Distributed Convex Optimization With State-Dependent (Social) Interactions and Time-Varying Topologies

Seyyed Shaho Alaviani, *Member, IEEE*, and Nicola Elia, *Fellow, IEEE*

***Abstract*—In this paper, an unconstrained collaborative optimization of a sum of convex functions is considered where agents make decisions using local information from their neighbors. The communication between nodes are described by a *time-varying* sequence of possibly *state-dependent* weighted networks. A new framework for modeling multi-agent optimization problems over networks with state-dependent interactions and time-varying topologies is proposed. A gradient-based discrete-time algorithm using diminishing step size is proposed for converging to the optimal solution under suitable assumptions. The algorithm is totally asynchronous without requiring B-connectivity assumption for convergence. The algorithm still works even if the weighted matrix of the graph is periodic and irreducible in synchronous protocol. Finally, a case study on a network of robots in an automated warehouse is provided in order to demonstrate the results.**

***Index Terms*—Distributed optimization, convex optimization, state-dependent networks, time-varying topologies.**

I. Introduction

DISTRIBUTED collaborative optimization of a sum of convex functions where agents make decisions using local information has been a hot topic to research during recent years due to its applications in many areas such as power systems, communication networks, sensor networks, smart buildings, and machine learning [1]–[4]. Many researchers have paid much effort to research on distributed optimization problems (see Surveys [1]–[4] and references therein).

Discrete-time algorithms have been presented to be an efficient way from existing optimization approaches to distributed networks [5]–[14], to cite a few. Furthermore, distributed optimization algorithms based on the framework of continuous-time dynamical systems have been proposed [15]–[20], to cite a few. Recently, Alternating Direction Method of Multipliers (ADMM) has been a suitable method for solving distributed optimization problems (see [21]–[27] and references therein).

In a *synchronous* protocol, all nodes activate at the same time and perform communication updates. This protocol requires a common notion of time among the nodes. On the other hand, in an *asynchronous* protocol, each node has its own concept of time defined by a local timer which randomly triggers either by the local timer or by a message from neighboring nodes. The algorithms guaranteed to work with no a priori bound on the time for updates are called *totally asynchronous*, and those that need a priori bound known also as B-connectivity[1] assumption are called *partially asynchronous* (see [28] and [29, Ch. 6-7]). As the dimension of the network increases, synchronization becomes an issue. Therefore, some investigators have considered asynchronous distributed optimization problems [30]–[36], to cite a few, where the authors assume B-connectivity. All these results are concerned under an assumption that the switching communication graph is *time-dependent*. The aforementioned papers consider networks where the information transmission weight between the adjacent agents is independent of the agents' states called *state-independent* networks.

However, in practice, the quality of a link between two agents not only depends on *time* but also depends on the agents' states called *state-dependent* networks. We need to emphasize that "time" and "states" are two different variables such that the weight of a link is a priori determined in time-dependent networks, whereas it cannot be a priori determined at any time in state-dependent networks (see [37] for more details).

State-dependent networks appear in several practical systems such as flocking of birds [38], opinion dynamics [39]–[47], mobile robotic networks (see [48] and references therein), wireless networks [49], and predator-prey interaction [50]. Flocking of birds in biological systems has been modeled as a state-dependent network of Cucker-Smale form in [38] via a transmission weight depending on the Euclidean distance between agents. The Cucker-Smale weight is designed similar to *Newton's law of gravitation*, i.e., as the distance between any two agents increases, the information that they receive from each other weakens but always exists. In opinion dynamics, revision of the opinion of a social agent depends on the differences

Manuscript received December 21, 2019; revised June 21, 2020, December 8, 2020, and February 28, 2021; accepted March 22, 2021. Date of publication March 31, 2021; date of current version May 14, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Subhro Das. This work was supported by NSF under Grant CCF-1320643, and AFOSR under Grant FA95501510119. A preliminary version of this paper has appeared without proofs in [11]. This work had been done primarily while Seyyed Shaho Alaviani and Nicola Elia were at Iowa State University, Ames, IA, USA. *(Corresponding author: Seyyed Shaho Alaviani.)*

Seyyed Shaho Alaviani is with the Department of Mechanical Engineering, Clemson University, Clemson, SC 29634 USA (e-mail: salavia@clemson.edu).

