a) $\langle x - P_X(x), y - P_X(x) \rangle \leqslant 0$, *for all* $y \in X$

(b) $\|P_X(x) - P_X(y)\| \leqslant \|x - y\|$, *for all* $x, y \in \mathcal{R}^m$.

(c) $\langle x - y, P_X(y) - P_X(x) \rangle \leqslant -\|P_X(x) - P_X(y)\|^2$, *for all* $y \in \mathcal{R}^m$.

(d) $\|x - P_X(x)\|^2 + \|y - P_X(x)\|^2 \leqslant \|x - y\|^2$, *for any* $y \in X$.

B. PROBABILITY THEORY

Denote $(\Omega, \mathcal{F}, \mathbb{P})$ as the probability space, where Ω is the whole event space, $\mathcal{F}$ is the σ-algebra on Ω, and $\mathbb{P}$ is the probability measure on $(\Omega, \mathcal{F})$. Then, definitions of convergence in $(\Omega, \mathcal{F}, \mathbb{P})$ and convergence of super-martingales theorem is given.

Definition 2 ([40] Convergence in $(\Omega, \mathcal{F}, \mathbb{P})$): (a) *$x_1, x_2, \ldots, x_k \ldots$ is a sequence of random variables (r. v.) in $(\Omega, \mathcal{F}, \mathbb{P})$. If $P(x_k \to x) = 1$, then x_k converges x almost surely (a. s.).*

(b) *$x_1, x_2, \ldots, x_k \ldots$ is a sequence of random variables (r. v.) in $(\Omega, \mathcal{F}, \mathbb{P})$. If $\mathbb{E}\|x_k - x\|^p \to 0$, then x_k converges to x in L^p.*

Lemma 4: [41]: In $(\Omega, \mathcal{F}, \mathbb{P})$, denote $\{F(k)\}_{k\geq 1}$ as a sequence of increasing sub-σ-algebras on $\mathcal{F}$. $\{h(k)\}_{k\geq 1}$, $\{v(k)\}_{k\geq 1}$ and $\{w(k)\}_{k\geq 1}$ are variable sequences in $\mathcal{R}$ such that for each k, $h(k)$, $v(k)$ and $w(k)$ are $F(k)$-measurable. Both $\{v(k)\}_{k\geq 1}$ and $\{w(k)\}_{k\geq 1}$ are nonnegative and $\sum_{k=1}^{\infty} w(k) < \infty$. Moreover, $\{h(k)\}_{k\geq 1}$ is bounded from below uniformly. If

$$\mathbb{E}[h(k+1)|F(k)] \leqslant (1 + \eta(k))h(k) - v(k) + w(k), \quad \forall k \geqslant 1$$

holds almost surely, where $\eta(k) \geqslant 0$ are constants with $\sum_{k=1}^{\infty} \eta(k) < \infty$, then $\{h(k)\}_{k\geq 1}$ converges almost surely with $\sum_{k=1}^{\infty} v(k) < \infty$.

C. INTERVAL OPTIMIZATION

1) ORDERINGS ON $\mathcal{C}(\mathcal{R})$ AND PROPERTIES OF INTERVAL-VALUED MAP G

Define $A = [a_L, a_R]$, $B = [b_L, b_R]$ are two non-empty compact intervals in $\mathcal{P}(\mathcal{R}^q)$. Now, we introduce some orderings on $\mathcal{P}(\mathcal{R})$.

Definition 3 [16], [18]: For any $A, B \in \mathcal{P}(\mathcal{R})$. denote:

(a) *$A \leqq_L B$ iff $a_L \leqslant b_L$;*

(b) *$A \leqq_U B$ iff $a_R \leqslant b_R$;*

(c) *$A \leqq B$ iff $A \leqq_L B$ and $A \leqq_U B$.*

Definition 4 [16], [18]: For any $A, B \in \mathcal{P}(\mathcal{R})$, denote:

(a) *$A <_L B$ iff $a_L < b_L$;*

(b) *$A <_U B$ iff $a_R < b_R$;*

(c) *$A < B$ iff $A <_L B$ and $A <_U B$;*

(d) *$A \leq B$ iff $A <_L B$ and $A \leqq_U B$, or $A \leqq_L B$ and $A <_U B$.*

Let $G : \mathcal{R}^p \rightrightarrows \mathcal{R}$ be any interval-valued map, where $G(\cdot)$ is an interval with respect to x. We give the definition of Lipschitz continuity and convexity of the interval-valued map $G : \mathcal{R}^p \rightrightarrows \mathcal{R}$ as follows:

Definition 5 ([42] Lipschitz Continuity): Let $G : \mathcal{R}^p \rightrightarrows \mathcal{R}$ be any interval-valued map. G is locally Lipschitz at x if there exist $K > 0$ and a neighborhood W of x such that

$$G(x_1) \subseteq G(x_2) + K\|x_1 - x_2\|, \quad \forall x_1, x_2 \in W.$$

We say that G is locally Lipschitz at x if there exist a neighbourhood W of x and a constant $K \geqslant 0$, such that

$$G(x_1) \subseteq B\big(G(x_2), K\|x_1 - x_2\|\big).$$

We denote by $B(A, \varrho) = \{y | d(y, A) \leqslant \varrho\}$, the ball of radius ϱ around subset A, where y is chosen from a metric space.

Definition 6 ([19] Convexity): Let $G : \mathcal{R}^p \rightrightarrows \mathcal{R}^q$ be any interval-valued map. G is convex (lower-convex or upper convex) on Ω if, $\forall x_1, x_2 \in \Omega$, $\forall \alpha \in [0, 1]$,

$$G\big(\alpha x_1 + (1-\alpha)x_2\big) \leqq (\leqq_L \text{ or } \leqq U)\alpha G(x_1) + (1-\alpha)G(x_2).$$

Remark 1: Suppose that G is compact-valued and convex, $G(\cdot) = [L(\cdot), R(\cdot)]$. Then according to Definitions 3 and 4, $L(\cdot)$, $R(\cdot) : \mathcal{R}^p \to \mathcal{R}$ are convex functions with respect to $x \in \mathcal{R}^p$. Namely, for any x_1, $x_2 \in \mathcal{R}^p$ and any $t \in [0, 1]$, the following inequalities hold:

$$L\big(tx_1 + (1-t)x_2\big) \leq slanttL(x_1) + (1-t)L(x_2),$$
$$R\big(tx_1 + (1-t)x_2\big) \leq slanttR(x_1) + (1-t)R(x_2).$$

2) INTERVAL OPTIMIZATION PROBLEMS WITH INTERVAL-VALUED MAP

Let $G : \mathcal{R}^p \rightrightarrows \mathcal{R}$ be any interval-valued map. Now we consider the following interval optimization problem:

$$(IOP) \qquad \min_x \quad G(x) \quad s.\,t. \quad x \in \Omega$$

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

Example 1: Motivated by the article [43], we give an example of interval valued function. Consider a function $G : \mathcal{R}^p \rightrightarrows \mathcal{R}$. Without loss of generality, consider c as an order set, which is influenced by orders maintained on *the presence of components of* $G(x)$. *If* $G(x_1, x_2) = c_1 x_1^2 + c_2 x_1 e^{c_3 x_2}$, *then* $c = [c_1, c_2, c_3]^\top$. *Suppose* c_1, c_2, c_3 *lies in intervals* C_1, C_2 *and* C_3, *respectively.* $C_i = [c_L^i, c_R^i]$, $c_i(t_i) = (1 - t_i)c_L^i + t_i c_R^i$, $t_i \in [0, 1]$, $i = 1, 2, 3$. *Thus* $C(t) = [c_1(t), c_2(t), c_3(t)]^\top \in C_v^3$, *where* C_v^k *stands for k-dimensinal column vecter whose elements are vectors. For the given interval vector* C_v^3, $\{G_{c(t)}(x_1, x_2) = c_1(t)x_1^2 + c_2(t)x_1 e^{c_3(t)x_2} : \mathcal{R}^2 \rightrightarrows \mathcal{R}, c(t) \in C_v^3\}$ *is an interval, where* $L(x) = \min_t G_{c(t)}(x_1, x_2)$ *and* $R(x) = \max_t G_{c(t)}(x_1, x_2)$.

In light of definitions of $L(x)$ and $R(x)$ of the example, we see that we can not get the explicit expressions of $L(x)$ and $R(x)$, and this (IOP) problem can be solved through set-valued optimization rather than vector valued optimization.

Based on quasi orderings of compact intervals in $\mathcal{C}(\mathcal{R})$ given in Definitions 3 and 4, we define the Parato optimal solution to (IOP):

Definition 7 [44]: A point $x^* \in \Omega$ *is said to be a Pareto optimal solution (PO) to (IOP) iff it holds that* $G(\bar{x}) \leqq G(x^*)$ *for some* $\bar{x} \in \Omega$ *implies* $G(x^*) \leqq G(\bar{x})$.

Remark 2: There is no solution to interval optimization problem given in fig. 2. However, $[x_1, x_2]$ *are Pareto optimal solutions to this given problem.*

(a) *For* $y \leqslant x_1$, *we have* $R(y) \geqslant R(x_1)$ *and* $L(y) \geqslant L(x_1)$, *which means that* $G(y) \geqq G(x_1)$.
(b) *For* $y \geqslant x_2$, *we have* $R(y) \geqslant R(x_2)$ *and* $L(y) \geqslant L(x_2)$, *which means that* $G(y) \geqq G(x_2)$.
(c) *For* $x_1 \leqslant y \leqslant x_2$, *we have* $R(y) \leqslant R(x_1)$, $L(y) \geqslant L(x_1)$, $R(y) \geqslant R(x_2)$ *and* $L(y) \leqslant L(x_2)$ *according to Definition 7. Therefore,* $[x_1, x_2]$ *are Pareto optimal solutions to this given problem.*

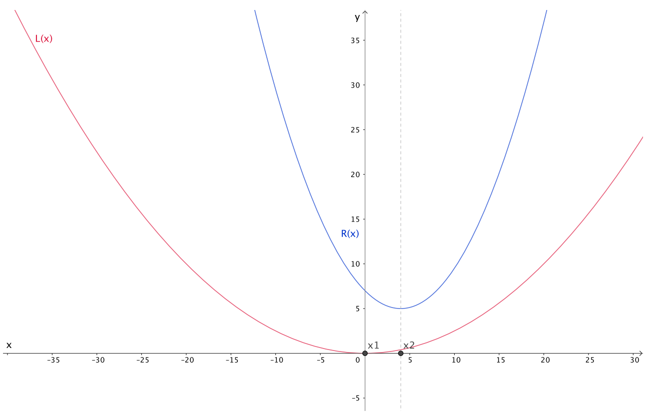

FIGURE 2. $L(x)$ **and** $R(x)$ **for vector** x**.**

Associated with (IOP), consider the following interval optimazation problem with its scalarization:

$$(SIOP) \quad \min_x \quad \lambda L(x) + (1 - \lambda)R(x)$$
$$s.t. \quad x \in \Omega$$

where $\lambda \in [0, 1]$.

