On the Convergence Properties of a Distributed Projected Subgradient Algorithm

Weijian Li, Zihan Chen, Youcheng Lou and Yiguang Hong

***Abstract*—A weight-balanced network plays an important role in the exact convergence of distributed optimization algorithms, but is not always satisfied in practice. Different from most of existing works focusing on designing distributed algorithms, we analyze the convergence of a well-known distributed projected subgradient algorithm over time-varying general graph sequences, i.e., the weight matrices of the network are only required to be row stochastic instead of doubly stochastic. We first show that there may exist a graph sequence such that the algorithm is not convergent when the network switches freely within finitely many graphs. Then to guarantee its convergence under any uniformly jointly strongly connected graph sequence, we provide a necessary and sufficient condition on the cost functions, i.e., the intersection of optimal solution sets to all local optimization problems is not empty. Furthermore, we surprisingly find that the algorithm is convergent for any periodically switching graph sequence, but the converged solution minimizes a weighted sum of local cost functions, where the weights depend on the Perron vectors of some product matrices of the underlying switching graphs. Finally, we consider a slightly broader class of quasi-periodically switching graph sequences, and show that the algorithm is convergent for any quasi-periodic graph sequence if and only if the network switches between only two graphs. This work helps us understand impacts of communication networks on the convergence of distributed algorithms, and complements existing results from a different viewpoint.**

***Index Terms*—Communication network, constrained distributed optimization, projected subgradient algorithm, convergence analysis**

I. INTRODUCTION

In the past decade, distributed convex optimization has received intensive research attention, motivated by its broad applications in various areas including distributed estimation [1], resource allocation [2], and machine learning [3]. The basic idea is that in a multi-agent network, all agents cooperate to solve an optimization problem with their local cost functions, constraints and neighbors' states. A variety of distributed algorithms have been designed. Distributed (sub)gradients-based algorithms have been greatly explored because they are easy to be implemented. By performing local averaging operations and taking subgradient descent steps, a distributed subgradient algorithm was proposed for unconstrained distributed optimization problems in [4]. Then a projected subgradient algorithm was developed to deal with a common set constraint in [5]. The case of nonidentical set constraints was studied in [6], [7]. Following that, by combining projected subgradient methods and primal-dual ideas, distributed primal-dual subgradient algorithms were designed to minimize a sum of local cost functions with set constraints, local inequality and equality constraints [8]. Moreover, the algorithms have also been explored for coupled constraints [9], [10].

The performance of a distributed algorithm is greatly affected by the underlying graph of a multi-agent network [11], [12]. Since the pioneering work for distributed optimization in [4], weight-balanced graphs have been widely employed to design distributed algorithms [8], [10], [13], [14] because there usually existed a common Lyapunov function to facilitate the convergence analysis [4], [13], [15]. Furthermore, most of (sub)gradient-based algorithms could achieve an optimal solution under weight-balanced graphs because Perron vectors of weight matrices were with identical entries [4], [6], [7]. In [10], undirected and connected graphs were adopted for distributed primal-dual algorithms. Based on saddle-point dynamics, a continuous-time algorithm was proposed under strongly connected and weight-balanced digraphs in [13]. Time-varying weight-balanced graphs have also been utilized for the distributed design [8], [14]. In [16], random weight-balanced networks were applied to a distributed subgradient algorithm, and then the convergence was analyzed.

However, a weight-balanced graph requires the in-degree of each node being equal to its out-degree, and is not always practical in real applications [11]. For instance, if agents use broadcast-based communications in a wireless network, they neither know their out-neighbors nor are able to adjust their outgoing weights. Thus, the weight-balance condition is difficult to be guaranteed in this case [7]. Motivated by the limitation, new mechanisms have been investigated to balance graphs, and then under weight-unbalanced graphs, distributed algorithms achieved exact convergence as balanced networks did [7], [17]–[19]. In [19], reweighting techniques were proposed to deal with fixed unbalanced networks, but Perron vectors of the networks, the global information, were required. In [15], a reweighting method was applied to seeking the Nash equilibrium of a zero-sum game problem. After that, adaptive algorithms were explored to estimate the Perron vectors in [20], [21], and then reweighting was used for the distributed design. Similar ideas were adopted for a continuous-time algorithm in [22]. However, a number of steps was required for the estimation. Under weight-unbalanced

This work was supported by Shanghai Municipal Science and Technology Major Project (No.2021SHZDZX0100), and by the National Natural Science Foundation of China (No. 62173250).

W. Li is with the School of System Design and Intelligent Manufacturing, South University of Science and Technology of China, Shenzhen, Guangdong, China (e-mail: ustcwjli@mail.ustc.edu.cn).

Z. Chen is with Huawei Technologies Co., Ltd., Beijing, China (e-mail: chenzihan9@huawei.com).

Y. Lou is with the MDIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China (e-mail: louyoucheng@amss.ac.cn).

Y. Hong is with the Department of Control Science and Engineering, Tongji University, and is also with Shanghai Research Institute for Intelligent Systems, Shanghai, China (e-mail: yghong@iss.ac.cn).

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3284599

networks, a distributed push-sum method was developed by combining a dual averaging algorithm with the push-sum consensus protocol in [18]. The push-sum protocol was integrated to continuous-time saddle point dynamics in [23], [24]. In [7], a distributed algorithm was proposed over directed graphs with row stochasticity and constraint regularity, where each agent was with a different identity, and knew the upper bound on the network size. With a row stochastic and a column stochastic matrices, a distributed push-pull algorithm was designed in [17], where the gradient of an agent was pushed to its neighbors, and the decision variable was pulled from its neighbors.

On the other hand, an interesting question is how unbalanced networks affect the performance of distributed algorithms. In fact, this question can provide us with a better understanding of existing works, and may further guide us to design effective distributed algorithms. For instance, convergence of a distributed projected subgradient algorithm was discussed in [6], and then heterogeneous stepsizes were designed to balance the network. In distributed learning, weight-unbalanced topologies have been explored to deal with the data heterogeneity, which means the different solutions between individual agent and the global optimization problem. By coupling the communication topology and the data heterogeneity, communication-efficient topologies were proposed for a distributed stochastic gradient descent algorithm in [25]. In [26], a novel topology, named D-Clique, was designed to deal with the data heterogeneity.

In this paper, we revisit a well-known distributed projected subgradient algorithm to minimize a sum of (nonsmooth) cost functions with a common set constraint [5]. Compared with the existing model in [5], the only difference is that the weight matrices of the time-varying network are only row stochastic, and not necessarily doubly stochastic. We aim to analyze convergence properties of this algorithm under such general network graphs. Focusing on analyzing instead of designing a distributed algorithm, this work complements existing results from a different perspective, and moreover, helps us understand how the communication network affects the convergence of distributed algorithms. Our main contributions are summarized as follows.

- We show that there generally exists a graph sequence such that the algorithm is not convergent if the time-varying network switches freely within finitely many graphs.
- To guarantee the convergence of this algorithm for any uniformly jointly strongly connected graph sequence, we provide a necessary and sufficient condition, namely, the intersection of optimal solution sets to all local optimization problems is not empty.
- We find that the algorithm is convergent for any periodically switching graph sequence, and moreover, the converged solution minimizes a weighted sum of the local cost functions. In addition, we introduce a broader class of quasi-periodic graph sequences, and show that the algorithm is always convergent for any quasi-periodic graph sequence if and only if the network switches between two graphs.

The remainder of this paper is organized as follows. Some preliminary knowledge is introduced in Section II, and then the problem is formulated in Section III. Our main results are presented in Sections IV, while their rigorous proofs are provided in Section V. Following that, illustrative examples are carried out in Section VI. Finally, concluding remarks are given in Section VII.

Notations: Let $\mathbb{R}$, $\mathbb{R}^m$ and $\mathbb{R}^{m\times n}$ be the set of real numbers, the set of m-dimensional real column vectors, and the set of m-by-n dimensional real matrices, respectively. Let $\mathbb{N}$ be the set of nonnegative integers. Vectors are column vectors by default. x' stands for the transpose of vector x. $[A]_{ij}$ means the (i,j)-th entry of matrix A. For a matrix X, $X=[x_{ij}]$ means its (i,j)-th entry being x_{ij}. The inner product of x and y is defined by $x'y$. Let $\|\cdot\|$, $\|\cdot\|_1$ be the Euclidean norm and l_1-norm of a vector, respectively. For a matrix $X\in\mathbb{R}^{m\times n}$, $\|X\|_*$ denote its nuclear norm, i.e., the sum of its singular values. Denote $\mathrm{dist}(x,\Omega)$ as the distance from a point x to a set Ω (that is, $\mathrm{dist}(x,\Omega)\triangleq\inf_{y\in\Omega}\|y-x\|$).

II. Preliminary Knowledge

In this section, we introduce some basic concepts related to convex analysis and graph theory.

A. Convex Analysis

A set $\Omega\subset\mathbb{R}^m$ is convex if $\lambda x+(1-\lambda)y\in\Omega$ for all $x,y\in\Omega$ and $\lambda\in[0,1]$. For a closed convex set $\Omega\subset\mathbb{R}^m$, we define $P_\Omega(\cdot):\mathbb{R}^m\to\Omega$ as a projection operator which maps $x\in\mathbb{R}^m$ to a unique point $P_\Omega(x)$ such that $P_\Omega(x)=\mathrm{argmin}_{y\in\Omega}\|x-y\|$. By Lemma 1 in [5], we have

$$\|P_\Omega(x)-y\|\le\|x-y\|,\ \forall x\in\mathbb{R}^m,\ \forall y\in\Omega, \tag{1}$$

and moreover,

$$\begin{aligned}\|P_\Omega(x)-y\|^2\le\|x-y\|^2-\|x-P_\Omega(x)\|^2,\\ \forall x\in\mathbb{R}^m,\ \forall y\in\Omega.\end{aligned} \tag{2}$$

A function $f:\Omega\to\mathbb{R}$ is convex if Ω is a convex set, and

$$\begin{aligned}f(\theta x+(1-\theta)y)\le\theta f(x)+(1-\theta)f(y),\\ \forall x,y\in\Omega,\ \forall\theta\in[0,1].\end{aligned}$$

Furthermore, it is strictly convex if the strict inequality holds whenever $x\neq y$ and $\theta\in(0,1)$. If $g_f(x)\in\mathbb{R}^m$ satisfies

$$f(y)-f(x)\ge(y-x)'g_f(x),$$

then $g_f(x)$ is the subgradient of f at x. Denoted by $\partial f(x)$ the set of all subgradients of f at x.

B. Graph Theory

The interaction topology of a multi-agent network can be modeled by a digraph $\mathcal{G}(\mathcal{V},\mathcal{E})$, where $\mathcal{V}$ is the node set, and $\mathcal{E}$ is the edge set. Then a nonnegative weight matrix $A=[a_{ij}]$ can be associated with $\mathcal{G}$, where $a_{ij}>0$ if and only if $(j,i)\in\mathcal{E}$. Conversely, a graph $\mathcal{G}$ can also be associated with a nonnegative matrix A. Node j is a neighbor of i, and can send information to i if $a_{ij}>0$. Graph $\mathcal{G}$ is said to be weight-balanced if $\sum_{j\in\mathcal{V}}a_{ij}=\sum_{j\in\mathcal{V}}a_{ji}$ for $i\in\mathcal{V}$, and is weight-unbalanced otherwise. A path from i_1 to i_p is defined by an edge sequence $(i_1, i_2), (i_2, i_3), \ldots, (i_{p-1}, i_p) \in \mathcal{E}$ with distinct nodes $i_1, \ldots, i_p$. $\mathcal{G}$ is strongly connected if there exists at least a path between every pair of nodes. If a network is time-varying, we denote $\mathcal{G}(\mathcal{V}, \mathcal{E}(k))$ or $\mathcal{G}(k)$ as the graph at time slot k. Furthermore, the joint graph over the time interval $[k_1, k_2)$ is given by $\mathcal{G}([k_1, k_2)) \triangleq \mathcal{G}(\mathcal{V}, \bigcup_{k\in[k_1,k_2)} \mathcal{E}(k))$.

A vector is said to be stochastic if it is with nonnegative entries and the sum of its entries is 1. Furthermore, it is also positive if all entries of the vector are positive. A matrix is row (column) stochastic if all of its row (column) vectors are stochastic, and is doubly stochastic if it is both row and column stochastic. A row stochastic matrix is also sometimes simply called a stochastic matrix. The following result, collected from Lemma 5.3 in [15], addresses the relationship of positive stochastic vectors and stochastic matrices.