Nicola Elia is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: nelia@umn.edu).

Digital Object Identifier 10.1109/TSP.2021.3070223

[1]There exists a bounded time interval such that union of the graphs is strongly connected and each edge transmits a message at least once.

between its opinion and opinions of other agents, resulting in state-dependent networks by considering the opinion as the state [39]–[47]. In mobile robotic networks or wireless networks, the quality of the link between two agents depends on the distance between them such that the links get weaker as the positions of agents are far from each other, resulting in state-dependent networks in reality (by considering the position as the state). The genome is viewed as a state-dependent network [51]. In a network of spacecrafts, communications depend on distances among spacecrafts, resulting in state-dependent networks (by considering the positions as states) [52], [53]. As stated, state-dependent networks appear in many real networks.

In social networks, an agent weighs the opinions of others differently depending on how close it is to other agents. The authors in [39]–[47] consider consensus problem for opinion dynamics. In [54], existence of consensus (as a special case of distributed optimization) in a multi-robot network has been investigated. The state-dependence of the communication introduces significant challenges and couples the study of information exchange with the analysis of an algorithm [12]. In fact, in state-dependent communications, weights of links are determined *endogenously* by the states of agents at each time so that no local property can be assumed or checked a priori. Therefore, designing distributed algorithms for consensus and optimization over state-dependent networks is still a challenging problem. Several authors have investigated distributed consensus [56]–[60] and distributed optimization [12], [15], over state-dependent networks.

Although state-dependent communication graphs may be viewed as a case of time-varying communication graphs studied in [55], we (and also the authors in [56]) believe that distributed consensus under state-dependent interactions has its specific properties and needs to be further studied since the dynamics in a state-dependent network is driven by *only local interactions*, an agent does *not* have *any* choice to weigh the information received from others, and the weight of each link and/or network connectivity is *a priori* unknown at each time while *only* determined by the states of the agents at the time. Indeed, not only *time* but also *agents' states* can change the topology of the network at each time (see Novelty in Section IV). For instance, the authors in [57] have developed an approach to predicting network connectivity in a state-dependent network by initial conditions, verifying that network connectivity is a priori unknown at each time and only determined by the states of the agents.

The synchronous discrete-time algorithms proposed in [56] can solve distributed consensus problems over non-switching undirected graphs where the weight of a link between two agents is nonincreasing with respect to Euclidean distance between their states and the algorithms need a parameter to be chosen based on a global information. The result in [56] allows to consider state-dependent networks of Cucker-Smale weight model [38] which does not have positive lower bound (like Newton's law of gravitation). The synchronous discrete-time algorithm proposed in [58] can solve distributed consensus problems over time-varying graphs *without* B-connectivity assumption for a specific weight model where states of agents are scalar. Continuous algorithms have been proposed in [56], [57], [59], [60] for solving distributed consensus problems over non-switching graphs with state-dependent interactions. Since consensus problems are special cases of distributed optimization problems, solving distributed convex optimization problems over state-dependent networks are very important and useful.

Recently, in [12] and [15], the authors have considered distributed multi-agent optimization problems over state-dependent networks. In [12], the authors consider constrained distributed optimization and assume that the weighted matrix of the graph is Markovian on the state variables at each iteration, where the weight of each link has a positive lower bound when activated. They assume that the probability of activating a link between two agents depends on the distance between their states. Moreover, they assume that weighted matrix of the graph is doubly stochastic with positive diagonals with probability one. In [15], a continuous-time system is proposed for unconstrained optimal consensus of convex optimization problem over directed time-varying networks; the authors assume that the weight of each link has positive lower bound. Furthermore, they assume that the intersection of the set of optimal solutions of each agent's cost function should be non-empty. The continuous-time algorithm they propose needs to project each agent's state into its optimal solution set at each time.