The following lemma holds according to [44]. We gave its proof here just for self-reminder.

Lemma 5: 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* $x^* \in \Omega$ *is an optimal solution to (SIOP), then* $x^* \in \Omega$ *is a Pareto optimization to (IOP).*
(b) *A point* $x^* \in \Omega$ *is a Pareto optimization to (IOP), then there exists a real number* $\lambda \in [0, 1]$ *such that* $x^* \in \Omega$ *is an optimal solution to (SIOP).*

Proof:

(a) Given a real number $\lambda \in (0, 1)$ and let $x^* \in \Omega$ be an optimal solution to (SIOP). Suppose that there exists a point $\bar{x} \in \Omega$, such that $G(\bar{x}) \leqq G(x^*)$, which implies $L(\bar{x}) \leqslant L(x^*)$ and $R(\bar{x}) \leqslant R(x^*)$. Therefore,
$$\lambda L(\bar{x}) + (1 - \lambda)R(\bar{x}) \leqslant \lambda L(x^*) + (1 - \lambda)R(x^*)$$
which contradicts that x^* is an optimal solution to (SIOP).
(b) Let $x^* \in \Omega$ be a Pareto optimal solution to (IOP). Since G is compact-valued and convex with respect to x, according to Remark 1, $L(x)$ and $U(x)$ are convex functions. Following Definition 7, there exists a vector $\boldsymbol{\lambda} = [a, b]^\top \neq 0$, $a \geqslant 0$, $b \geqslant 0$, such that
$$\boldsymbol{\lambda}^\top \begin{bmatrix} L(x^*) \\ R(x^*) \end{bmatrix} \leqslant \boldsymbol{\lambda}^\top \begin{bmatrix} L(x) \\ R(x) \end{bmatrix}$$
holds for all $x \in \Omega$. Define $\bar{\boldsymbol{\lambda}} = [\frac{a}{a+b}, \frac{b}{a+b}]$, we have
$$\bar{\boldsymbol{\lambda}}^\top \begin{bmatrix} L(x^*) \\ R(x^*) \end{bmatrix} \leqslant \bar{\boldsymbol{\lambda}}^\top \begin{bmatrix} L(x) \\ R(x) \end{bmatrix},$$
which proves that there exists a real number $\lambda \in [0, 1]$ such that $x^* \in \Omega$ is an optimal solution to (SIOP).

■

III. DISTRIBUTED INTERVAL OPTIMIZATION

Consider the following distributed interval optimization problem over an n-agent network:

$$(DIOP) \quad \min_x \; G(\boldsymbol{x}) = \sum_{i=1}^{n} G_i(x_i)$$
$$s.t. \quad x_i = x_j, \quad x_i \in X \tag{1}$$

where $\boldsymbol{x} = [x_1^\top, x_2^\top, \ldots, x_n^\top]^\top \in \mathcal{R}^{np}$, $x_i \in \mathcal{R}^p$, and $G_i : \mathcal{R}^p \rightrightarrows \mathcal{R}$ is a compact and convex interval-valued function. In this setting, the state of an agent i is the estimate of solution to problem (DIOP). Each agent i knows local functions G_i and global constraint X. We first make the following assumption on the local functions and constraints for the distributed interval optimal problem (DIOP):

Assumption 1: (a) $G_i(x)$ *is a convex, compact, Lipschitz continuous interval-valued function.*
(b) X *is a non-empty, compact, convex constraint set in* $\mathcal{R}^p$.

Consider solving (DIOP) over a time-varying multi-generator network. Define a directed network $\mathcal{G}(k) = (\mathcal{N}, \mathcal{E}(k), W(k))$ as the communication topology among generators, where $\mathcal{N} = \{1, 2, \ldots n\}$ is the agent set, the edge set $\mathcal{E}(k) \subset \mathcal{N} \times \mathcal{N}$ represents information communication at time k and $W(k) = [w_{ij}(k)]_{ij}$ represents adjacency matrix at time k. Each agent interacts with its neighbors in $\mathcal{G}(k) = (\mathcal{N}, \mathcal{E}(k), W(k))$ at time k. We make the following assumption about communication topology $\mathcal{G}(k) = (\mathcal{N}, \mathcal{E}(k), W(k))$.

Assumption 2: The graph $\mathcal{G}(k) = (\mathcal{N}, \mathcal{E}(k), W(k))$ satisfies:

(a) *There exists a constant η with $0 < \eta < 1$ such that, $\forall k \geqslant 0$ and $\forall i, j$, $w_{ii}(k) \geqslant \eta$; $w_{ij}(k) \geqslant \eta$ if $(j, i) \in \mathcal{E}(k)$.*
(b) *$W(k)$ is doubly stochastic, i. e. $\sum_{i=1}^{m} w_{ij}(k) = 1$ and $\sum_{j=1}^{m} w_{ij}(k) = 1$.*
(c) *There is an integer $\kappa \geqslant 1$ such that $\forall k \geqslant 0$ and $\forall (j, i) \in \mathcal{N} \times \mathcal{N}$,*

$$(j, i) \in \mathcal{E}(k) \cup \mathcal{E}(k+1) \cup \cdots \cup \mathcal{E}(k+\kappa-1).$$

Assumption 2 reveals that each agent i can periodically collect data from all its neighbours. It is also a common connectivity condition for time-varying distributed network designs (see [1], [3]).

Define the function $f : \mathcal{R}^{np} \times \mathcal{R}^{n} \to \mathcal{R}$ and $f_i : \mathcal{R}^{p} \times [0, 1] \to \mathcal{R}$ as

$$f(\boldsymbol{x}, \boldsymbol{\lambda}) \triangleq \sum_{i=1}^{n} f_i(x_i, \lambda_i) \tag{2}$$

$$f_i(x_i, \lambda_i) \triangleq \lambda_i L_i(x) + (1-\lambda_i) R_i(x) \tag{3}$$

where $i = 1, 2, \ldots, n$, $\boldsymbol{x} = [x_1^\top, x_2^\top, \ldots, x_n^\top]^\top \in \mathcal{R}^{nq}$ and $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_n]^\top \in \mathcal{R}^{n}$.

Remark 3: Note that both $L(\boldsymbol{x})$ and $R(\boldsymbol{x})$ are separable, that is

$$L(\boldsymbol{x}) = \sum_{i=1}^{n} L_i(x_i),$$

and

$$R(\boldsymbol{x}) = \sum_{i=1}^{n} R_i(x_i).$$

We can convert the distributed interval optimization problem into a standard distributed optimization problem with scalar values. Let $\boldsymbol{\lambda} = \lambda_0 \mathbf{1}_n$ with $\lambda_0 \in (0, 1)$. To solve problem (1), the following distributed optimization problem is solved:

$$\begin{aligned} \min_{x} \quad & f(\boldsymbol{x}, \boldsymbol{\lambda}) = \sum_{i=1}^{n} f_i(x_i, \lambda_i) \\ s.t. \quad & x_i = x_j, \quad x_i \in X \\ & \lambda_i = \lambda_j \end{aligned} \tag{4}$$

where agent i is acquainted with the information of f_i, x_i, $\lambda_i \in (0, 1)$, and its vicinity.

Remark 4: In Problem (4), We need $\lambda_i = \lambda_j$ through the design of algorithms for the following reasons:

(a) *The aforementioned articles on interval optimization are centralized. This paper provides a framework for solving distributedly centralized interval optimization problems.*
(b) *The distributed interval optimization problem may also be applied to the resolution of stochastic problems involving distributed stripe disturbances. The iteratively generated λ reflects the intrinsic stripe properties of such distributed stochastic problems.*
(c) *The problem of distributed interval optimization could also be applied to privacy protection problems. In current distributed privacy protection settings, agents offer stochastic function information with noises to other agents. If each agent chooses to provide others with a confidence region (interval-valued function) without revealing λ_i, these methods can better protect their data.*
(d) *Reformulated problem (4) degenerates to a typical distributed constrained optimization problem [4] if each agent i chooses a common parameter λ or λ_is don't not need to be common. Due to the need for $\lambda_i = \lambda_j$ via iterations, the reformulated problem is more challenging than typical distributed problems.*

Remark 5: According to Definitions 5-6 and Assumption 1, we have:

(a) *Each $f_i(x, \lambda)$ is convex with respect to x, i. e. for any x_1, x_2:*

$$f_i(\alpha x_1 + (1-\alpha)x_2, \lambda) \leqslant \alpha f_i(x_1, \lambda) + (1-\alpha) f_i(x_2, \lambda),$$

where $\alpha \in [0, 1]$.
(b) *Each $f_i(x, \lambda)$ is convex with respect to λ.*
(c) *Each $f_i(x, \lambda)$ is Lipschitz continuous with respect to x, i. e. for all x_1, x_2 and λ:*

$$\|f_i(x_1, \lambda) - f_i(x_2, \lambda)\| \leqslant L\|x_1 - x_2\|.$$

(d) *Each $f_i(x, \lambda)$ is Lipschitz continuous with respect to λ i. e. for all λ_1, λ_2 and x:*

$$\|f_i(x, \lambda_1) - f_i(x, \lambda_2)\| \leqslant K\|\lambda_1 - \lambda_2\|.$$

(e) *$\|\partial f_{i_x}(x, \lambda)\| \leqslant L$ and $\partial \|f_{i_\lambda}(x, \lambda)\| \leqslant K$.*

Proof of (e): Suppose there exists a vector x, such that we can choose a subgradient $\nabla f_{i_x}(x, \lambda) \in \partial f_{i_x}(x, \lambda)$, where $\|\nabla f_{i_x}(x, \lambda)\| > L$. Suppose $y = x + \nabla f_{i_x}(x, \lambda)$, according to the definition of subgradient given in Definition 1, we have

$$\begin{aligned} f_i(y, \lambda) - f_i(x, \lambda) &\geqslant \langle \nabla f_{i_x}(x, \lambda), y - x \rangle \\ &\geqslant \|\nabla f_{i_x}(x, \lambda)\|^2 > L\|\nabla f_{i_x}(x, \lambda)\| \\ &> L\|y - x\|, \end{aligned}$$

which contradicts the Lipschitz continuity of $f_i(x, \lambda)$ with respect to x. By an analogous proof, $\partial f_{i_\lambda}(x, \lambda) \leqslant K$. ■

Lemma 6: If $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{R}^{np} \times \mathcal{R}^{n}$, is an optimal solution to problem (4), then $\boldsymbol{x}^*$ is a Pareto solution to problem (1).