Lemma 1: For any positive stochastic vector $\mu \in \mathbb{R}^n$, there must be a stochastic matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$ such that $\mu' A = \mu'$, and moreover, the graph associated with A is strongly connected.

Let $B \in \mathbb{R}^{n\times n}$ be a stochastic matrix, and $\mathcal{G}_B$ be the associated graph. It follows from the Perron-Frobenius theorem [27] that there is a unique positive stochastic left eigenvector $\mu(B)$ of B associated with eigenvalue 1 if $\mathcal{G}_B$ is strongly connected. We call $\mu(B)$ the Perron vector of B.

III. Formulation and Algorithm

In this section, we revisit a projected subgradient algorithm for a constrained distributed optimization problem, and then give the problem statement.

Consider a network of n agents connected by a time-varying digraph $\mathcal{G}(\mathcal{V}, \mathcal{E}(k))$ (or simply $\mathcal{G}(k)$), where $\mathcal{V} = \{1, \ldots, n\}$ and $\mathcal{E}(k) \subset \mathcal{V}\times\mathcal{V}$. For each $i \in \mathcal{V}$, there is a local (nonsmooth) cost function $f_i : \mathbb{R}^m \to \mathbb{R}$ and a feasible constraint set $X \subset \mathbb{R}^m$. All agents cooperate to minimize the global cost function $\sum_{i\in\mathcal{V}} f_i(x)$ in X. To be strict, the problem can be formulated as

$$\min \ \sum_{i\in\mathcal{V}} f_i(x), \quad \text{s.t.} \ \ x \in X, \tag{3}$$

where x is the decision variable.

Similar to [5], [28], we consider a set constraint X in (3). Thus, the problem is more general with practical applications. We should mention that all results in this paper hold in the absence of X.

Let $x_i(k)$ be the estimation for a solution to (3) by agent i. Then a distributed algorithm is said to achieve a solution to (3) if for any initial condition $x_i(0) \in \mathbb{R}^m$, $\lim_{k\to\infty} \|x_i(k) - x_j(k)\| = 0$, and moreover, there exists $x^* \in X^*$ such that $\lim_{k\to\infty} \|x_i(k) - x^*\| = 0$, where

$$X^* = \Big\{ z \mid z = \operatorname{argmin}_{x\in X} \sum_{i\in\mathcal{V}} f_i(x) \Big\}.$$

To ensure the well-posedness of (3), we make the following standard assumptions.

Assumption 1: (Convexity) For each $i \in \mathcal{V}$, f_i is a convex function on an open set containing X, and X is a closed convex set.

Assumption 2: (Boundedness of Subgradients) For each $i \in \mathcal{V}$, the subgradient set of f_i is bounded over X, i.e., there exists a scalar $L > 0$ such that

$$\|d\| \le L, \ \forall d \in \partial f_i(x), \ \forall x \in X. \tag{4}$$

Assumption 3: (Connectivity) The graph sequence $\mathcal{G}(k)$ is uniformly jointly strongly connected (UJSC), i.e., there exist an integer $B > 0$ such that the joint graph $\mathcal{G}([k, k+B))$ is strongly connected for $k \in \mathbb{N}$.

Assumption 4: (Weight Rule)

(i) The weight matrix $A(k) = [a_{ij}(k)]$ associated with $\mathcal{G}(k)$ is stochastic, i.e., $\sum_{j\in\mathcal{V}} a_{ij}(k) = 1$ for $i \in \mathcal{V}$ and $k \in \mathbb{N}$.

(ii) There are self-loops in all $\mathcal{G}(k)$. Furthermore, if $a_{ij}(k) > 0$, then it is lower bounded by a constant $0 < \eta < 1$. To be specific, $a_{ij}(k) \ge \eta$ if $a_{ij}(k) > 0$, and $a_{ii}(k) \ge \eta$ for $i, j \in \mathcal{V}$ and $k \in \mathbb{N}$.

Note that (3) is a well-known constrained distributed optimization problem [5], [28], [29], and a pioneering distributed algorithm for the formulation is the projected subgradient method, which combines an average step with a local projected gradient update step [5]. For agent i, the specific form of this algorithm is given by

$$\begin{aligned} v_i(k) &= \sum_{j\in\mathcal{V}} a_{ij}(k) x_j(k), \\ x_i(k+1) &= P_X(v_i(k) - \alpha_k d_i(k)), \end{aligned} \tag{5}$$

where $d_i(k) \in \partial f_i(v_i(k))$, and $\alpha_k > 0$ is the stepsize. To guarantee the convergence of (5), the following assumption is made [5].

Assumption 5: (Stepsize Rule) $\sum_{k=0}^{\infty} \alpha_k = \infty$, and moreover, $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$.

Remark 1: In fact, Assumptions 1-5 have also been employed in [5], [8], [30]. As a comparison, we only suppose that the weight matrix $A(k)$ is row stochastic instead of doubly stochastic, i.e., the graph may be weight-unbalanced. Thus, the considered problem is more general.

The following result, proved in [5], shows a convergence property of (5).

Proposition 1: Let Assumptions 1-5 hold. If $A(k)$ is also column stochastic for $k \in \mathbb{N}$, then algorithm (5) achieves a solution to (3).

Proposition 1 indicates the convergence of (5) under graphs with doubly stochastic weight matrices. Following that, great efforts have been paid to develop distributed algorithms over weight-balanced graphs [8], [9], [13], [31]. Furthermore, new methods have been proposed to avoid the weight-balance assumption including the push-sum protocol [18] and the push-pull method [17]. An interesting question is how networks affect the performance of a distributed algorithm. In this paper, taking (5) as a starting point, we explore its convergence under general graphs, i.e., weight matrices of the multi-agent network are only row stochastic instead of doubly stochastic. To be specific, we are interested in the following three questions.

- Is algorithm (5) convergent under general graphs?
- If it is, which solution does it converge to? If not, under what condition on the cost functions it is convergent?
- Is there any class of graph sequences under which algorithm (5) is convergent?

IV. Main Results

In this section, we present the main results on the convergence of (5) under general graph sequences. At first, we show that there generally exists a graph sequence such that (5) is not convergent. Then we provide a necessary and sufficient condition to guarantee its convergence. Finally, we establish its convergence under periodic and quasi-periodic graph sequences.

A. Basic Results

Define $y(k) = \frac{1}{n}\sum_{i\in\mathcal{V}} x_i(k)$ as the average of agents' estimations. The following lemma, proved in Section V-B, shows consensus results of (5).

Lemma 2: Consider algorithm (5). Under Assumptions 2-4, the following statements hold.

(i) If the stepsize satisfies $\lim_{k\to\infty}\alpha_k = 0$, then

$$\lim_{k\to\infty}\|x_i(k)-y(k)\| = 0,\ \forall i\in\mathcal{V}.$$

(ii) If the stepsize satisfies $\sum_{k=0}^{\infty}\alpha_k^2 < \infty$, then

$$\sum_{k=0}^{\infty}\alpha_k\|x_i(k)-y(k)\| < \infty,\ \forall i\in\mathcal{V}.$$

Clearly, (5) can be rewritten as

$$x_i(k+1) = \sum_{j\in\mathcal{V}} a_{ij}(k)x_j(k) + \omega_i(k), \tag{6}$$

where $\omega_i(k) = P_X(v_i(k)-\alpha_k d_i(k)) - v_i(k)$. In fact, (6) is a consensus dynamics with disturbance $\omega_i(k)$. Combining (17) with $\lim_{k\to\infty}\alpha_k = 0$, we obtain $\lim_{t\to\infty}\omega_i(k) = 0$. For such a dynamics, consensus can be achieved under a UJSC graph sequence as discussed in [32], [33]. However, Lemma 2 (ii), indicating the consensus rate, has not been proved in [32], [33].

Referring to Theorem 1 in [6], we have the following result for (5) under a fixed digraph.

Lemma 3: Consider the graph sequence given by $\mathcal{G}(k) = \mathcal{G}_A$ for $k\in\mathbb{N}$, where $\mathcal{G}_A$ is a strongly connected graph associated with weight matrix $A = [a_{ij}]$. Suppose that $\sum_{j\in\mathcal{V}} a_{ij} = 1$ for all $i\in\mathcal{V}$. Under Assumptions 1, 2 and 5, algorithm (5) achieves a solution to

$$\min \quad \sum_{i\in\mathcal{V}}\mu_i(A)f_i(x), \quad \text{s.t.}\quad x\in X, \tag{7}$$

where $\mu(A) = [\mu_1(A),\ldots,\mu_n(A)]'$ is the Perron vector of A such that $\mu(A)'A = \mu(A)'$.

Lemma 3 implies that (5) optimizes a weighted sum of the local cost functions under a fixed weight-unbalanced network. Based on the result, reweighting techniques were proposed to deal with the weight-unbalanced graphs, and relevant works can be found in [6], [15], [19].

Remark 2: Lemma 3 discusses fixed graph sequences. In fact, the result can be directly extended as follows. Let $\mathcal{G}_{A_1},\ldots,\mathcal{G}_{A_p}$ be strongly connected graphs with an identical Perron vector $\mu(A)$. Consider a time-varying graph sequence $\mathcal{G}(k)$, which switches within $\{\mathcal{G}_{A_1},\ldots,\mathcal{G}_{A_p}\}$, i.e., $\mathcal{G}(k)\in\{\mathcal{G}_{A_1},\ldots,\mathcal{G}_{A_p}\}$ for all $k\in\mathbb{N}$. If Assumptions 1, 2, 4 and 5 hold, then algorithm (5) achieves a solution to (7). The proof is similar to that of Theorem 1 in [6], and is omitted here.

B. Convergence Analysis

In this section, we analyze whether (5) is still convergent in the absence of doubly stochastic weight matrices for the network.

Let $\mathcal{G}_{A_1},\ldots,\mathcal{G}_{A_p}$ be strongly connected graphs with weight matrices $A_1,\ldots,A_p$, respectively. We define $\mu(A_l) = [\mu_1(A_l),\ldots,\mu_n(A_l)]'$ as the Perron vectors of A_l for $l\in\{1,\ldots,p\}$, and moreover,

$$X^*_{\mu(A_l)} \triangleq \Big\{z \mid z = \text{argmin}_{x\in X}\sum_{i\in\mathcal{V}}\mu_i(A_l)f_i(x)\Big\}. \tag{8}$$

Then we have the following result, whose proof can be found in Section V-C.

Theorem 1: Let Assumptions 1, 2, 4(i) and 5 hold. Consider a time-varying graph $\mathcal{G}(k)$, which switches within $\{\mathcal{G}_{A_1},\ldots,\mathcal{G}_{A_p}\}$, i.e., $\mathcal{G}(k)\in\{\mathcal{G}_{A_1},\ldots,\mathcal{G}_{A_p}\}$ for all $k\in\mathbb{N}$. If $\bigcap_{l=1}^{p} X^*_{\mu(A_l)} = \emptyset$, then there exists a graph sequence $\{\mathcal{G}(k)\}$ such that algorithm (5) is not convergent.

It follows from Lemma 3 that (5) converges to a consensus solution in $X^*_{\mu(A_l)}$ if $\mathcal{G}(k) = \mathcal{G}_{A_l}$ for all $k\in\mathbb{N}$. As a result, $x_i(k)$ intends to oscillate if $\mathcal{G}(k)$ switches within $\{\mathcal{G}_{A_1},\ldots,\mathcal{G}_{A_p}\}$. This leads to the non-convergent result due to $\bigcap_{l=1}^{p} X^*_{\mu(A_l)} = \emptyset$.

Clearly, Theorem 1 indicates that there exist switching graph sequences such that algorithm (5) is not convergent. Moreover, its proof provides such a sequence. To the best of our knowledge, the result has not appeared in any existing literature. In fact, Theorem 1 is the basis for the analysis of remaining results in this paper.

C. Condition for Convergence

Here we explore a condition to guarantee the convergence of (5), and present the main result in the following theorem, whose proof is given in Section V-D.

Theorem 2: Let Assumptions 1-5 hold. Then algorithm (5) always achieves a solution to (3) for any UJSC graph sequence $\{\mathcal{G}(k)\}$ if and only if $\bigcap_{i\in\mathcal{V}} X_i^* \neq \emptyset$, where

$$X_i^* = \big\{z \mid z = \text{argmin}_{x\in X} f_i(x)\big\}. \tag{9}$$

Remark 3: Consider $\mathcal{G}(k)$ switching within $\{\mathcal{G}_{A_1},\ldots,\mathcal{G}_{A_p}\}$, where $\mathcal{G}_{A_l}$ is a strongly connected graph for $l\in\{1,\ldots,p\}$. From the proof of Theorem 2, if weight matrix A_l associated with $\mathcal{G}_{A_l}$ can be chosen freely under Assumptions 3 and 4, then $\bigcap_{l=1}^{p} X^*_{\mu(A_l)} = \emptyset$ if and only if $\bigcap_{i\in\mathcal{V}} X_i^* = \emptyset$. This implies that the Assumption $\bigcap_{i\in\mathcal{V}} X_i^* = \emptyset$ in Theorem 1 can be cast into $\bigcap_{l=1}^{p} X^*_{\mu(A_l)} = \emptyset$ in some sense.