Contribution: In this paper, we consider an unconstrained distributed convex optimization problem over *possibly* state-dependent networks. Due to uncertainty in a wireless channel, it is often impossible to ensure all time connectivity in practice. Motivated by this fact, we also consider *time-varying* topologies of the network. In contrast to [12], the *state-dependent weight* and the *time-dependent* switching of a link in our setting, like that of [58], are distinct. We propose a new framework[2] for modeling multi-agent optimization problems over state-dependent networks with time-varying topologies. We assume that weighted graph matrix of the network is doubly stochastic with respect to state variables for each interconnection graph and that the union of the interconnection graphs is strongly connected. This assumption allows us to consider weighted matrix of the graph to be *periodic and irreducible* in synchronous protocol (see Comment 2). We generalize the gradient-based algorithm using diminishing step size proposed for state-independent networks in [61] for distributed optimization with state-dependent networks and time-varying topologies. The differences with the work of [61] are: 1) the weight of a link in [61], and other work such as [9], to cite a few, takes values in a set having finite cardinality, whereas here the weight takes values in a continuum set; and 2) the weight of each link, taking values in this continuum set, is *a priori* unknown at each iteration while *only* determined by the states of the agents at the iteration. Therefore, the algorithm here cannot be directly derived from the algorithm in [61]. The algorithm[3] does not require the weights to have positive lower bounds. This allows us to consider networks with weights of Cucker-Smale model [38] which has significant potential applications in networks of spacecrafts [52], [53]. In

[2]The proposed framework in [11] is for non-switching graphs. Here we extend the framework to switching graphs.

[3]The proposed algorithm in this paper is an extension of that of [11] such that the algorithm here does not require the interconnection graph to be strongly connected at each time for convergence.

our algorithm, distributed consensus on an optimal solution can be reached under arbitrary initial states. We do *not* develop an approach to predicting network connectivity, so, based on states of agents, the weight of each link as well as network connectivity are completely unknown at each iteration. Our result does not need that intersection of the set of optimal solutions of each agent's cost function to be non-empty. The proposed algorithm is totally asynchronous. Therefore, to the best of our knowledge, the proposed algorithm is *not* comparable *even* for state-independent networks with existing algorithms since it does not assume B-connectivity assumption for convergence. We mention that this work is built on the work of [61] which considers *state-independent* networks.

The paper is organized as follows. In Sections II and III, preliminaries and formulation of the problem are given, respectively. Main results are given in Section IV. In Section V, a numerical example is given to show the results followed by Conclusions in Section VI.

Notations: $\Re$ denotes the set of all real numbers. $\Re^+$ denotes the set of positive real numbers. We use 2-norm for vectors and induced 2-norm for matrices, i.e., for any vector $z \in \Re^n$, $\|z\| = \|z\|_2 = \sqrt{z^T z}$, and for any matrix $Z \in \Re^{n\times n}$, $\|Z\| = \|Z\|_2 = \sqrt{\lambda_{max}(Z^T Z)} = \sigma_{max}(Z)$ where Z^T represents the transpose of matrix Z, λ_{max} represents maximum eigenvalue, and σ_{max} represents largest singular value. Sorted in an increasing order with respect to real parts, $\lambda_2(Z)$ represents the second eigenvalue of a matrix Z. $Re(r)$ represents the real part of the complex number r. For any matrix $Z \in \Re^{n\times n}$ with $Z = [z_{ij}]$, $\|Z\|_1 = max_{1\leq j\leq n}\{\sum_{i=1}^n z_{ij}\}$ and $\|Z\|_\infty = max_{1\leq i\leq n}\{\sum_{j=1}^n z_{ij}\}$. I_n represents Identity matrix of size $n \times n$ for some $n \in \mathbb{N}$ where $\mathbb{N}$ denotes the set of all natural numbers. $\nabla f(x)$ denotes the gradient of the function $f(x)$. $\otimes$ denotes the Kronecker product. $\mathbf{1}_m$ represents the vector of dimension m whose all entries are 1.