Due to the fact that the differentiability of $f(\boldsymbol{x}, \boldsymbol{\lambda})$ with respect to x does not hold in general, we propose a Distributed (sub)gradient-free inteval-valued algorithm 1 for solving reformulated problem (4).

Algorithm 1 Distributed (sub)Gradient-Free Inteval-Valued Algorithm

Input: Total numbers of iteration T, step-size $\iota(k)$.
Initialize: $\xi_i \in X$ for all $i = 1, 2, \ldots n$.

for $k = 0, \ldots T$ **do**
  Average of local observations $x_i(k)$:

$$\xi_i(k) = \sum_{j=1}^{n} w_{ij}(k) x_j(k). \tag{5}$$

  Calculation of local measurement $d_i(k)$

$$d_i(k) = \frac{\left[y_i^+(k) - y_i^-(k)\right] \Delta_i^-(k)}{2c(k)}, \tag{6}$$

  Descent Step:

$$\hat{\xi}_i(k) = \xi_i(k) - \iota(k) d_i(k). \tag{7}$$

  Projection Step:

$$x_i(k+1) = P_X\left(\hat{\xi}_i(k)\right). \tag{8}$$

  Average of local observations $\lambda_i(k)$:

$$\lambda_i(k+1) = \sum_{j=1}^{n} w_{ij}(k) \lambda_i(k). \tag{9}$$

end for
where $d_i(k)$ is used as an estimate for $\partial f_{i_{\epsilon_i(k)}}\left(\xi_i(k), \lambda_i(k)\right)$.

In (6), $\Delta_i(k) = \left[\Delta_i^1(k), \Delta_i^2(k), \ldots, \Delta_i^p(k)\right]^\top$. $\Delta_i^-(k) = \left[\frac{1}{\Delta_i^1(k)}, \frac{1}{\Delta_i^2(k)}, \ldots, \frac{1}{\Delta_i^p(k)}\right]^\top$, where $\left\{\Delta_i^q(k)\right\}_{k\geqslant 0}$, $q = 1, 2, \ldots, p$, $k = 1, 2, \ldots$ is a sequence of mutually independent and identically distributed random variables with zero mean. The measurements $y_i^+(k)$ and $y_i^-(k)$ are given by

$$y_i^+(k) = f_i\left(\xi_i(k) + c(k) \underset{i}{\Delta}(k), \lambda_i(k)\right),$$
$$y_i^-(k) = f_i\left(\xi_i(k) - c(k) \underset{i}{\Delta}(k), \lambda_i(k)\right).$$

Define $F(k) = \sigma\{x_i(k), x_i(k-1), \ldots, x_i(0), i = 1, 2, \ldots, n; \lambda_i(k), \lambda_i(k-1), \ldots, \lambda_i(0), i = 1, 2, \ldots, n; \Delta_i(k-1), \Delta_i(k-2), \ldots, \Delta_i(0), \quad i = 1, 2, \ldots, n\}$, where $F(k)$ is the σ-algebra created by the whole history of Distributed (sub)gradient-free inteval-valued algorithm (Algorithm 1) up to moment k (referring to [4]). We further make the following hypotheses on the dither signal $\Delta_i(k)$:

Assumption 3: (a) *For any fixed (i, q), let $\left\{\Delta_i^q(k)\right\}_{k\geqslant 0}$ be a sequence of independent and identically distributed (i. i. d.) random variables. For all $k \geqslant 0$ and for any (i, q)*

$$\left|\underset{i}{\overset{q}{\Delta}}(k)\right| < M_1, \quad \left|\frac{1}{\Delta_i^q(k)}\right| < M_2, \quad \mathbb{E}\left[\frac{1}{\Delta_i^q(k)}\right] = 0.$$

(b) *For $i \neq j$ or $q \neq r$, $\left\{\Delta_i^q(k)\right\}_{k\geqslant 0}$ and $\left\{\Delta_j^r(k)\right\}_{k\geqslant 0}$ are mutually independent of each other.*

Then, we introduce conditions on the step-size $\iota(k)$ of Algorithm 1 and $c(k)$ used in the randomized differences (6):

Assumption 4: (a) $\iota(k) > 0$, $\sum_{k=1}^{\infty} \iota(k) < \infty$.
(b) $c(k) > 0$, $c(k) \to 0$.
(c) $\sum_{k=1}^{\infty} \frac{\iota(k)}{c(k)} = \infty$, $\sum_{k=1}^{\infty} \frac{\iota^2(k)}{c^2(k)} < \infty$.

The chosen of unite parameter $\frac{\iota(k)}{c(k)}$ satisfies the stochastic approximation stepsize condition in [27] and [36].

IV. PROPERTIES OF DISTRIBUTED (SUB)GRADIENT-FREE STOCHASTIC ALGORITHM

In this section, we first analyse that the estimate $(x_i(k), \lambda_i(k))$ converges to a consensus optimal point (x^*, λ^*) almost surely of Algorithm 1. Then, the mean-square convergence rate of Algorithm 1 is provided.

Denote the transition matrix of $W(k)$ as $\Psi(k, s) = W(k)W(k-1)\cdots W(s), k \geqslant s$, where $\left[\Psi(k, s)\right]_{ij}$ is the ij-th element of $\Psi(k, s)$. The following lemma about $\Psi(k, s)$ holds ture, given in Proposition 1 of [1].

Lemma 7: Under Assumptions 2, $\left|\left[\Psi(k, s)\right]_{ij} - \frac{1}{n}\right| \leqslant \delta\beta^{k-s}$, $\forall k > s$, where $\delta = 2\left(1 + \eta^{-K_0}\right)/\left(1 - \eta^{-K_0}\right)$, with $K_0 = (n-1)\kappa$ and $\beta = \left(1 - \eta^{-K_0}\right)^{1/K_0} < 1$.

First, we present a theorem regarding the proposed algorithm's convergence analysis.

Theorem 1: With Assumptions 1-4,

(a) *all the sequences $\{\lambda_i(k)\}$, $i \in \mathcal{N}$ reach consensus (which is depended by initail parameters $\lambda_i(0)'s$) almost surely by Algorithm 1.*
(b) *all the sequences $\{x_i(k)\}$, $i \in \mathcal{N}$ by Algorithm 1 converge to the same optimal point x^* in consensus and almost surely.*

The proof of Theorem 1 relies on Lemmas 8-12. Lemma 8 provides an upper bound for the Euclidean norm of $d_i(k)$ in expection; Lemma 9 investigates the consensus in the L_1 norm of estimates $x_i(k)$ in Algorithm r1; Lemma 10 investigates the lower bound of the cross term of $d_i(k)$ and $(\xi_i(k) - x^*)$in expection and in conditional expection with respect to $F(k)$, where x^* is the optimal solution of (4) for fixed common point λ^*; Lemma 11 analyses that $\{x_i(k)\}$, $i \in \mathcal{N}$ converge to the same random variable almost surely and Lemma 12 analyses that $\{x_i(k)\}$, $i \in \mathcal{N}$ converge to x^* in L_2. The proofs for these lemmas can be found in the Appendix.

Lemma 8: Let Assumption 1 and 3 hold. Then the first order moments and second moments of $d_i(k)$ are bounded by

$$\mathbb{E}\left\|d_i(k)\right\| \leqslant L, \quad \mathbb{E}\left\|d_i(k)\right\|^2 \leqslant L^2,$$

where L is the Lipschitz constant with respect to x in Remark 5.

Lemma 9: With Assumptions 1-4, for given common point λ^*, the consensus of estimate $x_i(k)$ in L_1 is achieved by Algorithm 1, that is, for $i, j = 1, 2, \ldots, n$,

$$\lim_{k\to\infty} \mathbb{E}\left\|x_i(k) - x_j(k)\right\| = 0.$$

Lemma 10: With Assumptions 1 and 3, the cross term of $d_i(k)$ and $\xi_i(k) - \xi^*$ is lower bounded

(a) *in conditional expection with respect to $F(k)$ as follows:*

$$\begin{aligned}&\mathbb{E}\left[\left\langle d_i(k),\ x_i(k) - \xi^*\right\rangle \middle| F(k)\right]\\ &\geqslant f_i\left(\bar{x}(k), \bar{\lambda}(k)\right) - f_i\left(x^*, \lambda^*\right) - L\left\|\xi_i(k) - \bar{x}(k)\right\| - B\\ &- K\left\|\lambda_i(k) - \bar{\lambda}(k)\right\| - K\left\|\lambda_i(k) - \lambda^*\right\| - c(k)L\left\|\bigtriangleup_i(k)\right\|,\end{aligned}$$

(b) *in expection as follows:*

$$\begin{aligned}&\mathbb{E}\left[\left\langle d_i(k),\ \xi_i(k) - x^*\right\rangle\right]\\ &\geqslant \mathbb{E}\left[f_i\left(\bar{x}(k), \lambda^*\right) - f_i\left(x^*, \lambda^*\right)\right] - L\mathbb{E}\left\|\xi_i(k) - \bar{x}(k)\right\|\\ &- 2K\mathbb{E}\left\|\lambda_i(k) - \lambda^*\right\| - c(k)L\mathbb{E}\left\|\bigtriangleup_i(k)\right\| - B,\end{aligned}$$

where L is the Lipschitz constant with respect to x, K is the Lipschitz constant with respect to λ given in Remark 5, B is a positive constant.