Remark 4: Theorem 1 provides a necessary and sufficient condition for the convergence of (5), where necessity may follow from Theorem 1 by contradiction. However, it is not so straightforward because it fails for the assumption $\bigcap_{l=1}^{p} X^*_{\mu(A_l)}$ in Theorem 1. Thus, it is required to construct graphs such that $\bigcap_{l=1}^{p} X^*_{\mu(A_l)} = \emptyset$ by $\bigcap_{i\in\mathcal{V}} X_i^* = \emptyset$. For the sufficiency, it is considerable that all agents achieve a consensus solution in $\bigcap_{i\in\mathcal{V}} X_i^*$ because agent i intends to achieve consensus with its neighbors, and meanwhile, forces its state $x_i(k)$ to be in X_i^*. However, its proof is also not straightforward because the graphs switch freely. Its novelty

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3284599

is given as follows. Referring to [34], $\bigcap_{i\in\mathcal{V}} X_i^* = \emptyset$ means that there is no data heterogeneity. Distributed algorithms have been proposed to deal with the data heterogeneity [34], [35]. However, existing works neither have adopted the condition for weight-unbalanced graph sequences, nor have found that it was a sufficient condition for the convergence of (5).

It is worthwhile to mention that Theorem 2 is an extension of results for the convex intersection computation problem [36], [37] as follows. In a network of n nodes, all agents attempt to seek a consensus point in $\bigcap_{i\in\mathcal{V}} \Omega_i$ distributedly, where agent i only knows its local convex set Ω_i, and $\bigcap_{i\in\mathcal{V}} \Omega_i \neq \emptyset$. Suppose that $\sum_{k=0}^{\infty} \alpha_k = \infty$. By the algorithms proposed in [36], [37], the goal is achieved under Assumptions 3 and 4. Define $f_i(x) = \mathrm{dist}^2(x, \Omega_i)$ and $X \in \mathbb{R}^m$. Then the convex intersection computation problem can be cast into (3). The assumption $\bigcap_{i\in\mathcal{V}} \Omega_i \neq \emptyset$ in [36], [37] is a special case $\bigcap_{i\in\mathcal{V}} X_i^* \neq \emptyset$ here. Therefore, the sufficiency in Theorem 2 is an extension for convergence results shown in [36], [37], and the necessity has not appeared in existing works.

D. Periodic Graph Sequences

Theorems 1 and 2 indicate that the convergence of (5) cannot be guaranteed in general if the graph sequence $\{\mathcal{G}(k)\}$ can be chosen and switched freely. However, is it still convergent for some special graph sequences? In this subsection, we investigate the convergence of (5) under periodic and quasi-periodic graph sequences.

Let $\mathcal{G}_{A_l}$ be a graph associated with weight matrix A_l for $l \in \{1, \ldots, p\}$, where $p \geq 2$. Consider $\mathcal{G}(k)$ switching periodically within $\{\mathcal{G}_{A_1}, \ldots, \mathcal{G}_{A_p}\}$. To be specific, the graph sequence is given by $\mathcal{G}(tp + l - 1) = \mathcal{G}(l - 1) = \mathcal{G}_{A_l}$ for all $t \in \mathbb{N}$. For simplicity, we write the sequence as

$$\mathcal{G}_{A_1} \to \cdots \to \mathcal{G}_{A_p} \to \mathcal{G}_{A_1} \to \cdots \to \mathcal{G}_{A_p} \to \cdots .$$

Let $\mu^1, \mu^2, \ldots, \mu^p$ be Perron vectors of

$$(A_p A_{p-1} \cdots A_1),\ (A_{p-1} A_{p-2} \cdots A_1 A_p), \ldots, \\ (A_1 A_p A_{p-1} \cdots A_2),$$

respectively. Then we have the following result, whose proof is provided in Section V-E.

Theorem 3: Let Assumptions 1-5 hold. If the set constraint X is compact and the stepsize sequence $\{\alpha_k\}$ is non-increasing, then algorithm (5) achieves a solution to

$$\min \sum_{i\in\mathcal{V}} \frac{1}{p}(\mu_i^1 + \cdots + \mu_i^p) f_i(x), \quad \text{s.t. } x \in X, \tag{10}$$

where μ_i^l is the i-th entry of μ^l.

Remark 5: In Theorem 3, the assumption that $\{\alpha_k\}$ is non-increasing is adopted to guarantee M_0 being bounded in (30), and thus, it can be relaxed by $\sum_{k=0}^{\infty}(\alpha_{2k} - \alpha_{2k+1}) < \infty$. For instance, there are only finite k such that $\alpha_{k+1} > \alpha_k$.

Remark 6: Theorem 3 indicates that (5) is convergent under a periodically switching graph sequence $\{\mathcal{G}(k)\}$, and moreover, the converged solution minimizes a weighted sum of local cost functions, where the weights depend on the Perron vectors of some product matrices of the underlying periodically switching graphs.

In fact, periodic graph sequences have been applied to multi-agent systems. For instance, the consensus problem was addressed for a class of feedback nonlinear multi-agent systems under periodic networks [38]. It studied the stability of a consensus dynamics under periodic graphs in [39].

Theorem 3 may lead to some applications. For instance, (5) may also be convergent even though underlying graphs in a sequence are weight-unbalanced. If a graph sequence is periodic and $\mu_i^l = \mu_j^l$ in (10) for all $i, j \in \mathcal{V}$, then (5) achieves the exact convergence. Furthermore, the result provides a way to design heterogenous stepsizes to balance the sequences, and achieves the exact convergence under any periodic sequences. Similar to [25], [26], our results may have potential applications to deal with the data heterogeneity in distributed learning.

Because the joint graph $\mathcal{G}([tp, (t+1)p))$ is time-invariant at each time interval $[tp, (t+1)p)$, (5) is convergent under a periodic graph sequence by Lemma 3. However, it should be noted that Theorem 3 is not so straightforward. Intuitively, we may think that (5) tends to achieve a solution to

$$\min \sum_{i\in\mathcal{V}} \hat{\mu}_i f_i(x), \quad \text{s.t. } x \in X, \tag{11}$$

where $\hat{\mu} = [\hat{\mu}_1, \ldots, \hat{\mu}_n]'$ is the Perron vector of $(A_1 A_2 \cdots A_p)$. Clearly, the solution set to (11) is generally different from that of (10), which makes a contradiction. Take $p = 2$ for interpretation. Consider $f_i(x)$ being strictly convex. Then there is a unique solution to any weighted sum of the local cost functions. If (5) achieves a solution to a weighted optimization problem, $\lim_{k\to\infty} x_i(k)$ is independent on the initial state. As a result, (5) reaches the same solution under both graph sequences

$$\mathcal{G}_{A_1} \to \mathcal{G}_{A_2} \to \mathcal{G}_{A_1} \to \mathcal{G}_{A_2} \to \cdots ,$$

and

$$\mathcal{G}_{A_2} \to \mathcal{G}_{A_1} \to \mathcal{G}_{A_2} \to \mathcal{G}_{A_1} \to \cdots .$$

However, the Perron vectors of $A_1 A_2$ and $A_2 A_1$ are generally not identical, which leads to (11) with different solutions under the above two sequences. This implies the incorrectness of (11). By the proof of Theorem 3, we conclude that $\sum_{l=1}^{p} \mu^l$ is independent of the initial point of the graph sequence, which verifies the correctness of Theorem 3.

In [16], the authors considered the weight matrices $A(k)$ drawn independently from a probability space, and explored convergence properties of a distributed subgradient algorithm. It was shown that (5) achieved an optimal solution to (3) with probability 1 if weight matrices $A(k)$ were doubly stochastic with probability 1 and the mean connectivity graph was strongly connected. Intuitively, we can infer that convergence of (5) is dependent on the expectation of the graph sequence. Specifically, if the graph sequence switches randomly between $\{\mathcal{G}_{A_1}, \ldots, \mathcal{G}_{A_p}\}$ by a uniform distribution. Then (5) reaches a solution to

$$\min \sum_{i\in\mathcal{V}} \mu_i^e f_i(x), \quad \text{s.t. } x \in X, \tag{12}$$

where $\mu^e = [\mu_1^e, \ldots, \mu_n^e]'$ is the Perron vector of the expectation for $\sum_{l=1}^{p} \frac{1}{p} A_l$. Clearly, for (10) and (12), it cannot always hold that $\frac{1}{p}(\mu_i^1 + \cdots + \mu_i^p) = \mu_i^e$. Thus, under periodic and random graph sequences, convergence properties of (5) are very different.

It follows from Theorem 3 that a periodic graph sequence is a sufficient condition to guarantee the convergence of (5). It is natural to consider whether the condition is also necessary. We relax the periodic condition slightly, and define a broader class of quasi-periodic graph sequences as follows. Let $\mathcal{G}_{A_l}$ be a graph associated with weight matrix A_l for $l \in \{1, \ldots, p\}$, where $p \geq 2$. $\{\mathcal{G}(k)\}$ is called a quasi-periodic sequence if it switches within $\{\mathcal{G}_{A_1}, \ldots, \mathcal{G}_{A_p}\}$ at each time interval $[tp, (t+1)p)$ for $t \in \mathbb{N}$, but the order of $\mathcal{G}_{A_l}$ can be changed over t. For instance, we consider $p = 3$. Then we can take the graph sequence as $\mathcal{G}_{A_1} \to \mathcal{G}_{A_2} \to \mathcal{G}_{A_3}$ at the time interval $[0, 3)$, and $\mathcal{G}_{A_1} \to \mathcal{G}_{A_3} \to \mathcal{G}_{A_2}$ at the time interval $[3, 6)$. The following theorem, proved in Section V-F, addresses a property of (5) under quasi-periodic graph sequences.

Theorem 4: Let Assumptions 1-5 hold. Suppose that the set constraint X is compact, the stepsize sequence $\{\alpha_k\}$ is non-increasing, and $\bigcap_{i\in\mathcal{V}} X_i^* = \emptyset$. Then algorithm (5) is convergent for any quasi-periodic graph sequence if and only if $p = 2$. Moreover, if $p = 2$, (5) achieves a solution to

$$\min \sum_{i\in\mathcal{V}} \frac{1}{2}(\mu_i^1 + \mu_i^2) f_i(x), \quad \text{s.t. } x \in X, \tag{13}$$

where μ_i^1, μ_i^2 are the i-th entries of Perron vectors of A_2A_1 and A_1A_2, respectively, and A_1 and A_2 are the adjacency matrices of the two graphs.

Under some specific graph sequences, algorithm (5) can also be convergent. For instance, in a sequence, all graphs are the same, all graphs have a same Perron vector, or different orders of the sequence give a same Perron vectors. However, it should be noted that in Theorem 4, it is required that under any graph sequences, the algorithm is convergent. Thus, the result holds.

Remark 7: By Theorem 3, the optimization problem (10) relies on the order of the graph sequence at each time interval $[tp, (t+1)p)$ if $p \geq 3$. With the help of Theorem 1, Theorem 4 can be obtained.

At each time interval $[tL, (t+1)L)$ for $t \in \mathbb{N}$, we consider a graph sequence switching freely between $\{\mathcal{G}_{A_1}, \mathcal{G}_{A_2}, \ldots, \mathcal{G}_{A_p}\}$, where $p \geq 2$ and $L \geq p$. Furthermore, we take the frequency of G_{A_i} be p_i. Then by Theorems 3 and 4, the following statements hold.

1) If the graph sequence is periodic, then (5) converges to a solution of (10).
2) If $p = 2$ and $L = 2$, the algorithm always achieves a solution to (13).
3) If $p > 2$ and $L \geq p$, then there exits a sequence such that (5) is not convergent.