Lemma 11: With Assumptions 1-4, for fixed common point λ^*, all the sequences $\{x_i(k)\}$, $i \in \mathcal{N}$ converge to the same random variable consensusly and almost surely by Algorithm 1.

Lemma 12: Set $\iota(k) = \frac{1}{k^{1+\epsilon}}$ and $c(k) = \frac{1}{k^{\delta}}$ with $\frac{1}{2} + \epsilon > \delta \geqslant \epsilon > 0$. With Assumptions 1-4, we have

$$\sum_{i=1}^{n} \mathbb{E}\left\|x_i(k) - x^*\right\|^2 \leqslant \frac{M_1}{k^{2+2\epsilon}} + \frac{M_2}{k^{\epsilon}} + \frac{M_3}{k^{\epsilon+\delta}},$$

where M_1, M_2 and M_3 are constants.

Proof of Theorem 1:

(a) We prove that for $i, j = 1, 2, \ldots, n$,

$$\lim_{k\to\infty} \left\|\lambda_i(k) - \lambda_j(k)\right\| = 0 \quad a.\ s.$$

According to the definition of transition matrix $\Psi(k, s)$ and the definition of $\lambda_i(k+1)$ in (9), we have:

$$\lambda_i(k+1) = \sum_{j=1}^{n} \left[\Psi(k, 0)\right]_{ij} \lambda_j(0). \tag{10}$$

Define $\bar{\lambda}(k+1) = \frac{1}{n}\sum_{i=1}^{n} \lambda_i(k+1)$. According to Assumption 1 and by an analoglous induction, the following equality holds:

$$\bar{\lambda}(k+1) = \frac{1}{n}\sum_{i=1}^{n} \lambda_i(0). \tag{11}$$

Therefore, $\forall i \in \mathcal{N}$,

$$\left\|\lambda_i(k+1) - \bar{\lambda}(k+1)\right\| \leqslant \sum_{j=1}^{n} \left|\left[\Psi(k, 0)\right]_{ij} - \frac{1}{n}\right| \left\|\lambda_j(0)\right\|. \tag{12}$$

Plugging in the estimate of $\Psi(k, s)$ in Lemma 7, we have

$$\left\|\lambda_i(k+1) - \bar{\lambda}(k+1)\right\| \leqslant n\delta\beta^k \max_{1\leqslant i\leqslant n} \left\|\lambda_i(0)\right\|. \tag{13}$$

Therefore,

$$\lim_{k\to\infty} \left\|\lambda_i(k) - \bar{\lambda}(k)\right\| = 0,\ \forall i \in \mathcal{N}. \tag{14}$$

(b) According to Lemma 11, $\lim_{k\to\infty} \sum_{i=1}^{n} \left\|x_i(k) - x^*\right\|^2$ converges to a non-negative random variable almost surely. According to Lemma 12, we have

$$\lim_{k\to\infty} \mathbb{E}\left\|x_i(k) - x^*\right\|^2 = 0,$$

which means that $\{x_i(k)\}$, $i \in \mathcal{N}$ generated from Algorithm 1 converge to the optimal solution x^* in L_2. Therefore,

$$\lim_{k\to\infty} \sum_{i=1}^{n} \left\|x_i(k) - x^*\right\|_2^2 = 0, \quad \text{a.s.} \tag{15}$$

■

Then we give the following convergence rate result of the proposed algorithm.

Theorem 2: Set $\iota(k) = \frac{1}{k^{1+\epsilon}}$ and $c(k) = \frac{1}{k^{\delta}}$ with $\frac{1}{2} + \epsilon > \delta \geqslant \epsilon > 0$. With Assumptions 1-4, for distributed (sub)gradient-free inteval-valued algorithm (Algorithm 1), we have

$$\sum_{i=1}^{n} \mathbb{E}\left\|x_i(k) - x^*\right\|^2 \sim O\left(\max\left\{\frac{1}{k^{\epsilon}}, \frac{1}{k^{1+2\epsilon-2\delta}}\right\}\right)$$

$$\sum_{i=1}^{n} \mathbb{E}\left\|\lambda_i(k) - \lambda^*\right\|^2 \sim O\left(\beta^k\right).$$

Proof: According to Lemma 12, we have

$$\sum_{i=1}^{n} \mathbb{E}\left\|x_i(k) - x^*\right\|^2 \sim O\left(\max\left\{\frac{1}{k^{\epsilon}}, \frac{1}{k^{1+2\epsilon-2\delta}}\right\}\right). \tag{16}$$

Following (13) in the proof of Theorem 1, we obtain

$$\sum_{i=1}^{n} \mathbb{E}\left\|\lambda_i(k) - \lambda^*\right\|^2 \sim O\left(\beta^k\right). \tag{17}$$

The proof is completed. ■

Remark 6: It directly follows from Theorem 2 that the optimal values for ϵ and δ are $\epsilon = \dfrac{1}{2}$ and $\delta = \dfrac{1}{2}$, respectively, which in turn indicate that $\iota_k = \frac{1}{k^{\frac{3}{2}}}$, $c_k = \frac{1}{k^{\frac{1}{2}}}$, and

$$\sum_{i=1}^{n} \mathbb{E}\left\|x_i(k) - x^*\right\|^2 \sim O\left(\frac{1}{\sqrt{k}}\right)$$

$$\sum_{i=1}^{n} \mathbb{E}\left\|\lambda_i(k) - \lambda^*\right\|_2^2 \sim O\left(\frac{1}{\sqrt{k}}\right).$$

Not only does the rate match the best rate for centralized stochastic approximation algorithms, see [36] and references therein, but it also matches the best rate given for first-order stochastic subgradient algorithms [37] with diminishing step-size..

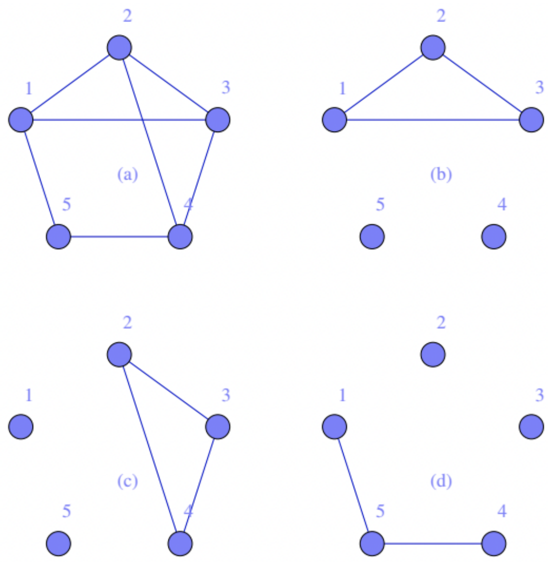

FIGURE 3. Topology of the 5-agent network.

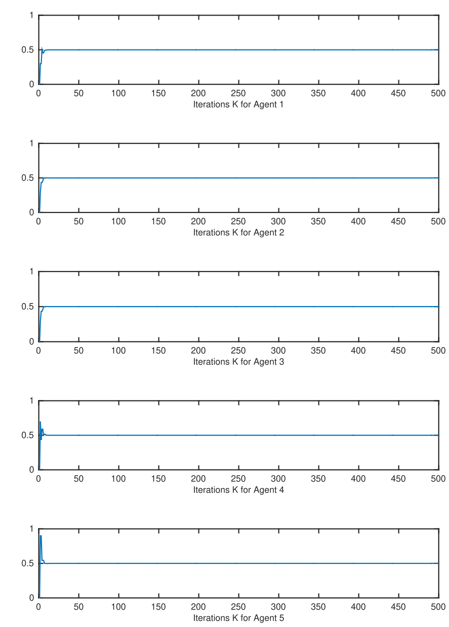

FIGURE 4. $\lambda_i(k)$ for agent i.

V. SIMULATION

Simulations of the distributed (sub)gradient-free inteval-valued algorithm are presented in this section. In particular, we consider the following distributed interval-valued quadratic problem:

$$
\begin{aligned}
\min \quad & G(x) = \sum_{i=1}^{5} [\upsilon_{1i}, \upsilon_{2i}] \|x - \rho_i\|^2 \\
s.t. \quad & \|x\| \leqslant X,
\end{aligned}
$$

where $\upsilon_{1i}, \upsilon_{1i} \in \mathcal{R}$ and $\rho_i \in \mathcal{R}^p$. This problem is motivited from centralized quadratic interval-valued learning [43] and distributed optimization [45].

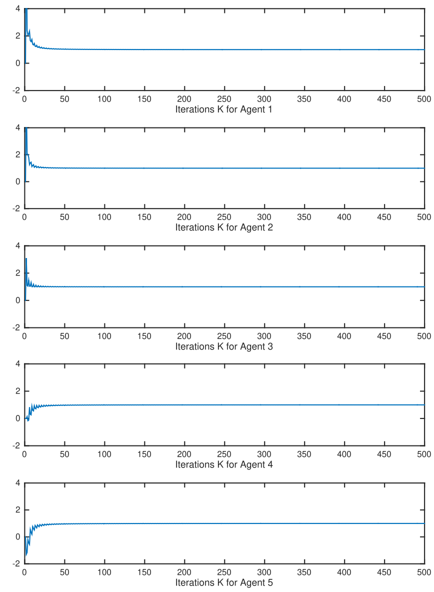

FIGURE 5. $x_i(k)$ for agent i.

We demonstrate the proposed algorithm with the expression $X := \{x \big| \|x\| \leqslant 100\}$, assuming $[\upsilon_{1i}, \upsilon_{2i}] = [0.5, 2]$, where $\rho_1 = 3$, $\rho_2 = 2$, $\rho_3 = 1$, $\rho_4 = 0$, $\rho_5 = -1$, respectively. In addition, we detail the parameter choices utilized in simulations of the proposed algorithm. Initially, we set the step size $\iota(k) = \frac{1}{k^{\frac{3}{2}}}$ and the parameter $c(k) = \frac{1}{k^{\frac{1}{2}}}$ for the random differences. $\lambda_1(0) = 0.1$, $\lambda_2(0) = 0.3$, $\lambda_3(0) = 0.5$, $\lambda_4(0) = 0.7$, $\lambda_5(0) = 0.9$, and $x_i(0)$'s $= 0$ are then respectively initialized.

Then, let us investigate the convergence performance of the distributed (sub)gradient-free interval-valued algorithm. The simulation results are based on a 5-agent time-varying network whose communication topology is depicted in Fig. 2. Both Fig.4 and Fig. 5 illustrate the convergence performance of the proposed algorithm. For 500 iterations, we can attain a (0.500, 0.996) pareto solution.

VI. CONCLUSION

This paper investigated the problem of distributed interval optimization subject to local convex constraints. The objective functions for the distributed design are compact interval-valued functions, and the network is time-varying. A distributed subgradient-free methodology for finding a Pareto-optimal solution to a distributed interval optimization problem was developed by constructing random differences. In addition, we showed that a Pareto optimal solution can be achieved with probability one over a time-varying network and provided a numerical illustration of the algorithm's efficacy.

APPENDIX A
PROOF OF LEMMA 8

According to the definition of $d_i(k)$ in (6), we have

$$
d_i(k) = \frac{\left[y_i^+(k) - y_i^-(k)\right] \Delta_i^-(k)}{2c(k)} \tag{18}
$$

where $\|y_i^+(k) - y_i^-(k)\| = \|f_i(\xi_i(k) + c(k)\triangle_i(k), \lambda_i(k)) - f_i(\xi_i(k) - c(k)\triangle_i(k), \lambda_i(k)\| \leqslant 2Lc(k)\|\triangle_i(k)\|$ according to Remark 5. Due to Assumption 3(a), we have

$$\mathbb{E}\left\|\frac{[y_i^+(k) - y_i^-(k)]\triangle_i^-(k)}{2c(k)}\right\| \leqslant L, \tag{19}$$

and

$$\mathbb{E}\left\|\frac{[y_i^+(k) - y_i^-(k)]\triangle_i^-(k)}{2c(k)}\right\|^2 \leqslant L^2. \tag{20}$$

APPENDIX B
PROOF OF LEMMA 9

For all $i \in \mathcal{N}$, $k \geqslant 0$, define

$$p_i(k+1) = x_i(k+1) - \xi_i(k) = x_i(k+1) - \sum_{j=1}^n w_{ij}(k)x_j(k). \tag{21}$$

as the error between $x_i(k+1)$ and $\xi_i(k)$. According to Lemma 5(b) and the fact that X is a closed, convex set, we get

$$\begin{aligned}&\|p_i(k+1)\|\\&= \left\|P_X\Big(\sum_{j=1}^n w_{ij}(k)x_j(k) - \iota(k)d_i(k)\Big) - \sum_{j=1}^n w_{ij}(k)x_j(k)\right\|\\&\leqslant \iota(k)\|d_i(k)\|.\end{aligned} \tag{22}$$

Rewrite (9) compactly in terms of $\Psi(k,s)$ and the definition of $p_i(k+1)$ as follows:

$$\begin{aligned}x_i(k+1) &= \sum_{j=1}^n [\Psi(k,0)]_{ij} x_j(0) + p_i(k+1)\\&\quad + \sum_{s=1}^k\sum_{j=1}^n [\Psi(k,s)]_{ij} p_j(s),\end{aligned} \tag{23}$$

for $k \geqslant s$. Define $\bar{x}(k+1) = \frac{1}{n}\sum_{i=1}^n x_i(k+1)$. Moreover, with Assumption 1(b) and by an analoglous induction, the following equality holds:

$$\bar{x}(k+1) = \frac{1}{n}\sum_{i=1}^n x_i(0) + \frac{1}{n}\sum_{s=1}^{k+1}\sum_{j=1}^n p_j(s) \tag{24}$$

Therefore, $\forall i \in \mathcal{N}$,

$$\begin{aligned}&\|x_i(k+1) - \bar{x}(k+1)\|\\&\leqslant \sum_{j=1}^n \left|[\Psi(k,0)]_{ij} - \frac{1}{n}\right|\|x_j(0)\|\\&\quad + \|p_i(k+1)\| + \frac{1}{n}\sum_{j=1}^n\|p_j(k+1)\|\\&\quad + \sum_{s=1}^k\sum_{j=1}^n\left|[\Psi(k,s)]_{ij} - \frac{1}{n}\right|\|p_j(s)\|.\end{aligned} \tag{25}$$

Taking the expectation of (25), we get

$$\mathbb{E}\|x_i(k+1) - \bar{x}(k+1)\| \tag{26}$$

$$\begin{aligned}&\leqslant \sum_{j=1}^n\left|[\Psi(k,0)]_{ij} - \frac{1}{n}\right|\|x_j(0)\| + \mathbb{E}\|p_i(k+1)\|\\&\quad + \frac{1}{n}\sum_{j=1}^n\mathbb{E}\|p_j(k+1)\| + \sum_{s=1}^k\sum_{j=1}^n\left|[\Psi(k,s)]_{ij}\right.\\&\quad \left. - \frac{1}{n}\right|\mathbb{E}\|p_j(s)\|.\end{aligned} \tag{27}$$

Plugging in the estimate of $\Psi(k,s)$ in Lemma 7 and the estimate of $p_i(k+1)$ in (22), we have

$$\begin{aligned}&\mathbb{E}\|x_i(k+1) - \bar{x}(k+1)\|\\&\leqslant n\delta\beta^k\max_{1\leqslant i\leqslant n}\|x_i(0)\| + \iota(k)\mathbb{E}\|d_i(k)\| + \frac{\iota(k)}{n}\sum_{i=1}^n\mathbb{E}\|d_i(k)\|\\&\quad + \delta\sum_{s=1}^k\beta^{k-s}\sum_{i=1}^n\iota(s-1)\mathbb{E}\|d_i(s-1)\|.\end{aligned} \tag{28}$$

From Theorem 1, we have $\mathbb{E}\|d_i(k)\| \leqslant L$. Therefore,

$$\begin{aligned}\mathbb{E}\|x_i(k+1) - \bar{x}(k+1)\| &\leqslant n\delta\beta^k\max_{1\leqslant i\leqslant n}\|x_i(0)\| + 2\iota(k)L\\&\quad + \delta n\sum_{s=1}^k\iota(s-1)\beta^{k-s}L.\end{aligned} \tag{29}$$

Since $\sum_{k=1}^\infty\iota(k)^2 < \infty$ with Assumption 4(a) and $\sum_{k=1}^\infty\frac{\iota(k)}{c(k)} < \infty$ with Assumption 4(c), we obtain $\lim_{k\to\infty}\iota(k) = 0$ and $\lim_{k\to\infty}\frac{\iota(k)}{c(k)} = 0$. According to Lemma 3.1 in [4], we obtain $\lim_{k\to\infty}\sum_{s=1}^k\iota(s-1)\beta^{k-s} = 0$ and $\lim_{k\to\infty}\sum_{s=1}^k\frac{\iota(s-1)}{c(s-1)}\beta^{k-s} = 0$. Therefore,

$$\lim_{k\to\infty}\mathbb{E}\|x_i(k+1) - \bar{x}(k+1)\| = 0,\ \forall i \in \mathcal{N}. \tag{30}$$

APPENDIX C
PROOF OF LEMMA 10

(a) Define $[C_i(k)]_1 = \xi_i(k) + c(k)\triangle_i(k)$ and $[C_i(k)]_2 = \xi_i(k) - c(k)\triangle_i(k)$. According to Lemma 1:

$$\begin{aligned}&f_i([C_i(k)]_1, \lambda_i(k)) - f_i([C_i(k)]_2, \lambda_i(k))\\&\in\Big\langle\partial f_{i_{\xi_i(k)+\theta_i c(k)\triangle_i(k)}}\big(\xi_i(k) + \theta_i c(k)\triangle_i(k), \lambda_i(k)\big),\\&\quad 2c(k)\triangle_i(k)\Big\rangle,\end{aligned} \tag{31}$$

where $\theta_i \in [-1, 1]$ is a constant. Therefore, there exists $\varsigma_i \in \partial f_{i_{\xi_i(k)+\theta_i c(k)\triangle_i(k)}}\big(\xi_i(k) + \theta_i c(k)\triangle_i(k), \lambda_i(k)\big)$ such that

$$\begin{aligned}&f_i([C_i(k)]_1, \lambda_i(k)) - f_i([C_i(k)]_2, \lambda_i(k))\\&\quad = \langle\varsigma_i,\ 2c(k)\triangle_i(k)\rangle.\end{aligned} \tag{32}$$

By taking conditional expectation of $\langle d_i(k), \xi_i(k) - x^*\rangle$ with respect to $F(k)$ and noticing (32), we obtain the following inequality:

$$\mathbb{E}\big[\langle d_i(k), \xi_i(k) - x^*\rangle \big| F(k)\big] = D_i(k), \quad (33)$$

where

$$D_i(k) = \mathbb{E}\Big[(\varsigma_i)^\top \bigwedge_i(k)\big[\bigwedge_i(k)\big]^{-\top}(\xi_i(k) - x^*)\big|F(k)\Big].$$

We further formulate $D_i(k)$ as follows:

$$D_i(k) = \mathbb{E}\Big[(\varsigma_i)^\top \Big(\bigwedge_i(k)\big[\bigwedge_i(k)\big]^{-\top} - I\Big)(\xi_i(k) - x^*) \times \big|F(k)\Big] + \mathbb{E}\Big[\langle \varsigma_i,\ \xi_i(k) - x^*\rangle\big|F(k)\Big]. \quad (34)$$