Here we relax the periodic graph in another way. Let $\mathcal{G}_{A_l}$ be a graph associated with weight matrix A_l for $l \in \{1, \ldots, p\}$, where $p \geq 2$. Consider $\mathcal{G}(k)$ switching within $\{\mathcal{G}_{A_1}, \ldots, \mathcal{G}_{A_p}\}$ at each time interval $[tD, (t+1)D)$, where $D > p$. However, $\mathcal{G}_{A_l}$ may appear with different frequencies at time intervals $[tD, (t+1)D)$ for $t \in \mathbb{N}$. For instance, when $D = 3$ and $p = 2$, we can take the graph sequence as $\mathcal{G}_{A_1} \to \mathcal{G}_{A_1} \to \mathcal{G}_{A_2}$ at the time interval $[0, 3)$, and $\mathcal{G}_{A_1} \to \mathcal{G}_{A_2} \to \mathcal{G}_{A_2}$ at the time interval $[3, 6)$. In this case, we have the following corollary.

Corollary 1: Let Assumptions 1-5 hold. Suppose that $\bigcap_{i\in\mathcal{V}} X_i^* = \emptyset$. If A_l can be chosen freely, then there exists a graph sequence such that algorithm (5) is not convergent.

The main idea on the proof focuses on discussing the Perron vector at each time interval $[tD, (t+1)D)$. It is similar to that of the case of $p \geq 3$ in Theorem 4, and is omitted here.

Remark 8: Theorem 4 and Corollary 1 indicate that (5) is not convergent in general if a periodic graph sequence is with a slight modification. Therefore, the periodic condition is very important to guarantee the convergence of (5).

V. Proofs

In this section, we introduce several useful lemmas, and then prove the results presented in the last section.

A. Supporting Lemmas

Referring to [4], [15], we define the transition matrix as

$$\Phi(k, s) = A(k)A(k-1)\cdots A(s)$$

for $s, k \in \mathbb{N}$ with $k \geq s$, where $\Phi(k, k) = A(k)$. Then $\Phi(k, s)$ is a stochastic matrix. In light of Lemma 2 in [4], we have the following result.

Lemma 4: Under Assumptions 3 and 4, $[\Phi(s + (n-1)B - 1, s)]_{ij} \geq \eta^{(n-1)B}$ for all $i, j \in \mathcal{V}$ and $s \in \mathbb{N}$.

The following lemma, found from Lemma 3 in [40], will be used for the consensus analysis.

Lemma 5: For $\mu = [\mu_1, \ldots, \mu_n]' \in \mathbb{R}^n$, we define $g(\mu) = \max_{1\leq i,j\leq n} \|\mu_i - \mu_j\|$. If $P = [p_{ij}] \in \mathbb{R}^{n\times n}$ is a stochastic matrix, then $g(P\mu) \leq \tau(P)g(\mu)$, where $\tau(P) = 1 - \min_{i,j} \sum_{s=1}^n \min\{p_{is}, p_{js}\}$.

We introduce two lemmas about infinite series for the convergence analysis. The first one is a deterministic version of Lemma 11 on page 50 in [41], while the second one is collected from Lemma 7 in [5].

Lemma 6: Let $\{a_k\}, \{b_k\}$ and $\{c_k\}$ be non-negative sequences with $\sum_{k=0}^\infty b_k < \infty$. If $a_{k+1} \leq a_k + b_k - c_k$ holds for all $k \in \mathbb{N}$, then the limit $\lim_{k\to\infty} a_k$ exists and is a finite number.

Lemma 7: Let $0 < \beta < 1$ and $\{\gamma_k\}$ be a positive scalar sequence.

(i) If $\lim_{k\to\infty} \gamma_k = 0$, then $\lim_{k\to\infty} \sum_{s=0}^k \beta^{k-s}\gamma_s = 0$.
(ii) If $\sum_{k=0}^\infty \gamma_k < \infty$, then $\sum_{k=0}^\infty \sum_{s=0}^k \beta^{k-s}\gamma_s < \infty$.

Similar to Lemma 6 in [4], we have the following result for (5). Since it will be frequently used later, we provide a concise proof here.

Lemma 8: Let $x_i(k)$ be generated by algorithm (5). Suppose that Assumptions 1, 2 and 4 (i) hold. For $z \in X$ and $k \in \mathbb{N}$, we have

$$\begin{aligned} \|x_i(k+1) - z\|^2 \leq &\sum_{j\in\mathcal{V}} a_{ij}(k)\|x_j(k) - z\|^2 + \alpha_k^2 L^2 \\ &- 2\alpha_k(f_i(v_i(k)) - f_i(z)). \end{aligned} \tag{14}$$

Proof. Clearly, (5) can be rewritten as

$$x_i(k+1) = v_i(k) - \alpha_k d_i(k) + \varphi_i(k),$$

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3284599

where $\varphi_i(k) = P_X(v_i(k) - \alpha_k d_i(k)) - (v_i(k) - \alpha_k d_i(k))$. Recalling (2) gives

$$\begin{aligned}\|x_i(k+1)-z\|^2 \leq& \|v_i(k) - \alpha_k d_i(k) - z\|^2 - \|\varphi_i(k)\|^2\\ \leq&\|v_i(k) - z\|^2 + \alpha_k^2\|d_i(k)\|^2\\ &- 2\alpha_k(f_i(v_i(k)) - f_i(z)) - \|\varphi_i(k)\|^2.\end{aligned} \tag{15}$$

Because the norm square function is convex,

$$\sum_{j\in\mathcal{V}} a_{ij}(k)\|x_j(k) - z\|^2 \geq \|v_i(k) - z\|^2.$$

By combining (4) and (15), the conclusion follows. □

B. Proof of Lemma 2

Consider $m = 1$ to simplify the proof. Otherwise, the Kronecker product can be adopted when necessary. By (6), we have

$$\begin{aligned}x(k+1) =& A(k)x(k) + \omega(k)\\ =&\Phi(k, s)x(s) + \sum_{r=s}^{k-1}\Phi(k, r + 1)\omega(r) + \omega(k),\end{aligned} \tag{16}$$

where $x(k) = [x_1(k), \ldots, x_n(k)]'$, and $\omega(k) = [\omega_1(k), \ldots, \omega_n(k)]'$. Recalling (1) yields

$$\|\omega_i(k)\| \leq \|v_i(k) - \alpha_k d_i(k) - v_i(k)\| \leq \alpha_k L. \tag{17}$$

Take $h(k) = \max_{i,j\in\mathcal{V}} \|x_i(k) - x_j(k)\|$ and $T = (n - 1)B$. It follows from (16), Lemma 5 and $g(\mu + \nu) \leq g(\mu) + 2\max_i \|\nu_i\|$ that

$$\begin{aligned}h(s + (t + 1)T) \leq& \sum_{r=s+tT}^{s+(t+1)T-1} 2L\alpha_r\\ &+ \tau\big(\Phi(s + (t + 1)T - 1, s + tT)\big)h(s + tT).\end{aligned} \tag{18}$$

In light of Lemma 4, $\Phi(s + (t + 1)T - 1, s + tT) \geq \eta^T$, and then $\tau\big(\Phi(s + (t + 1)T - 1, s + tT)\big) \leq 1 - \eta^T$. Define $\beta_{s,t} = \sum_{r=s+tT}^{s+(t+1)T-1} \alpha_r$. By (18), we obtain

$$\begin{aligned}h(s+(t + 1)T) \leq& (1 - \eta^T)h(s + tT) + 2L\beta_{s,t}\\ \leq&(1 - \eta^T)^{t+1}h(s) + \sum_{r=0}^{t} 2L(1 - \eta^T)^{t-r}\beta_{s,r}.\end{aligned} \tag{19}$$

If $\lim_{t\to\infty} \alpha_t = 0$, then $\lim_{t\to\infty} \beta_{s,t} = 0$. Due to Lemma 7 (i), $\lim_{t\to\infty} \sum_{r=0}^{t}(1 - \eta^T)^{t-r}\beta_{s,r} = 0$. Clearly, $\lim_{t\to\infty}(1 - \eta^T)^{t+1}h(s) = 0$. Thus, $\lim_{k\to\infty} h(k) = 0$. By the definition of $y(k)$, $h(k) \geq \|x_i(k) - y(k)\|$, and then $\lim_{k\to\infty} \|x_i(k) - y(k)\| = 0$.

Combining (19) with $2\alpha_{s+tT}\beta_{s,r} \leq \alpha_{s+tT}^2+\beta_{s,r}^2$, we derive

$$\begin{aligned}\sum_{k=0}^{\infty}\alpha_k h(k) =& \sum_{t=0}^{\infty}\sum_{s=0}^{T-1} \alpha_{s+tT}h(s + tT)\\ \leq& \sum_{s=0}^{T-1}\sum_{t=0}^{\infty}\Big[\alpha_{s+tT}(1 - \eta^T)^t h(s) + \sum_{r=0}^{t-1}(1 - \eta^T)^{(t-1)-r}\beta_{s,r}^2 L\\ &+ \alpha_{s+tT}^2\sum_{r=0}^{t-1}(1 - \eta^T)^{(t-1)-r}L\Big].\end{aligned}$$

Because of $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$, α_k is bounded, and then $\sum_{t=0}^{\infty} \alpha_{s+tT}(1 - \eta^T)^t h(s)$ is also bounded. Obviously, $\beta_{s,t}^2 \leq \sum_{r=s+tT}^{s+(t+1)T-1} 2\alpha_r^2$. By $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$, $\sum_{t=0}^{\infty} \beta_{s,t}^2 < \infty$. Recalling Lemma 7 (ii) gives $\sum_{t=0}^{\infty}\sum_{r=0}^{t-1}(1 - \eta^T)^{(t-1)-r}\beta_{s,r}^2 L < \infty$. In addition, $\sum_{r=0}^{t-1}(1-\eta^T)^{(t-1)-r}L \leq L/\eta^T$, and then $\sum_{t=0}^{\infty} \alpha_{s+tT}^2 \sum_{r=0}^{t-1}(1 - \eta^T)^{(t-1)-r}L < \infty$. Therefore, $\sum_{k=0}^{\infty} \alpha_k h(k) < \infty$, and moreover, $\sum_{k=0}^{\infty} \alpha_k\|x_i(k) - y(k)\| < \infty$. This completes the proof. □

C. Proof of Theorem 1

Because of the convexity of f_i and $\mu_i(A_l) > 0$, $\sum_{i\in\mathcal{V}} \mu_i(A_l)f_i(x)$ is a convex function, and as a result, $X^*_{\mu(A_l)}$ is a closed convex set. Define $x(k) = [x_1'(k), \ldots, x_n'(k)]'$. It follows from Lemma 3 that (5) converges to a point in $X^*_{\mu(A_l)}$ under the fixed graph sequence $\mathcal{G}(k) = \mathcal{G}_{A_l}$ for $k \in \mathbb{N}$. Thus, for any $\epsilon > 0$ and initial point $x(0) \in \mathbb{R}^{mn}$, there is $T_l(\epsilon, x(0)) \in \mathbb{N}$ such that

$$\mathrm{dist}(x_i(t), X^*_{\mu(A_l)}) < \epsilon,\ \forall i \in \mathcal{V},\ \forall t \geq T_l(\epsilon, x(0)).$$

Due to $\bigcap_{l=1}^{p} X^*_{\mu(A_l)} = \emptyset$, there exists $X^*_{\mu(A_i)}$ such that

$$X^*_{\mu(A_i)}\bigcap\big(\bigcap_{j\neq i} X^*_{\mu(A_j)}\big) = \emptyset.$$

In the following, we discuss two cases including $\bigcap_{j\neq i} X^*_{\mu(A_j)} \neq \emptyset$ and $\bigcap_{j\neq i} X^*_{\mu(A_j)} = \emptyset$, respectively.

Case 1. Consider $X^*_{sc} \triangleq \bigcap_{j\neq i} X^*_{\mu(A_j)} \neq \emptyset$. Define $d \triangleq \mathrm{dist}(X^*_{\mu(A_i)}, \bigcap_{j\neq i} X^*_{\mu(A_j)}) > 0$. It follows from the sufficiency of Theorem 2 that there exists a switching graph sequence $\mathcal{G}_{sc}$ such that algorithm (5) converges to X^*_{sc}.

We construct time sequences $\{t_k\}$ and $\{s_k\}$, and a switching graph sequence $\{\mathcal{G}(k)\}$ as follows.