By Definition 1 and Remark 5, we obtain

$$\begin{aligned}
&\mathbb{E}[\langle \varsigma_i,\ \xi_i(k) - x^*\rangle|F(k)] \\
&= \mathbb{E}[\langle \varsigma_i, \xi_i(k)+\theta_i c(k)\bigwedge_i(k)-\theta_i c(k)\bigwedge_i(k)-x^*\rangle|F(k)] \\
&\geqslant \mathbb{E}[f_i\big(\xi_i(k)+\theta_i c(k)\bigwedge(k)_i, \lambda_i(k)\big)-f_i\big(x^*, \lambda_i(k)\big)|F(k)] \\
&\quad - \big|c(k)\big|L\mathbb{E}\big\|\theta_i \bigwedge_i(k)\big\| \\
&\geqslant \mathbb{E}[f_i\big(\xi_i(k)+\theta_i c(k)\bigwedge_i(k), \lambda_i(k)\big)-f_i\big(\bar{x}(k), \lambda_i(k)\big)|F(k)] \\
&\quad + f_i\big(\bar{x}(k), \lambda_i(k)\big) - f_i\big(x^*, \lambda_i(k)\big) - \big|c(k)\big|L\mathbb{E}\big\|\theta_i \bigwedge_i(k)\big\| \\
&\geqslant f_i\big(\bar{x}(k), \bar{\lambda}(k)\big) - f_i\big(x^*, \lambda^*\big) + f_i\big(\bar{x}(k), \lambda_i(k)\big) \\
&\quad - f_i\big(\bar{x}(k), \bar{\lambda}(k)\big) - f_i\big(x^*, \lambda_i(k)\big) + f_i\big(x^*, \lambda^*\big) \\
&\quad - L\big\|\xi_i(k) - \bar{x}(k)\big\| - 2\big|c(k)\big|L\mathbb{E}\big\|\theta_i \bigwedge_i(k)\big\| \\
&\geqslant f_i\big(\bar{x}(k), \bar{\lambda}(k)\big) - f_i\big(x^*, \lambda^*\big) - L\big\|\xi_i(k) - \bar{x}(k)\big\| \\
&\quad - K\big\|\lambda_i(k)-\bar{\lambda}(k)\big\| - K\|\lambda_i(k)-\lambda^*\| - 2c(k)L\mathbb{E}\big\|\bigwedge_i(k)\big\|,
\end{aligned} \quad (35)$$

and

$$\begin{aligned}
&\Big|\mathbb{E}\Big[(\varsigma_i)^\top\Big(\bigwedge_i(k)\big[\bigwedge_i(k)\big]^{-\top} - I\Big)(\xi_i(k) - x^*)\big|F(k)\Big]\Big| \\
&= \Big|\mathbb{E}\Big[(\varsigma_i - \partial f_{i_{x_i(k)}}(x_i(k), \lambda_i(k)))^\top \\
&\quad \times \Big(\bigwedge_i(k)\big[\bigwedge_i(k)\big]^{-\top} - I\Big)(\xi_i(k) - x^*)\big|F(k)\Big]\Big| \leqslant B.
\end{aligned} \quad (36)$$

where B is a positive constant. Combining (35), (28) with (33) gives

$$\begin{aligned}
&\mathbb{E}\big[\langle d_i(k),\ x_i(k) - \xi^*\rangle\big|F(k)\big] \\
&\geqslant f_i\big(\bar{x}(k), \bar{\lambda}(k)\big) - f_i\big(x^*, \lambda^*\big) - L\big\|\xi_i(k) - \bar{x}(k)\big\| - B \\
&- K\big\|\lambda_i(k)-\bar{\lambda}(k)\big\| - K\|\lambda_i(k)-\lambda^*\| - c(k)L\big\|\bigwedge_i(k)\big\|.
\end{aligned} \quad (37)$$

(b) By Definition 1 and Remark 5, we obtain

$$\begin{aligned}
&\mathbb{E}[\langle \varsigma_i,\ \xi_i(k) - x^*\rangle|F(k)] \\
&= \mathbb{E}[\langle \varsigma_i, \xi_i(k)+\theta_i c(k)\bigwedge_i(k)-\theta_i c(k)\bigwedge_i(k)-x^*\rangle|F(k)] \\
&\geqslant \mathbb{E}[f_i\big(\xi_i(k)+\theta_i c(k)\bigwedge(k)_i, \lambda_i(k)\big)-f_i\big(x^*, \lambda_i(k)\big)|F(k)] \\
&\quad - \big|c(k)\big|L\mathbb{E}\big\|\theta_i \bigwedge_i(k)\big\| \\
&\geqslant \mathbb{E}[f_i\big(\xi_i(k) + \theta_i c(k)\bigwedge_i(k), \lambda_i(k)\big) \\
&\quad - f_i\big(\bar{x}(k), \lambda_i(k)\big)|F(k)] \\
&\quad + f_i\big(\bar{x}(k), \lambda_i(k)\big) - f_i\big(x^*, \lambda_i(k)\big) - \big|c(k)\big|L\mathbb{E}\big\|\theta_i \bigwedge_i(k)\big\| \\
&\geqslant f_i\big(\bar{x}(k), \lambda^*\big) - f_i\big(x^*, \lambda^*\big) + f_i\big(\bar{x}(k), \lambda_i(k)\big) \\
&\quad - f_i\big(\bar{x}(k), \lambda^*\big) - f_i\big(x^*, \lambda_i(k)\big) + f_i\big(x^*, \lambda^*\big) \\
&- L\big\|\xi_i(k) - \bar{x}(k)\big\| - 2\big|c(k)\big|L\mathbb{E}\big\|\theta_i \bigwedge_i(k)\big\| \\
&\geqslant f_i\big(\bar{x}(k), \lambda^*\big) - f_i\big(x^*, \lambda^*\big) - L\big\|\xi_i(k) - \bar{x}(k)\big\| \\
&- 2K\|\lambda_i(k) - \lambda^*\| - 2c(k)L\mathbb{E}\big\|\bigwedge_i(k)\big\|.
\end{aligned} \quad (38)$$

Similar to the proof of part (a), we can get

$$\begin{aligned}
&\mathbb{E}\big[\langle d_i(k),\ x_i(k) - \xi^*\rangle\big|F(k)\big] \\
&\geqslant f_i\big(\bar{x}(k), \lambda^*\big) - f_i\big(x^*, \lambda^*\big) - L\big\|\xi_i(k) - \bar{x}(k)\big\| - 2L \\
&- 2K\|\lambda_i(k) - \lambda^*\| - c(k)L\big\|\bigwedge_i(k)\big\|.
\end{aligned} \quad (39)$$

The proof of second part of Lemma 10 can be given by taking expection to both side of (39).

APPENDIX D
PROOF OF LEMMA 11

(a) We prove that for $i, j = 1, 2, \ldots, n$,

$$\lim_{k\to\infty} \big\|x_i(k) - x_j(k)\big\| = 0 \quad a.\ s.$$

From Lemma 9, $\lim_{k\to\infty} \mathbb{E}\big\|x_i(k+1) - \bar{x}(k+1)\big\| = 0$ holds. Still

$$\begin{aligned}
0 &\leqslant \mathbb{E}\Big[\liminf_{k\to\infty}\big\|x_i(k+1) - \bar{x}(k+1)\big\|\Big] \\
&\leqslant \liminf_{k\to\infty}\mathbb{E}\big\|x_i(k+1) - \bar{x}(k+1)\big\| = 0,
\end{aligned} \quad (40)$$

which yields $\mathbb{E}\Big[\liminf_{k\to\infty}\big\|x_i(k+1) - \bar{x}(k+1)\big\|\Big] = 0$. Therefore, $\liminf_{k\to\infty}\big\|x_i(k+1) - \bar{x}(k+1)\big\| = 0$ holds almost surely. Since $\sum_{i=1}^n \big\|x_i(k+1) - \bar{x}(k+1)\big\|^2 \leqslant \sum_{i=1}^n \big\|x_i(k+1) - \bar{x}(k)\big\|^2$ according to Lemma 2 and $\big\|x_i(k+1) - \bar{x}(k)\big\|^2 \leqslant \big\|\hat{\xi}_i(k) - \bar{x}(k)\big\|^2$ according to Lemma 3,

$$\sum_{i=1}^n \big\|x_i(k+1) - \bar{x}(k+1)\big\|^2$$

$$\leqslant \sum_{i=1}^{n} \left\|\hat{\xi}_i(k) - \bar{x}(k)\right\|^2$$
$$\leqslant \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(k)\left\|x_j(k) - \bar{x}(k)\right\|^2 + \iota^2(k)\sum_{i=1}^{n}\left\|d_i(k)\right\|^2$$
$$+ 2\iota(k)\sum_{i=1}^{n}\left\|d_i(k)\right\|\sum_{j=1}^{n} w_{ij}(k)\left\|x_j(k) - \bar{x}(k)\right\|. \quad (41)$$

According to Assumption 2(b),

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(k)\left\|x_j(k) - \bar{x}(k)\right\|^2 = \sum_{i=1}^{n}\left\|x_i(k) - \bar{x}(k)\right\|^2. \quad (42)$$

Thus, taking the conditional expectation of both side of (41) yields

$$\sum_{i=1}^{n}\mathbb{E}\Big[\left\|x_i(k+1) - \bar{x}(k+1)\right\|^2\Big|F(k)\Big]$$
$$\leqslant \sum_{i=1}^{n}\left\|x_j(k) - \bar{x}(k)\right\|^2 + \sum_{i=1}^{n}\iota^2(k)\mathbb{E}\left\|d_i(k)\right\|^2$$
$$+ \sum_{i=1}^{n} 2n\iota(k)\mathbb{E}\left\|d_i(k)\right\|\mathbb{E}\left\|x_i(k) - \bar{x}(k)\right\|. \quad (43)$$

According to Assumption 4 and Lemma 8(b)

$$\sum_{k=1}^{\infty}\sum_{i=1}^{n}\iota^2(k)\mathbb{E}\left\|d_i(k)\right\|^2 < \infty.$$

According to Theorem 6.2 of [4], $\sum_{k=1}^{\infty}\iota(k)\left\|x_i(k) - \bar{x}(k)\right\| < \infty$ with probability 1. Through Lemma 8(a), $\sum_{k=1}^{\infty}\sum_{i=1}^{n} 2n\iota(k)\mathbb{E}\left\|d_i(k)\right\|\mathbb{E}\left\|x_i(k) - \bar{x}(k)\right\| < \infty$. Therefore, $\lim_{k\to\infty}\left\|x_i(k) - \bar{x}(k)\right\| = 0$ holds almost surely by Lemma 4.