Let $s_0 = 0, t_0 = s_0$ and $x(0) \in \mathbb{R}^{mn}$. Furthermore,

$$\begin{aligned}s_1 =& T_1(d/3, x(t_0)), t_1 = t_0 + s_1,\\ &\text{and } \mathcal{G}(k) = \mathcal{G}_{A_i} \text{ for } k = t_0 + 1, \ldots, t_1;\\ s_2 =& T_2(d/3, x(t_1)), t_2 = t_1 + s_2,\\ &\text{and } \mathcal{G}(k) = \mathcal{G}_{sc} \text{ for } k = t_1 + 1, \ldots, t_2;\\ &\vdots\\ s_{2k+1} =& T_1(d/3, x(t_{2k})), t_{2k+1} = t_{2k} + s_{2k+1},\\ &\text{and } \mathcal{G}(k) = \mathcal{G}_{A_i} \text{ for } k = t_{2k} + 1, \ldots, t_{2k+1};\\ s_{2k+2} =& T_2(d/3, x(t_{2k+1})), t_{2k+2} = t_{2k+1} + s_{2k+2},\\ &\text{and } \mathcal{G}(k) = \mathcal{G}_{sc} \text{ for } k = t_{2k+1} + 1, \ldots, t_{2k+2}.\end{aligned} \tag{20}$$

Then $\|x(t_{2k+1}) - x(t_{2k+2})\| > d/3$ for all $k \in \mathbb{N}$. Therefore, $x(t)$ is not convergent.

Case 2. Consider $\bigcap_{j\neq i} X^*_{\mu(A_j)} = \emptyset$. Clearly, there are $k, l \in \{1, \ldots, p\}$ such that

$$d \triangleq \mathrm{dist}(X^*_{\mu(A_k)}, X^*_{\mu(A_l)}) > 0.$$

By a similar procedure as (20), we can also construct a graph sequence switching between $\mathcal{G}_{A_k}$ and $\mathcal{G}_{A_l}$ such that $x(t)$ is not convergent.

By combining the above results, the conclusion holds. □

D. Proof of Theorem 2

(Necessity). The necessity is shown by contradiction. To be specific, if $\bigcap_{i\in\mathcal{V}} X_i^* = \emptyset$, there always exists a graph sequence such that (5) is not convergent.

By Lemma 1, for any positive stochastic vector $\mu \in \mathbb{R}^n$, there exists a stochastic matrix A, whose Perron vector is μ. Moreover, the graph $\mathcal{G}_A$ associated with A is strongly connected. Here we take two positive stochastic vectors $\mu(A_1)$ and $\mu(A_2)$ associated with matrices A_1 and A_2, and graphs $\mathcal{G}_{A_1}$ and $\mathcal{G}_{A_2}$ as follows. Define $X^*_{\mu(A_1)}$ and $X^*_{\mu(A_2)}$ by (8). Take $\mu(A_1) = [1/n, \ldots, 1/n]$. In light of Lemma 3, all agents converge to a point $\hat{x}$ in $X^*_{\mu(A_1)}$ if $\mathcal{G}(k) = \mathcal{G}_{A_1}$. If $\bigcap_{i\in\mathcal{V}} X_i^* = \emptyset$, then there must be $i_0 \in \mathcal{V}$ and $\tilde{x} \in X^*_{i_0}$ such that $f_{i_0}(\tilde{x}) < f_{i_0}(\hat{x})$. Take $\mu(A_2)$ such that

$$\mu_{i_0}(A_2)\big(f_{i_0}(\hat{x}) - f_{i_0}(\tilde{x})\big) > \sum_{i\in\mathcal{V}, i\neq i_0} \mu_i(A_2)(f_i(\tilde{x}) - f_i(\hat{x})),$$

where $\mu_i(A_2)$ is the i-th entry of $\mu(A_2)$. Consequently, $\sum_{i\in\mathcal{V}} \mu_i(A_2) f_i(\tilde{x}) < \sum_{i\in\mathcal{V}} \mu_i(A_2) f_i(\hat{x})$. Therefore, $X^*_{\mu(A_1)} \bigcap X^*_{\mu(A_2)} = \emptyset$. In view of Theorem 1, there exists a graph sequence such that (5) is not convergent.

(Sufficiency). The sufficiency is proved by the following three steps.

Step 1. We first show that $\{x_i(k)\}$ is bounded.

Define $X_s^* \triangleq \bigcap_{i\in\mathcal{V}} X_i^* \neq \emptyset$ and take $x^* \in X_s^*$. By setting $z = x^*$ in (14), we derive

$$\begin{aligned}\|x_i(k+1) - x^*\|^2 \leq &\sum_{j\in\mathcal{V}} a_{ij}(k)\|x_j(k) - x^*\|^2 + \alpha_k^2 L^2 \\ &- 2\alpha_k(f_i(v_i(k)) - f_i(x^*)).\end{aligned} \tag{21}$$

Define $\xi(k) = \max_{i\in\mathcal{V}} \|x_i(k) - x^*\|^2$. Then

$$\xi(k+1) \leq \xi(k) + \alpha_k^2 L^2 - 2\alpha_k \min\{f_i(v_i(k)) - f_i(x^*)\}.$$

Notice that $f_i(v_i(k)) - f_i(x^*) \geq 0$ and $\sum_{k=0}^\infty \alpha_k^2 L^2 < \infty$. It follows from Lemma 6 that there exists ξ^* such that $\lim_{k\to\infty} \xi(k) = \xi^*$. As a result, $\{x_i(k)\}$ is bounded.

Additionally,

$$\begin{aligned}\max_{i\in\mathcal{V}} \|x_i(k) - x^*\| &- \max_{i,j\in\mathcal{V}} \|x_i(k) - x_j(k)\| \\ &\leq \min_{i\in\mathcal{V}} \|x_i(k) - x^*\| \leq \max_{i\in\mathcal{V}} \|x_i(k) - x^*\|.\end{aligned}$$

Recalling $\lim_{k\to\infty} \max_{i,j\in\mathcal{V}} \|x_i(k) - x_j(k)\| = 0$ gives

$$\lim_{k\to\infty} \|x_i(k) - x^*\| = \xi^*, \ \forall i \in \mathcal{V}.$$

Clearly, $y(k)$ is also bounded. In light of $\lim_{k\to\infty} \|x_i(k) - y(k)\| = 0$, the sequence $\{\|y(k) - x^*\|\}$ is convergent for any $x^* \in X_s^*$.

Step 2. Define $\zeta_i(k) = \|x_i(k) - x^*\|^2$, and analyze $\zeta_i(k)$. Recalling (21), we obtain

$$\begin{aligned}\zeta_i(k+1) \leq &\sum_{j\in\mathcal{V}} [\Phi(k,s)]_{ij}\zeta_j(s) + \alpha_k^2 L^2 \\ &+ \sum_{r=s}^{k-1}\sum_{j\in\mathcal{V}} [\Phi(k,r+1)]_{ij}\alpha_r^2 L^2 - 2\alpha_k\big(f_i(v_i(k)) - f_i(x^*) \\ &- \sum_{r=s}^{k-1}\sum_{j\in\mathcal{V}} 2[\Phi(k,r+1)]_{ij}\alpha_r\big(f_j(v_j(r)) - f_j(x^*)\big).\end{aligned} \tag{22}$$

Clearly,

$$\begin{aligned}&-\sum_{r=s}^{k-1}\sum_{j\in\mathcal{V}} [\Phi(k,r+1)]_{ij}\alpha_r\big(f_j(v_j(r)) - f_j(x^*)\big) \\ &\quad= -\sum_{r=s}^{k-1}\sum_{j\in\mathcal{V}} [\Phi(k,r+1)]_{ij}\alpha_r\big(f_j(v_j(r)) - f_j(y(r))\big) \\ &\quad\quad - \sum_{r=s}^{k-1}\sum_{j\in\mathcal{V}} [\Phi(k,r+1)]_{ij}\alpha_r\big(f_j(y(r)) - f_j(x^*)\big).\end{aligned} \tag{23}$$

By (4), $\|f_j(v_j(r)) - f_j(y(r))\| < L\|v_j(r) - y(r)\|$. In view of Lemma 2 (ii), we obtain

$$\Big\|\sum_{r=s}^{k-1}\sum_{j\in\mathcal{V}} [\Phi(k,r+1)]_{ij}\alpha_r\big(f_j(v_j(r)) - f_j(y(r))\big)\Big\| < \infty.$$

Step 3. We show that $x_i(k)$ converges to a point in X_s^* by contradiction.

For any $x \in X$ and $\epsilon > 0$ such that $\mathrm{dist}(x, X_i^*) > \epsilon$, there must be $\delta > 0$ such that $f_i(x) - f_i(x^*) > \delta$ due to the convexity and continuity of f_i. By Lemma 4, $\Phi(k,s) \geq \eta^{(n-1)B}$ for all $k \geq s + (n-1)B - 1$. For any $\epsilon > 0$, we suppose $\mathrm{dist}(y(r), X_j^*) > \epsilon$. For $k \geq s + (n-1)B$,

$$\begin{aligned}&-\sum_{r=s}^{k-1}\sum_{j\in\mathcal{V}} [\Phi(k,r+1)]_{ij}\alpha_r\big(f_j(y(r)) - f_j(x^*)\big) \\ &\leq -\sum_{r=s}^{k-(n-1)B}\sum_{j\in\mathcal{V}} [\Phi(k,r+1)]_{ij}\alpha_r\big(f_j(y(r)) - f_j(x^*)\big) \\ &\leq -\delta\eta^{(n-1)B}\sum_{r=s}^{k-(n-1)B} \alpha_r.\end{aligned} \tag{24}$$

Substituting (23) and (24) into (22), we obtain $\lim_{k\to\infty} \zeta(k) = -\infty$ by Assumption 5. This contradicts with the boundedness of $x_i(k)$ proved in Step 1. Thus,

$$\lim_{k\to\infty} \inf \mathrm{dist}(y(k), X_s^*) = 0.$$

Since $y(k)$ is bounded, the sequence has at least one limit point. In view of $\lim_{k\to\infty} \inf \mathrm{dist}(y(k), X_s^*) = 0$, one of the limit points, denoted by y^*, must be in X_s^*. As shown in Step 1, the sequence $\{\|y(k) - y^*\|\}$ is convergent, and as a result, the limit point is unique, i.e. $\lim_{k\to\infty} y(k) = y^*$. By $\lim_{k\to\infty} \|x_i(k) - y(k)\| = 0$, we conclude that for all $i \in \mathcal{V}$, $x_i(k)$ converges to $y^* \in X_s^*$. Thus, the proof is completed. □

E. Proof of Theorem 3

Consider $p = 2$ for simplicity, and note that the idea can be directly extended to the case of $p > 2$. We prove the theorem by the following three steps.

Step 1. Analyze the sequences $\{x(2k)\}$ and $\{x(2k+1)\}$.

Without loss of generality, we consider $\mathcal{G}(2k) = \mathcal{G}_{A_1}$ and $\mathcal{G}(2k+1) = \mathcal{G}_{A_2}$ for $k \in \mathbb{N}$. Recalling (14) gives

$$\begin{aligned}\|x_i(2k+1) - z\|^2 \leq &\sum_{j\in\mathcal{V}} [A_1]_{ij}\|x_j(2k) - z\|^2 \\ &+ \alpha_{2k}^2 L^2 - 2\alpha_{2k}\big(f_i(v_i(2k)) - f_i(z)\big),\end{aligned} \tag{25}$$

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3284599

and moreover,

$$\|x_i(2k+2)-z\|^2 \leq \sum_{j\in\mathcal{V}}[A_2]_{ij}\|x_j(2k+1)-z\|^2 + \alpha_{2k+1}^2L^2 - 2\alpha_{2k+1}\big(f_i(v_i(2k+1))-f_i(z)\big). \tag{26}$$

Notice that $\sum_{j\in\mathcal{V}}[A_2]_{ij}=1$. Substituting (25) into (26), we obtain

$$\begin{aligned}\|x_i(2k+2)-z\|^2 \leq& \sum_{j\in\mathcal{V}}[A_2A_1]_{ij}\|x_j(2k)-z\|^2 + (\alpha_{2k}^2 \\ &+ \alpha_{2k+1}^2)L^2 - \sum_{j\in\mathcal{V}}2\alpha_{2k}[A_2]_{ij}(f_j(v_j(2k))-f_j(z)) \\ &- 2\alpha_{2k+1}(f_i(v_i(2k+1))-f_i(z)).\end{aligned} \tag{27}$$

Let $\mu^1=[\mu_1^1,\ldots,\mu_n^1]'$, $\mu^2=[\mu_1^2,\ldots,\mu_n^2]'$ be the Perron vectors of A_2A_1 and A_1A_2 such that $(\mu^1)'A_2A_1=(\mu^1)'$ and $(\mu^2)'A_1A_2=(\mu^2)'$, respectively. As a result,

$$[(\mu^1)'A_2](A_1A_2)=[(\mu^1)'A_2],$$

and

$$[(\mu^2)'A_1](A_2A_1)=[(\mu^2)'A_1].$$

Therefore, $(\mu^1)'A_2$, $(\mu^2)'A_1$ are the Perron vectors of A_1A_2 and A_2A_1, respectively. Because the joint graph $\mathcal{G}_{A_1}\cup\mathcal{G}_{A_2}$ is strongly connected, the Perron vectors of both A_2A_1 and A_1A_2 are unique by the Perron-Frobenius theorem. Thus,

$$(\mu^2)'=(\mu^1)'A_2, \text{ and } (\mu^1)'=(\mu^2)'A_1.$$

Define $X_p^*=\{z \mid z=\operatorname{argmin}_{x\in X}\sum_{i\in\mathcal{V}}\frac{1}{2}(\mu_i^1+\mu_i^2)f_i(x)\}$. Let $x^*\in X_p^*$ and take $z=x^*$. Multiplying μ_i^1 to both sides of (27) and summing all $i\in\mathcal{V}$, we obtain

$$\begin{aligned}\sum_{i\in\mathcal{V}}\mu_i^1\|x_i(2k+2)-x^*\|^2 \leq& \sum_{i\in\mathcal{V}}\mu_i^1\|x_i(2k)-x^*\|^2 \\ &+ \sum_{i\in\mathcal{V}}\mu_i^1(\alpha_{2k}^2+\alpha_{2k+1}^2)L^2 - \sum_{i\in\mathcal{V}}2\mu_i^2\alpha_{2k}\big(f_i(v_i(2k)) \\ &- f_i(x^*)\big) - \sum_{i\in\mathcal{V}}2\mu_i^1\alpha_{2k+1}\big(f_i(v_i(2k+1))-f_i(x^*)\big).\end{aligned}$$

By a similar procedure for discussing $\|x_i(2k+2)-x^*\|$, we also have

$$\begin{aligned}\sum_{i\in\mathcal{V}}\mu_i^2\|x_i(2k+3)-x^*\|^2 \leq& \sum_{i\in\mathcal{V}}\mu_i^2\|x_i(2k+1)-x^*\|^2 \\ &+ \sum_{i\in\mathcal{V}}\mu_i^2(\alpha_{2k+1}^2+\alpha_{2k+2}^2)L^2 \\ &- \sum_{i\in\mathcal{V}}2\mu_i^1\alpha_{2k+1}\big(f_i(v_i(2k+1))-f_i(x^*)\big) \\ &- \sum_{i\in\mathcal{V}}2\mu_i^2\alpha_{2k+2}\big(f_i(v_i(2k+2))-f_i(x^*)\big).\end{aligned}$$

Define $\chi_k=\sum_{i\in\mathcal{V}}\mu_i^1\|x_i(2k)-x^*\|^2+\sum_{i\in\mathcal{V}}\mu_i^2\|x_i(2k+1)-x^*\|^2$. Notice that $\sum_{i\in\mathcal{V}}\mu_i^1=1$ and $\sum_{i\in\mathcal{V}}\mu_i^2=1$. Combining the above two inequalities, we derive

$$\begin{aligned}\chi_{k+1} \leq& \chi_k + (\alpha_{2k}^2+2\alpha_{2k+1}^2+\alpha_{2k+2}^2)L^2 \\ &- \sum_{i\in\mathcal{V}}2(\mu_i^1+\mu_i^2)\alpha_{2k}\big(f_i(v_i(2k))-f_i(x^*)\big) \\ &- \sum_{i\in\mathcal{V}}2(\mu_i^1+\mu_i^2)\alpha_{2k+1}\big(f_i(v_i(2k+1))-f_i(x^*)\big) \\ &+ M_1(k)+M_2(k),\end{aligned} \tag{28}$$

where

$$\begin{aligned}M_1(k)=&\sum_{i\in\mathcal{V}}2\mu_i^1\big[\alpha_{2k}\big(f_i(v_i(2k))-f_i(x^*)\big) \\ &- \alpha_{2k+1}\big(f_i(v_i(2k+1))-f_i(x^*)\big)\big],\end{aligned}$$

and moreover,

$$\begin{aligned}M_2(k)=&\sum_{i\in\mathcal{V}}2\mu_i^2\big[\alpha_{2k+1}\big(f_i(v_i(2k+1))-f_i(x^*)\big) \\ &- \alpha_{2k+2}\big(f_i(v_i(2k+2))-f_i(x^*)\big)\big].\end{aligned}$$

Step 2. Analyze $M_1(k)$ and $M_2(k)$.

For all $N\in\mathbb{N}$, we have

$$\begin{aligned}\sum_{k=0}^N M_1(k)=&\sum_{k=0}^N\sum_{i\in\mathcal{V}}2\mu_i^1\big[\alpha_{2k}\big(f_i(v_i(2k))-f_i(v_i(2k+1))\big) \\ &+ (\alpha_{2k}-\alpha_{2k+1})\big(f_i(v_i(2k+1))-f_i(x^*)\big)\big].\end{aligned} \tag{29}$$

By (4), we obtain

$$\begin{aligned}&\alpha_{2k}\big\|f_i(v_i(2k))-f_i(v_i(2k+1))\big\| \\ &\leq\alpha_{2k}L\|v_i(2k)-v_i(2k+1)\| \\ &\leq\alpha_{2k}L\big(\|v_i(2k)-x_i(2k)\|+\|v_i(2k+1)-x_i(2k+1)\| \\ &\quad+\|x_i(2k)-x_i(2k+1)\|\big).\end{aligned}$$

It follows from Lemma 2 (ii) that

$$\begin{aligned}\lim_{N\to\infty}\sum_{k=0}^N\sum_{i\in\mathcal{V}}&\alpha_{2k}L\big(\|v_i(2k)-x_i(2k)\| \\ &+\|v_i(2k+1)-x_i(2k+1)\|\big)<\infty.\end{aligned}$$

According to (5) and (1),

$$\begin{aligned}&\|x_i(2k)-x_i(2k+1)\|\leq\|x_i(2k)-v_i(2k)\| \\ &\quad+\|x_i(2k+1)-v_i(2k)\|\leq\|x_i(2k)-v_i(2k)\|+\alpha_{2k}L.\end{aligned}$$

In view of Assumption 5,

$$\lim_{N\to\infty}\sum_{k=0}^N\sum_{i\in\mathcal{V}}\alpha_{2k}\|x_i(2k)-x_i(2k+1)\|<\infty.$$

As a result,

$$\lim_{N\to\infty}\sum_{k=0}^N\sum_{i\in\mathcal{V}}\mu_i^1\alpha_{2k}\big\|f_i(v_i(2k))-f_i(v_i(2k+1))\big\|<\infty.$$

Revisit the last term of (29). Notice that (5) implies that both $x_i(k)$ and $y(k)$ are bounded because X is a compact set. Due to the continuity of f_i, there is a positive constant M_0 such that $\|f_i(v_i(k))-f_i(x^*)\|\leq M_0$. Because α_k is non-increasing, we have

$$\begin{aligned}&\sum_{k=0}^N\sum_{i\in\mathcal{V}}\mu_i^1(\alpha_{2k}-\alpha_{2k+1})\big\|f_i(v_i(2k+1))-f_i(x^*)\big\| \\ &\leq\sum_{k=0}^N(\alpha_{2k}-\alpha_{2k+1})M_0 \\ &\leq(\alpha_0-\alpha_1)M_0+\sum_{k=1}^N(\alpha_{2k-1}-\alpha_{2k+1})M_0 \\ &<\alpha_0M_0<\infty.\end{aligned} \tag{30}$$

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3284599

Therefore, $\lim_{N\to\infty}\sum_{k=0}^{N} M_1(k)$ is bounded. By a similar procedure for discussing $M_1(k)$, we also have $\lim_{N\to\infty}\sum_{k=0}^{N} M_2(k) < \infty$.

Step 3. We show that $x_i(k)$ converges to a point in X_p^*.

Note that

$$f_i(v_i(k)) - f_i(x^*) = f_i(v_i(k)) - f_i(y(k)) + f_i(y(k)) - f_i(x^*).$$

By re-arranging the terms of (28) and summing these relations over the time interval $k=0$ to N, we have

$$\begin{aligned}\chi_{N+1} \leq \chi_0 &+ \sum_{k=0}^{N}(\alpha_{2k}^2 + 2\alpha_{2k+1}^2 + \alpha_{2k+2}^2)L^2 \\ &- \sum_{k=0}^{2N+1}\sum_{i\in\mathcal{V}} 2(\mu_i^1+\mu_i^2)\alpha_k\big(f_i(v_i(k)) - f_i(y(k))\big) \\ &- \sum_{k=0}^{2N+1}\sum_{i\in\mathcal{V}} 2(\mu_i^1+\mu_i^2)\alpha_k\big(f_i(y(k)) - f_i(x^*)\big) \\ &+ \sum_{k=0}^{N}\big(M_1(k)+M_2(k)\big).\end{aligned} \tag{31}$$

In light of (4) and Lemma 2 (ii), we obtain

$$\begin{aligned}\lim_{N\to\infty} &- \sum_{k=0}^{2N+1}\sum_{i\in\mathcal{V}} 2(\mu_i^1+\mu_i^2)\alpha_k\big(f_i(v_i(k)) - f_i(y(k))\big) \\ &\leq \lim_{N\to\infty}\sum_{k=0}^{2N+1}\sum_{i\in\mathcal{V}} 2(\mu_i^1+\mu_i^2)\alpha_k\|v_i(k)-y(k)\| < \infty.\end{aligned}$$

Due to the convexity and continuity of f_i, for any $x^*\in X_p^*$ and $\epsilon>0$ such that $\mathrm{dist}(y(k),X_p^*)>\epsilon$, there exists $\delta>0$ such that $\sum_{i\in\mathcal{V}} 2(\mu_{1,i}+\mu_{2,i})\alpha_k\big(f_i(y(k))-f_i(x^*)\big) > \delta$. Then

$$\begin{aligned}\sum_{k=0}^{2N+1}\sum_{i\in\mathcal{V}} &2(\mu_{1,i}+\mu_{2,i})\alpha_k\big(f_i(y(k)) - f_i(x^*)\big) \\ &\geq \sum_{k=0}^{2N+1}\sum_{i\in\mathcal{V}} 2\delta(\mu_{1,i}+\mu_{2,i})\alpha_k.\end{aligned}$$

Under Assumption 5, if $\mathrm{dist}(y(k),X_p^*) > \epsilon$ for any $\epsilon>0$, the right hand of (31) tends to $-\infty$ as N tends to infinity. This contradicts with $\chi_{N+1}\geq 0$. Therefore, $\lim_{k\to\infty}\inf \mathrm{dist}(y(k),X_p^*) = 0$. Since $y(k)$ is bounded, the sequence has at least a limit point. In view of $\lim_{k\to\infty}\inf \mathrm{dist}(y(k),X_p^*) = 0$, one of the limit points, denoted by y^*, must be in X_p^*.

It follows from (28) that

$$\lim_{k\to\infty}\sup \chi_k \leq \lim_{k\to\infty}\inf \chi_k.$$

As a result, the scalar sequence χ_k is convergent. Then the limit point y^* is unique. Due to $\lim_{k\to\infty}\|x_i(k)-y(k)\|=0$, the sequence $\{x_i(k)\}$ converges to the same point y^* in X_p^*. This completes the proof.