(b) Clearly, $\left\|x_i(k+1) - x^*\right\|^2 \leqslant \left\|\hat{\xi}_i(k) - x^*\right\|^2$ according to the properties of Euclidean norm in Lemma 3. Then,

$$\left\|x_i(k+1) - x^*\right\|^2 \leqslant \left\|\xi_i(k) - x^*\right\|^2 + \iota^2(k)\left\|d_i(k)\right\|^2 - 2\iota(k)\big\langle d_i(k),\ \xi_i(k) - x^*\big\rangle. \quad (44)$$

Taking conditional expection of both sides of (44), we obtain for all $k = 0, 1, 2, \ldots$,

$$\mathbb{E}\Big[\left\|x_i(k+1) - x^*\right\|^2\Big|F(k)\Big]$$
$$\leqslant\mathbb{E}\Big[\left\|\xi_i(k) - x^*\right\|^2\Big|F(k)\Big] + \iota^2(k)\mathbb{E}\Big[\left\|d_i(k)\right\|^2\Big|F(k)\Big]$$
$$- 2\iota(k)\mathbb{E}\Big[\big\langle d_i(k),\ \xi_i(k) - x^*\big\rangle\Big|F(k)\Big]. \quad (45)$$

By the double stochasticity of matrix $W(k)$ in Assumption 2(b),

$$\sum_{i=1}^{n}\mathbb{E}\Big[\left\|\xi_i(k) - x^*\right\|^2\Big|F(k)\Big] \leqslant \sum_{i=1}^{n}\left\|x_i(k) - x^*\right\|^2,$$
$$\sum_{i=1}^{n}\mathbb{E}\Big[\left\|\xi_i(k) - \bar{x}(k)\right\|\Big|F(k)\Big] \leqslant \sum_{i=1}^{n}\left\|x_i(k) - \bar{x}(k)\right\|. \quad (46)$$

Then, with probability 1, for $i \in \mathcal{N}$, it holds

$$\sum_{i=1}^{n}\mathbb{E}\Big[\left\|x_i(k+1) - x^*\right\|^2\Big|F(k)\Big]$$
$$\leqslant \sum_{i=1}^{n}\Big[\left\|x_i(k) - x^*\right\|^2 + \big[O_i(k)\big]_1 + \big[O_i(k)\big]_2$$
$$+ \big[O_i(k)\big]_3 + \big[O_i(k)\big]_4 + \big[O_i(k)\big]_5 + \big[O_i(k)\big]_6 - J_i(k)\Big], \quad (47)$$

where

$$\begin{cases} \big[O_i(k)\big]_1 & = \iota^2(k)\mathbb{E}\Big[\left\|d_i(k)\right\|^2\Big|F(k)\Big] \\ \big[O_i(k)\big]_2 & = 2\iota(k)L\mathbb{E}\left\|x_i(k) - \bar{x}(k)\right\| \\ \big[O_i(k)\big]_3 & = 4\iota(k)c(k)L\mathbb{E}\left\|\bigtriangleup(k)_i\right\| \\ \big[O_i(k)\big]_4 & = 2\iota(k)L\mathbb{E}\left\|\lambda_i(k) - \bar{\lambda}(k)\right\| \\ \big[O_i(k)\big]_5 & = 2\iota(k)L\mathbb{E}\left\|\lambda_i(k) - \lambda^*\right\| \\ \big[O_i(k)\big]_6 & = 2\iota(k)B \\ J_i(k) & = 2\iota(k)\big[f_i\big(\bar{x}(k), \bar{\lambda}(k)\big) - f_i\big(x^*, \lambda^*\big)\big]. \end{cases}$$

According to Assumption 4 and Lemma 8,

$$\sum_{k=1}^{\infty}\big[O_i(k)\big]_1 < \infty.$$

By the proof in part (a), $\sum_{k=1}^{\infty}\big[O_i(k)\big]_2 < \infty$. By Assumption 3-4, $\sum_{k=1}^{\infty}\big[O_i(k)\big]_3 < \infty$. By Theorem 1, $\sum_{k=1}^{\infty}\big[O_i(k)\big]_4 < \infty$ and $\sum_{k=1}^{\infty}\big[O_i(k)\big]_5 < \infty$. By Assumption 4, $\sum_{k=1}^{\infty}\big[O_i(k)\big]_6 < \infty$. Therefore, $\sum_{k=1}^{\infty}\sum_{i=1}^{n}\Big[\big[O_i(k)\big]_1 + \big[O_i(k)\big]_2 + \big[O_i(k)\big]_3 + \big[O_i(k)\big]_4 + \big[O_i(k)\big]_5 + \big[O_i(k)\big]_6\Big] < \infty$. From Lemma 7, the sequence $\sum_{i=1}^{n}\left\|x_i(k) - x^*\right\|^2$ converges almost surely with $\sum_{k=1}^{\infty}\sum_{i=1}^{n} J_i(k) < \infty$. Therefore, the sequence $\sum_{i=1}^{n}\left\|\xi_i(k) - \xi^*\right\|^2$ converges to a random variable with probability 1. The proof is completed.

APPENDIX E
PROOF OF LEMMA 12

By taking expectation to both sides of (44), we obtain

$$\mathbb{E}\left\|x_i(k+1) - x^*\right\|^2 \leqslant \mathbb{E}\left\|\xi_i(k) - x^*\right\|^2 + \iota^2(k)\mathbb{E}\left\|d_i(k)\right\|^2 - 2\iota(k)\mathbb{E}\big[\big\langle d_i(k),\ \xi_i(k) - x^*\big\rangle\big]. \quad (48)$$

By the double stochasticity of matrix $W(k)$ given in Assumption 2(b), we have the following inequalities

$$\sum_{i=1}^{n}\mathbb{E}\left\|\xi_i(k) - x^*\right\|^2 = \sum_{i=1}^{n}\mathbb{E}\Big\|\sum_{j=1}^{n} w_{ij}(k)x_j(k) - x^*\Big\|^2$$
$$\leqslant \sum_{i=1}^{n}\mathbb{E}\left\|x_i(k) - x^*\right\|^2, \quad (49)$$

$$\sum_{i=1}^{n}\mathbb{E}\big\|\xi_i(k)-\bar{x}(k)\big\| = \sum_{i=1}^{n}\mathbb{E}\Big\|\sum_{j=1}^{n}w_{ij}(k)x_j(k)-\bar{x}(k)\Big\| \leqslant \sum_{i=1}^{n}\mathbb{E}\big\|x_i(k)-\bar{x}(k)\big\|. \quad (50)$$

By taking summation of both sides of (48) for $k = 1, 2, \ldots T$ and $i = 1, 2, \ldots n$ and noticing (49), (50) and Lemma 10, we have

$$\begin{aligned}
&\sum_{s=1}^{k}\sum_{i=1}^{n}\mathbb{E}\big\|x_i(s+1)-x^*\big\|^2 \\
&\leqslant \sum_{s=1}^{k}\sum_{i=1}^{n}\mathbb{E}\big\|x_i(s)-x^*\big\|^2 + 4K\sum_{s=1}^{k}\sum_{i=1}^{n}\iota(s)\mathbb{E}\big\|\lambda_i(s)-\lambda^*\big\| \\
&\quad + 2L\sum_{s=1}^{k}\sum_{i=1}^{n}\iota(s)\mathbb{E}\big\|x_i(s)-\bar{x}(s)\big\| + 2nB\sum_{s=1}^{k}\iota(s) \\
&\quad + 4L\sum_{s=1}^{k}\sum_{i=1}^{n}\iota(s)c(s)\mathbb{E}\Big\|\bigtriangleup_i(s)\Big\| + \sum_{s=1}^{k}\sum_{i=1}^{n}\iota^2(k)\mathbb{E}\big\|d_i(s)\big\|^2 \\
&\quad - 2\sum_{s=1}^{k}\sum_{i=1}^{n}\iota(s)\mathbb{E}\Big[f_i\big(\bar{x}(s),\lambda^*\big)-f_i\big(x^*,\lambda^*\big)\Big]. \quad (51)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{i=1}^{n}\mathbb{E}\big\|x_i(k+1)-x^*\big\|^2 \\
&\leqslant \sum_{s=1}^{k}\sum_{i=1}^{n}\iota^2(k)\mathbb{E}\big\|d_i(s)\big\|^2 + 4K\sum_{s=1}^{k}\sum_{i=1}^{n}\iota(s)\mathbb{E}\big\|\lambda_i(s)-\lambda^*\big\| \\
&\quad + 2L\sum_{s=1}^{k}\sum_{i=1}^{n}\iota(s)\mathbb{E}\big\|x_i(s)-\bar{x}(s)\big\| + 2nB\sum_{s=1}^{k}\iota(s) \\
&\quad + 4L\sum_{s=1}^{k}\sum_{i=1}^{n}\iota(s)c(s)\mathbb{E}\Big\|\bigtriangleup_i(s)\Big\|. \quad (52)
\end{aligned}$$

Noticing that $\iota(s) = \frac{1}{s^{1+\epsilon}}$, $c(s) = \frac{1}{s^{\delta}}$, and $\frac{1}{2}+\epsilon > \delta > 0$. By Lemma 8, for the first term on the right hand side of (52), we have

$$\sum_{s=1}^{k}\sum_{i=1}^{n}\iota^2(s)\mathbb{E}\big\|d_i(s)\big\|^2 \leqslant n\sum_{s=1}^{k}\iota^2(k)L^2 \leqslant \frac{M_1}{k^{2+2\epsilon}}. \quad (53)$$

Since X is bounded in $\mathcal{R}^m$, for $x \in X$, there exists a constant M_x such that $\|x\| \leqslant M_x$. Still, $\lambda \in [0, 1]$. For the terms on the right hand side of (52), we have

$$2L\sum_{s=1}^{k}\sum_{i=1}^{n}\iota(s)\mathbb{E}\big\|x_i(s)-\bar{x}(s)\big\| \leqslant 4nLM_x\sum_{s=1}^{k}\iota(s) \leqslant \frac{M_{21}}{k^{\epsilon}}, \quad (54)$$

$$4K\sum_{s=1}^{k}\sum_{i=1}^{n}\iota(s)\mathbb{E}\big\|\lambda_i(s)-\lambda^*\big\| \leqslant 2nK\sum_{s=1}^{k}\iota(s) \leqslant \frac{M_{22}}{k^{\epsilon}}. \quad (55)$$

According to Assumption 3-4, for the fouth term on the right hand side of (52), we have $2nB\sum_{s=1}^{k}\iota(s) \leqslant \frac{M_{23}}{k^{\epsilon}}$. For the last term on the right hand side of (52), we have

$$4L\sum_{s=1}^{k}\sum_{i=1}^{n}\iota(s)c(s)\mathbb{E}\Big\|\bigtriangleup_i(s)\Big\| \leqslant \frac{M_3}{k^{\epsilon+\delta}}. \quad (56)$$

Since $M_1, M_{21}, M_{22}, M_{23}, M_3$ are positive constants in the above inequalities, we have

$$\sum_{i=1}^{n}\mathbb{E}\big\|x_i(k)-x^*\big\|^2 \leqslant \frac{M_1}{k^{2+2\epsilon}} + \frac{M_2}{k^{\epsilon}} + \frac{M_3}{k^{\epsilon+\delta}}, \quad (57)$$

where $M_{21}+M_{22}+M_{23} = M_2$.