F. Proof of Theorem 4

We first show that if $p=2$, (5) is convergent under quasi-periodic graph sequences. Here, notations are the same as those in the proof of Theorem 3. Consider $\mathcal{G}(k)$ switching between $\mathcal{G}_{A_1}$ and $\mathcal{G}_{A_2}$. For quasi-periodic graphs, there are two cases: $\mathcal{G}(2k)=\mathcal{G}_{A_1}$ and $\mathcal{G}(2k+1)=\mathcal{G}_{A_2}$; $\mathcal{G}(2k)=\mathcal{G}_{A_2}$ and $\mathcal{G}(2k+1)=\mathcal{G}_{A_1}$. If $\mathcal{G}(2k)=\mathcal{G}_{A_1}$ and $\mathcal{G}(2k+1)=\mathcal{G}_{A_2}$. As proved in (28), we have

$$\begin{aligned}\chi_{k+1} \leq \chi_k &+ (\alpha_{2k}^2 + 2\alpha_{2k+1}^2 + \alpha_{2k+2}^2)L^2 \\ &- \sum_{i\in\mathcal{V}} 2(\mu_i^1+\mu_i^2)\alpha_{2k}\big(f_i(v_i(2k)) - f_i(x^*)\big) \\ &- \sum_{i\in\mathcal{V}} 2(\mu_i^1+\mu_i^2)\alpha_{2k+1}\big(f_i(v_i(2k+1)) - f_i(x^*)\big) \\ &+ M_1(k) + M_2(k).\end{aligned} \tag{32}$$

If $\mathcal{G}(2k)=\mathcal{G}_{A_2}$ and $\mathcal{G}(2k+1)=\mathcal{G}_{A_1}$, by a similar way for discussing χ_k, we obtain

$$\begin{aligned}\tilde{\chi}_{k+1} \leq \tilde{\chi}_k &+ (\alpha_{2k}^2 + 2\alpha_{2k+1}^2 + \alpha_{2k+2}^2)L^2 \\ &- \sum_{i\in\mathcal{V}} 2(\mu_i^1+\mu_i^2)\alpha_{2k}\big(f_i(v_i(2k)) - f_i(x^*)\big) \\ &- \sum_{i\in\mathcal{V}} 2(\mu_i^1+\mu_i^2)\alpha_{2k+1}\big(f_i(v_i(2k+1)) - f_i(x^*)\big) \\ &+ M_3(k) + M_4(k),\end{aligned} \tag{33}$$

where $\tilde{\chi}_k = \sum_{i\in\mathcal{V}}\mu_i^2\|x_i(2k)-x^*\|^2 + \sum_{i\in\mathcal{V}}\mu_i^1\|x_i(2k+1)-x^*\|^2$,

$$\begin{aligned}M_3(k) = \sum_{i\in\mathcal{V}} 2\mu_i^2\big[&\alpha_{2k}\big(f_i(v_i(2k)) - f_i(x^*)\big) \\ &- \alpha_{2k+1}\big(f_i(v_i(2k+1)) - f_i(x^*)\big)\big],\end{aligned}$$

and moreover,

$$\begin{aligned}M_4(k) = \sum_{i\in\mathcal{V}} 2\mu_i^1\big[&\alpha_{2k+1}\big(f_i(v_i(2k+1)) - f_i(x^*)\big) \\ &- \alpha_{2k+2}\big(f_i(v_i(2k+2)) - f_i(x^*)\big)\big].\end{aligned}$$

Define $\iota_k = \max\{\chi_k,\tilde{\chi}_k\}$, $M_5(k)=\max\{M_1(k),M_3(k)\}$ and $M_6(k)=\max\{M_2(k),M_4(k)\}$. Note that $M_3(k)$ and $M_4(k)$ have similar properties as $M_1(k)$ proved in Section V-E. Combining (32) and (33), we derive

$$\begin{aligned}\iota_{k+1} \leq \iota_k &+ (\alpha_{2k}^2 + 2\alpha_{2k+1}^2 + \alpha_{2k+2}^2)L^2 \\ &- \sum_{i\in\mathcal{V}} 2(\mu_i^1+\mu_i^2)\alpha_{2k}\big(f_i(v_i(2k)) - f_i(x^*)\big) \\ &- \sum_{i\in\mathcal{V}} 2(\mu_i^1+\mu_i^2)\alpha_{2k+1}\big(f_i(v_i(2k+1)) - f_i(x^*)\big) \\ &+ M_5(k) + M_6(k).\end{aligned}$$

Clearly, $\lim_{N\to\infty}\sum_{k=0}^{\infty}\big(M_5(k)+M_6(k)\big) < \infty$. With a similar procedure for discussing χ_k and $x_i(k)$ in Section V-E, we prove that $x_i(k)$ converges to a point in X_p^* for all $i\in\mathcal{V}$, and X_p^* is the optimal solution set to

$$\min \sum_{i\in\mathcal{V}}\frac{1}{2}(\mu_i^1+\mu_i^2)f_i(x), \quad \text{s.t. } x\in X,$$

where μ_i^1, μ_i^2 are the i-th entries of Perron vectors of A_2A_1 and A_1A_2, respectively, and moreover, A_1 and A_2 are the adjacency matrices of the two graphs.

In the following, we show that there exists a graph sequence such that (5) is not convergent if $p\geq 3$. We consider $p=3$ for simplicity, and note that the result can be easily extended to

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3284599

cases of $p > 3$. In view of Theorem 3, for the periodic graph sequence with the order $\mathcal{G}_{A_1} \to \mathcal{G}_{A_2} \to \mathcal{G}_{A_3}$, (5) converges to X^*_{sp}, where X^*_{sp} is the solution set of

$$\min \sum_{i\in\mathcal{V}} (\mu_i^1 + \mu_i^2 + \mu_i^3) f_i(x), \qquad \text{s.t.} \quad x \in X,$$

where μ^1, μ^2 and μ^3 are the Perron vectors of $A_3A_2A_1$, $A_2A_1A_3$ and $A_1A_3A_2$, respectively. Similarly, for the periodic graph sequence with the order $\mathcal{G}_{A_1} \to \mathcal{G}_{A_3} \to \mathcal{G}_{A_2}$, (5) converges to $\tilde{X}^*_{sp}$, where $\tilde{X}^*_{sp}$ is the solution set of

$$\min \sum_{i\in\mathcal{V}} (\tilde{\mu}_i^1 + \tilde{\mu}_i^2 + \tilde{\mu}_i^3) f_i(x), \qquad \text{s.t.} \quad x \in X,$$

where $\tilde{\mu}^1$, $\tilde{\mu}^2$ and $\tilde{\mu}^3$ are the Perron vectors of $A_2A_3A_1$, $A_3A_1A_2$, and $A_1A_2A_3$, respectively. If $\bigcap_{i\in\mathcal{V}} X_i^* = \emptyset$, it follows from Remark 3 that there exist $(\mu_i^1 + \mu_i^2 + \mu_i^3)$ and $(\tilde{\mu}_i^1 + \tilde{\mu}_i^2 + \tilde{\mu}_i^3)$ such that $X^*_{sp} \bigcap \tilde{X}^*_{sp} = \emptyset$. Because the weight matrices can be chosen freely under Assumptions 3 and 4, there will always be A_1, A_2 and A_3 such that $X^*_{sp} \bigcap \tilde{X}^*_{sp} = \emptyset$.

By a similar way used in the proof of Theorem 1, we construct time sequences $\{t_k\}$ and $\{s_k\}$, and a graph sequence $\{\mathcal{G}(k)\}$ switching between $\mathcal{G}_{A_1} \to \mathcal{G}_{A_2} \to \mathcal{G}_{A_3}$ and $\mathcal{G}_{A_1} \to \mathcal{G}_{A_3} \to \mathcal{G}_{A_2}$ at time intervals $[3t, 3(t+1))$. Then (5) is not convergent under $\{\mathcal{G}(k)\}$. This completes the proof. □

VI. Numerical Simulations

Here, an illustrative example is provided to verify the theoretical results presented in Section IV.

Similar to [42], [43], we solve a low-rank matrix completion problem in a distributed way by (5). There is a network of n agents that obtains incomplete and corrupted observations of a matrix $Z^o \in \mathbb{R}^{m_1\times m_2}$. Agent i knows observations from a set Ω_i, and all agents cooperate to recover a true low-rank matrix Z based on their observations and a low-rank constraint. In practice, the low-rank constraint can be approximated by a nuclear norm constraint. To be strict, the problem can be formulated as

$$\min \frac{1}{N} \sum_{i\in\mathcal{V}} \sum_{(s,r)\in\Omega_i} \Big([Z_i]_{s,r} - [Z^{\mathrm{o}}]_{s,r}\Big)^2, \quad \text{s.t.} \quad \|Z\|_* \le \sigma \tag{34}$$

where σ is a positive real constant, N is the number of observations, and $\|\cdot\|_*$ is the nuclear norm.

Take $m_1 = 120$, $m_2 = 200$, $\sigma = 20$, and $\alpha_k = 1/k^{0.6}$. The observations in Z^o are generated from the interval $[1, 5]$ by a uniform distribution. Moreover, there are 50 agents in total. We generate three graphs $\mathcal{G}_{A_1}$, $\mathcal{G}_{A_2}$ and $\mathcal{G}_{A_3}$ randomly, where all their nodes communicate with others with probabilities 0.2, 0.3 and 0.4, respectively. Moreover, Assumptions 3 and 4 hold. For algorithm (5), all entries of $Z_i(0)$ are set as zero.

Let $g_A(Z)$ be the global cost of (34) under a fixed graph $\mathcal{G}_A$. Furthermore, we generate a Erdos-Renyi graph $\mathcal{G}_{er}$, which is an undirected random graph, and each node communicates with others with probability 0.3. Fig. 1 shows trajectories of $c(k) = g_{A_i}(Z(k)) - g_{er}(Z(k))$. The different converged points indicate that (5) achieves a biased optimizer under a weight-unbalanced graph.

Fig. 2 presents the trajectories of $e(k) = \frac{1}{n^2}\sum_{i\in\mathcal{V}} \|x_i(k) - y(k)\|$ under $\mathcal{G}_{A_1}, \mathcal{G}_{A_2}, \mathcal{G}_{A_3}$ and $\mathcal{G}_{er}$. Due to $\lim_{k\to\infty} e(k) = 0$, (5) achieves consensus even though the graphs are only row stochastic.

For the illustration of Theorem 2, we consider all agents knowing Z^o to guarantee $\bigcap_{i\in\mathcal{V}} X_i^* \neq \emptyset$. Furthermore, we define a graph sequence $\mathcal{G}_s(k)$, which switches freely between $\mathcal{G}_{A_1}$, $\mathcal{G}_{A_2}$ and $\mathcal{G}_{A_3}$. Let Z^* be the optimal solution to (34), and it is computed by a centralized projected gradient descent algorithm. Define $r(k) = |g(Z(k)) - g(Z^*)|$. Fig. 3 shows the trajectories of $r(k)$ and $e(k)$ under $\mathcal{G}_s(k)$, and it indicates the consensus and convergence of (5) in this case.

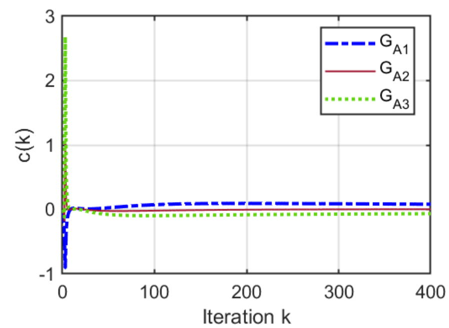


Fig. 1. Trajectories of $c(k)$ under $\mathcal{G}_{A_1}$, $\mathcal{G}_{A_2}$ and $\mathcal{G}_{A_3}$.

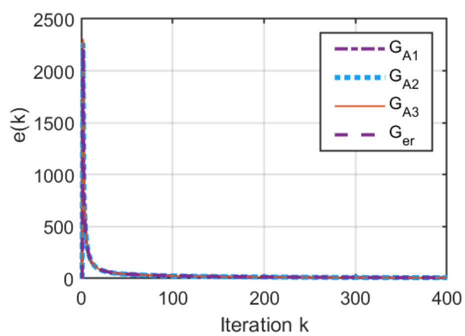


Fig. 2. Consensus results of (5) under $\mathcal{G}_{A_1}$, $\mathcal{G}_{A_2}$, $\mathcal{G}_{A_3}$ and $\mathcal{G}_{er}$.

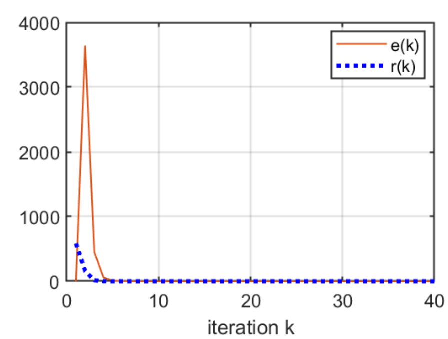


Fig. 3. Trajectory of $r(k)$ and $e(k)$ under $\mathcal{G}_s(k)$.

Finally, we consider the following two cases.

1) The graph sequence switches periodically as

$$\mathcal{G}_{A_1} \to \mathcal{G}_{A_2} \to \mathcal{G}_{A_3} \to \mathcal{G}_{A_1} \to \mathcal{G}_{A_2} \to \mathcal{G}_{A_3} \to \cdots.$$

2) The graph sequence switches freely between $\mathcal{G}_{A_1} \to \mathcal{G}_{A_2}$ and $\mathcal{G}_{A_2} \to \mathcal{G}_{A_1}$ at each time interval $[2t, 2(t+1))$ for all $t \in \mathbb{N}$.

Fig. 4 shows trajectories of $g(Z)$ under the above two sequences. Clearly, algorithm (5) achieves convergence, and the results imply the correctness of Theorems 3 and 4.

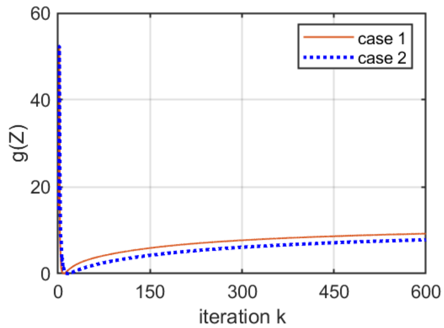


Fig. 4. Trajectory of $g(Z)$ under the two sequences.

VII. Conclusions

This paper investigated convergence properties of a distributed projected subgradient algorithm whose graphs may be time-varying and weight-unbalanced. Firstly, it was proved that there might exist a graph sequence such that the algorithm was not convergent if the network switched freely within finitely many graphs. Then to guarantee the convergence of this algorithm for any uniformly strongly connected graph sequence, it was provided a necessary and sufficient condition, i.e., the intersection of optimal solution sets to all local optimization problems was not empty. Following that, it was found that the algorithm was convergent under periodically switching graph sequences, and optimized a weighted sum of local cost functions. Furthermore, it was shown that the algorithm was always convergent for any quasi-periodic graph sequence if and only if the network switched between two graphs. Finally, numerical simulations on a low-rank matrix completion problem were carried out for illustration.

References

[1] S. Kar, J. M. F. Moura, and K. Ramanan, “Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication,” *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3575–3605, 2012.

[2] D. K. Molzahn, F. Dörfler, H. Sandberg, S. H. Low, S. Chakrabarti, R. Baldick, and J. Lavaei, “A survey of distributed optimization and control algorithms for electric power systems,” *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2941–2962, 2017.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[4] A. Nedić and A. Ozdaglar, “Distributed subgradient methods for multi-agent optimization,” *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, 2009.

[5] A. Nedić, A. Ozdaglar, and P. A. Parrilo, “Constrained consensus and optimization in multi-agent networks,” *IEEE Transactions on Automatic Control*, vol. 55, no. 4, pp. 922–938, 2010.

[6] P. Lin, W. Ren, and Y. Song, “Distributed multi-agent optimization subject to nonidentical constraints and communication delays,” *Automatica*, vol. 65, pp. 120–131, 2016.

[7] V. S. Mai and E. H. Abed, “Distributed optimization over directed graphs with row stochasticity and constraint regularity,” *Automatica*, vol. 102, pp. 94–104, 2019.

[8] M. Zhu and S. Martínez, “On distributed convex optimization under inequality and equality constraints,” *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, 2012.

[9] Y. Zhu, W. Ren, W. Yu, and G. Wen, “Distributed resource allocation over directed graphs via continuous-time algorithms,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 2, pp. 1097–1106, 2021.

[10] X. Zeng, P. Yi, Y. Hong, and L. Xie, “Distributed continuous-time algorithms for nonsmooth extended monotropic optimization problems,” *SIAM Journal on Control and Optimization*, vol. 56, no. 6, pp. 3973–3993, 2018.

[11] A. Nedić, A. Olshevsky, and M. G. Rabbat, “Network topology and communication-computation tradeoffs in decentralized optimization,” in *Proceedings of the IEEE*, 2018, pp. 953–976.

[12] T. Yang, X. Yi, J. Wu, Y. Yuan, D. Wu, Z. Meng, Y. Hong, H. Wang, Z. Lin, and K. H. Johansson, “A survey of distributed optimization,” *Annual Reviews in Control*, vol. 47, pp. 278–305, 2019.

[13] B. Gharesifard and J. Cortés, “Distributed continuous-time convex optimization on weight-balanced digraphs,” *IEEE Transactions on Automatic Control*, vol. 59, no. 3, pp. 781–786, 2014.

[14] D. Yuan, D. W. Ho, and S. Xu, “Regularized primal–dual subgradient method for distributed constrained optimization,” *IEEE Transactions on Cybernetics*, vol. 46, no. 9, pp. 2109–2118, 2016.

[15] Y. Lou, Y. Hong, L. Xie, G. Shi, and K. H. Johansson, “Nash equilibrium computation in subnetwork zero-sum games with switching communications,” *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2920–2935, 2016.

[16] I. Lobel and A. Ozdaglar, “Distributed subgradient methods for convex optimization over random networks,” *IEEE Transactions on Automatic Control*, vol. 56, no. 6, pp. 1291–1306, 2011.

[17] S. Pu, W. Shi, J. Xu, and A. Nedić, “Push-pull gradient methods for distributed optimization in networks,” *IEEE Transactions on Automatic Control*, vol. 66, no. 1, pp. 1–16, 2021.

[18] K. I. Tsianos, S. Lawlor, and M. G. Rabbat, “Push-sum distributed dual averaging for convex optimization,” in *51st IEEE Conference on Decision and Control*. Maui, HI, USA: IEEE, 2012, pp. 5453–5458.

[19] K. I. Tsianos and M. G. Rabbat, “Distributed dual averaging for convex optimization under communication delays,” in *American Control Conference*. Montreal, QC, Canada: IEEE, 2012, pp. 1067–1072.

[20] Z. Qu, C. Li, and F. Lewis, “Cooperative control based on distributed estimation of network connectivity,” in *Proceedings of the 2011 American control conference*. IEEE, 2011, pp. 3441–3446.

[21] A. Priolo, A. Gasparri, E. Montijano, and C. Sagues, “A decentralized algorithm for balancing a strongly connected weighted digraph,” in *2013 American Control Conference*. IEEE, 2013, pp. 6547–6552.

[22] D. Yue, S. Baldi, J. Cao, and B. De Schutter, “Distributed adaptive optimization with weight-balancing,” *IEEE Transactions on Automatic Control*, vol. 67, no. 4, pp. 2068–2075, 2021.

[23] B. Touri and B. Gharesifard, “Continuous-time distributed convex optimization on time-varying directed networks,” in *2015 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2015, pp. 724–729.

[24] ——, “Saddle-point dynamics for distributed convex optimization on general directed graphs,” in *2016 54th IEEE Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 862 – 866.

[25] L. Bars, Batiste, A. Bellet, M. Tommasi, and E. Lavoie, “Refined convergence and topology learning for decentralized optimization with heterogeneous,” *Arxiv: 2204.04452*, 2022.

[26] A. Bellet, A.-M. Kermarrec, and E. Lavoie, “D-cliques: Compensating for data heterogeneity with topology in decentralized federated learning,” 2021.

[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge university press, 2012.

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

[29] W. Li, X. Zeng, S. Liang, and Y. Hong, “Exponentially convergent algorithm design for constrained distributed optimization via non-smooth approach,” *IEEE Transactions on Automatic Control*, 2021, to be published, doi: 10.1109/TAC.2021.3075666.

[30] S. Liu, Z. Qiu, and L. Xie, “Convergence rate analysis of distributed optimization with projected subgradient algorithm,” *Automatica*, vol. 83, pp. 162–169, 2017.

[31] X. Li and L. Xie, "Distributed algorithms for computing a fixed point of multi-agent nonexpansive operators," *Automatica*, vol. 122, p. 109286, 2020.

[32] G. Shi and K. H. Johansson, "Robust consensus for continuous-time multiagent dynamics," *SIAM Journal on Control and Optimization*, vol. 51, no. 5, pp. 3673–3691, 2013.

[33] L. Wang and L. Guo, "Robust consensus and soft control of multi-agent systems with noises," *Journal of Systems Science and Complexity*, vol. 21, no. 3, pp. 406–415, 2008.

[34] W. Shi, Q. Ling, G. Wu, and W. Yin, "EXTRA: An exact first-order algorithm for decentralized consensus optimization," *SIAM Journal on Optimization*, vol. 25, no. 2, pp. 944–966, 2015.

[35] J. Xu, S. Zhu, C. Soh, Yeng, and L. Xie, "Augmented distributed gradient methods for multi-agent optimization under uncoordinated constant stepsizes," in *54th IEEE Conference on Decision and Control*. IEEE, 2015, pp. 2055–2060.

[36] Y. Lou, Y. Hong, and S. Wang, "Distributed continuous-time approximate projection protocols for shortest distance optimization problems," *Automatica*, vol. 69, pp. 289–297, 2016.

[37] G. Shi, K. H. Johansson, and Y. Hong, "Reaching an optimal consensus: Dynamical systems that compute intersections of convex sets," *IEEE Transactions on Automatic Control*, vol. 58, no. 3, pp. 610–622, 2013.

[38] H. Kim, H. Shim, J. Back, and H. Seo, Jin, "Consensus of multi-agent systems under periodic time-varying network," in *IFAC Proceedings Volumes*, vol. 43, no. 14. Elsevier, 2010, pp. 155–160.

[39] M. Mohammad and B. Eric, A, "Consensus stability in multi-agent systems with periodically switched communication topology using floquet theory," *Transactions of the Institute of Measurement and Control*, vol. 43, no. 5, pp. 1239–1254, 2021.

[40] J. Hajnal and M. S. Bartlett, "Weak ergodicity in non-homogeneous markov chains," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 54, no. 2, pp. 233–246, 1958.

[41] B. T. Polyak, *Introduction to Optimization*. New York, NY, USA: Optimization Software, Inc., 1987.

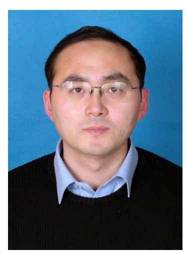
[42] H.-T. Wai, J. Lafond, A. Scaglione, and E. Moulines, "Decentralized frank–wolfe algorithm for convex and nonconvex problems," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 5522–5537, 2017.

[43] W. Li, X. Zeng, Y. Hong, and H. Ji, "Distributed design for nuclear norm minimization of linear matrix equations with constraints," *IEEE Transactions on Automatic Control*, vol. 66, no. 2, pp. 745–752, 2020.

Weijian Li received the B.Sc. degree in mechanical engineering from Wuhan University of Technology in 2016, and the Ph.D. degree in operations research and cybernetics from the University of Science and Technology of China in 2021. He was also a joint Ph. D. student in Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences. He worked at the Decision Intelligence Lab, DAMO Academy, Alibaba Group from December, 2021 to March, 2023. At present, he is a visiting scholar at the South University of Science and Technology of China. His research interests include distributed optimization, game and learning, network control and Bayesian signal processing.

Zihan Chen received the B.Sc. degree in mathematics and applied mathematics from University of Science and Technology of China, Hefei, China, in 2015, and the Ph.D. degree in complex systems and control from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2020. He is currently a researcher in Huawei Technologies Co., Ltd., Beijing, China. His research interests include social network analysis and distributed optimization.

Youcheng Lou received the B.Sc. degree in mathematics and applied mathematics from Zhengzhou University, Zhengzhou, China, in 2008 and the Ph.D. degree in complex systems and control from the Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), Beijing, China, in 2013. From 2013 to 2018, he was a Postdoctoral Researcher with the National Center for Mathematics and Interdisciplinary Sciences, AMSS, CAS, where from 2016 to 2018, he was also a Postdoctoral Fellow with the Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong, funded by Hong Kong Scholars Program. From 2018 to 2020, he was an assistant professor with the AMSS, CAS. He is currently an associate professor with the AMSS, CAS. Currently he serves as an associate editor of International Journal of Economic Theory. His current research areas include rational expectations equilibrium theory, market microstructure, social economic networks and distributed optimization.



Yiguang Hong received his B.Sc. and M.S. degrees from Peking University, China, and the Ph.D. degree from the Chinese Academy of Sciences (CAS), China. He is currently a professor of Shanghai Institute of Intelligent Science and Technology, Tongji University from Oct 2020, and was a professor of Academy of Mathematics and Systems Science, CAS. His current research interests include nonlinear control, multi-agent systems, distributed optimization/game, machine learning, and social networks. Prof. Hong serves as Editor-in-Chief of Control Theory and Technology. He also serves or served as Associate Editors for many journals, including the IEEE Transactions on Automatic Control, IEEE Transactions on Control of Network Systems, and IEEE Control Systems Magazine. He is a recipient of the Guan Zhaozhi Award at the Chinese Control Conference, Young Author Prize of the IFAC World Congress, Young Scientist Award of CAS, the Youth Award for Science and Technology of China, and the National Natural Science Prize of China. He is also a Fellow of IEEE, a Fellow of Chinese Association for Artificial Intelligence, and a Fellow of Chinese Association of Automation (CAA). Additionally, he was a board of governor of IEEE Control Systems Society.