REFERENCES

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

[2] P. Yi, Y. Hong, and F. Liu, "Distributed gradient algorithm for constrained optimization with application to load sharing in power systems," *Syst. Control Lett.*, vol. 83, pp. 45–52, Sep. 2015.

[3] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Autom. Control*, vol. 55, no. 4, pp. 922–938, Apr. 2010.

[4] S. S. Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *J. Optim. Theory Appl.*, vol. 147, no. 3, pp. 516–545, Dec. 2010.

[5] Y. Zhang, Y. Lou, Y. Hong, and L. Xie, "Distributed projection-based algorithms for source localization in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3131–3142, Jun. 2015.

[6] A. Cherukuri and J. Cortés, "Distributed generator coordination for initialization and anytime optimization in economic dispatch," *IEEE Trans. Control Netw. Syst.*, vol. 2, no. 3, pp. 226–237, Sep. 2015.

[7] G. Shi, B. D. O. Anderson, and U. Helmke, "Network flows that solve linear equations," *IEEE Trans. Autom. Control*, vol. 62, no. 6, pp. 2659–2674, Jun. 2017.

[8] X. Zeng, S. Liang, Y. Hong, and J. Chen, "Distributed computation of linear matrix equations: An optimization perspective," *IEEE Trans. Autom. Control*, vol. 64, no. 5, pp. 1858–1873, May 2019.

[9] Y. Wang, P. Lin, and Y. Hong, "Distributed regression estimation with incomplete data in multi-agent networks," *Sci. China Inf. Sci.*, vol. 61, no. 9, Sep. 2018, Art. no. 092202.

[10] X. Zeng, P. Yi, and Y. Hong, "Distributed algorithm for robust resource allocation with polyhedral uncertain allocation parameters," *J. Syst. Sci. Complex.*, vol. 31, no. 1, pp. 103–119, Feb. 2018.

[11] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1617–1626, May 2013.

[12] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for Machine Learning*. Cambridge, MA, USA: MIT Press, 2012.

[13] A. Neumaier, *Interval Methods for Systems of Equations*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[14] J. Rohn, "Positive definiteness and stability of interval matrices," *SIAM J. Matrix Anal. Appl.*, vol. 15, no. 1, pp. 175–184, Jan. 1994.

[15] V. I. Levin, "Nonlinear optimization under interval uncertainty," *Cybern. Syst. Anal.*, vol. 35, no. 2, pp. 297–306, Mar. 1999.

[16] B. Qing Hu and S. Wang, "A novel approach in uncertain programming—Part I: New arithmetic and order relation for interval numbers," *J. Ind. Manage. Optim.*, vol. 2, no. 4, pp. 351–371, 2006.

[17] L. Wu, M. Shahidehpour, and Z. Li, "Comparison of scenario-based and interval optimization approaches to stochastic SCUC," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 913–921, May 2012.

[18] H. Ishibuchi and H. Tanaka, "Multiobjective programming in optimization of the interval objective function," *Eur. J. Oper. Res.*, vol. 48, no. 2, pp. 219–225, Sep. 1990.

[19] H.-C. Wu, "On interval-valued nonlinear programming problems," *J. Math. Anal. Appl.*, vol. 338, no. 1, pp. 299–316, Feb. 2008.

[20] S.-T. Liu and R.-T. Wang, "A numerical solution to interval quadratic programming exists, however it is generally not used due to the lack of efficiency," *Appl. Math. Comput.*, vol. 189, no. 2, pp. 1274–1281, Jun. 2007.
[21] C. Jiang, X. Han, G. R. Liu, and G. P. Liu, "A nonlinear interval number programming method for uncertain optimization problems," *Eur. J. Oper. Res.*, vol. 188, no. 1, pp. 1–13, Jul. 2008.
[22] A. Jayswal, I. Stancu-Minasian, and I. Ahmad, "On sufficiency and duality for a class of interval-valued programming problems," *Appl. Math. Comput.*, vol. 218, no. 8, pp. 4119–4127, Dec. 2011.
[23] M. Hladık, "Interval linear programming: A survey," in *Linear Programming: New Frontiers in Theory and Applications*. Hauppauge, NY, USA: Nova Science, 2012, pp. 85–120.
[24] A. Bellet, Y. Liang, A. B. Garakani, M.-F. Balcan, and F. Sha, "A distributed Frank–Wolfe algorithm for communication-efficient sparse learning," in *Proc. SIAM Int. Conf. Data Mining*. Philadelphia, PA, USA: SIAM, Jun. 2015, pp. 478–486.
[25] R. Tempo, G. Calafiore, and F. Dabbene, *Randomized Algorithms for Analysis and Control of Uncertain Systems: With Applications*. London, U.K.: Springer, 2012.
[26] H. Ishii and R. Tempo, "Distributed randomized algorithms for the PageRank computation," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 1987–2002, Sep. 2010.
[27] H. F. Chen, T. E. Duncan, and B. Pasik-Duncan, "A Kiefer–Wolfowitz algorithm with randomized differences," *IEEE Trans. Autom. Control*, vol. 44, no. 3, pp. 442–453, Mar. 1999.
[28] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to Derivative-Free Optimization*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2009.
[29] J. C. Duchi, M. I. Jordan, M. J. Wainwright, and A. Wibisono, "Optimal rates for zero-order convex optimization: The power of two function evaluations," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2788–2806, May 2015.
[30] X. Gao, B. Jiang, and S. Zhang, "On the information-adaptive variants of the ADMM: An iteration complexity perspective," *J. Sci. Comput.*, vol. 76, no. 1, pp. 327–363, Jul. 2018.
[31] Y. Nesterov and V. Spokoiny, "Random gradient-free minimization of convex functions," *Found. Comput. Math.*, vol. 17, no. 1, pp. 527–566, 2017.
[32] A. K. Sahu, D. Jakovetic, D. Bajovic, and S. Kar, "Distributed zeroth order optimization over random networks: A Kiefer–Wolfowitz stochastic approximation approach," in *Proc. IEEE Conf. Decis. Control (CDC)*, Dec. 2018, pp. 4951–4958.
[33] D. Hajinezhad, M. Hong, and A. Garcia, "Zone: Zeroth-order nonconvex multiagent optimization over networks," *IEEE Trans. Autom. Control*, vol. 64, no. 10, pp. 3995–4010, Oct. 2019.
[34] D. Yuan and D. W. C. Ho, "Randomized gradient-free method for multiagent optimization over time-varying networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1342–1347, Jun. 2015.
[35] D. Yuan, D. W. C. Ho, and S. Xu, "Zeroth-order method for distributed optimization with approximate projections," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 2, pp. 284–294, Feb. 2016.
[36] H.-F. Chen, *Stochastic Approximation and Its Applications*. New York, NY, USA: Springer, 2006.
[37] E. Hazan, "Introduction to online convex optimization," *Found. Trends Optim.*, vol. 2, nos. 3–4, pp. 157–325, 2016.
[38] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of Convex Analysis*. Berlin, Germany: Springer, 2012.
[39] F. H. Clarke, R. J. Stern, Y. S. Ledyaev, and R. R. Wolenski, *Nonsmooth Analysis and Control Theory* (Graduate Texts in Mathematics), vol. 178. New York, NY, USA: Springer, 1998, pp. 137–151.
[40] R. Durrett, *Probability: Theory and Examples*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[41] B. T. Polyak, *Introduction to Optimization*. London, U.K.: Chapman & Hall, 1987.
[42] J.-P. Aubin and A. Cellina, *Differential Inclusions Set-Valued Maps and Viability Theory*. Berlin, Germany: Springer, 2012.
[43] A. K. Bhurjee and G. Panda, "Efficient solution of interval optimization problem," *Math. Methods Oper. Res.*, vol. 76, no. 3, pp. 273–288, Dec. 2012.
[44] T. Maeda, "On optimization problems with set-valued objective maps: Existence and optimality," *J. Optim. Theory Appl.*, vol. 153, no. 2, pp. 263–279, May 2012.
[45] A. Nedic and A. Olshevsky, "Stochastic gradient-push for strongly convex functions on time-varying directed graphs," *IEEE Trans. Autom. Control*, vol. 61, no. 12, pp. 3936–3947, Dec. 2016.

YINGHUI WANG received the B.S. degree in mathematics and applied mathematics from Shandong University, in 2010, and the Ph.D. degree from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, in 2019. She is currently a Lecturer with the 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, China. Her current research interests include distributed optimization, distributed control, and distributed computation of network systems.

JIUWEI WANG received the B.S. degree in mathematics and applied mathematics from Sichuan University, in 2010, and the Ph.D. degree from the Academy of Mathematics and Systems Science, Chinese Academy of Sciences, in 2019. His current research interests include resource allocation, econometrics, and management science.

XIAOBO SONG received the B.S. degree in engineering (automation) from Yanshan University, Qinhuangdao, China, in 2017. She is currently pursuing the M.S. degree in control engineering with the University of Science and Technology Beijing, Beijing, China. Her current research interests include distributed optimization, distributed control, and distributed computation of network systems.

YANPENG HU received the B.S. degree in mechanical engineering from the Beijing Institute of Technology, China, and the M.S. and Ph.D. degrees in aircraft design engineering from Beihang University. He is currently an Associate Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing. His research interests include aircraft design, aircraft flight dynamics, and control.