II. Preliminaries

A vector $v \in \Re^n$ is said to be a *stochastic vector* when its components $v_i, i = 1, 2, \ldots, n$, are non-negative and their sum is equal to 1; a square $n \times n$ matrix V is said to be a *stochastic matrix* when each row of V is a stochastic vector. A square $n \times n$ matrix V is said to be *doubly stochastic matrix* when both V and V^T are stochastic matrices.

Let X be a real Hilbert space with norm $\|.\|$ and inner product $\langle ., . \rangle$. An operator $A : X \longrightarrow X$ is said to be *monotone* if $\langle x - y, Ax - Ay\rangle \geq 0$ for all $x, y \in X$. $A : X \longrightarrow X$ is called *ρ-strongly monotone* if $\langle x - y, A(x) - A(y)\rangle \geq \rho\|x - y\|^2$ for all $x, y \in X$. A function $f(x)$ is *ρ-strongly convex* if $\langle x - y, \nabla f(x) - \nabla f(y)\rangle \geq \rho\|x - y\|^2$ for all $x, y \in X$. Therefore, a function is ρ-strongly convex if its gradient is ρ-strongly monotone. A function $f(x)$ is *$\mathcal{L}$-strongly smooth* if

$$\langle x - y, \nabla f(x) - \nabla f(y)\rangle \leq \mathcal{L}\|x - y\|^2, \forall x, y \in X.$$

A mapping $B : X \longrightarrow X$ is said to be *K-Lipschitz continuous* if there exists a $K > 0$ such that $\|B(x) - B(y)\| \leq K\|x - y\|$ for all $x, y \in X$.

Definition 1: Let the operator $T : X \longrightarrow X$ be a self map. A $x \in X$ is said to be a *fixed point* of T if $T(x) = x$, and $Fix(T)$ denotes the set of all fixed points of T.

Definition 2: A self map operator $H : X \longrightarrow X$ is said to be *nonexpansive* if for any $x, y \in X$ we have

$$\|H(x) - H(y)\| \leq \|x - y\|.$$

Lemma 1: [62]: Let $W \in \Re^{m\times m}$. Then $\|W\|_2 \leq \sqrt{\|W\|_1\|W\|_\infty}$.

Lemma 2: [63]: Let $\{a_t\}_{t=0}^\infty$ be a sequence of nonnegative real numbers satisfying

$$a_{t+1} \leq (1 - b_t)a_t + b_t h_t, \quad t \geq 0$$

where $b_t \in [0, 1], \sum_{t=0}^\infty b_t = \infty$, and $\limsup_{t\longrightarrow\infty} h_t \leq 0$. Then $\lim_{t\longrightarrow\infty} a_t = 0$.

Lemma 3: [64] *(The Banach Fixed Point Theorem):* Assume that S is a closed, nonempty set in the Banach space X, and the operator $H : S \longrightarrow S$ is a *contraction*, i.e.,

$$\|Hu - Hv\| \leq \kappa\|u - v\|, \quad \forall u, v \in S$$

and $0 \leq \kappa < 1$. Then the equation $u = Hu, u \in S$, has a unique fixed point on the set S.

III. Problem Formulation

People's *beliefs* (or *opinions*) are not fixed. In fact, individuals form beliefs (or opinions) on various economic, political, or social variables ("state") based on information they receive from others, including friends, neighbors, coworkers, local leaders, news sources, and political actors [41]–[47]. A key question faced by any society is whether this process of information exchange will lead to the formation of more accurate beliefs [41]. A famous idea going back to Condorcet's Jury Theorem [65] encapsulates the idea that exchange of dispersed information will enable socially beneficial aggregation of information [41].

To precisely formulate the above social problem, consider a set of $\mathcal{V} = \{1, 2, \ldots, m\}$ agents interacting over a social network. Each agent $i \in \mathcal{V}$ has its own opinion about an underlying state but is open to some extent to revise it when being informed about the opinions of other agents [39]. Knowing the revisions may lead to further revisions and the question then is if this iterative process of changing opinions will tend to a consensus among the agents [39]. Denote by $x_{i,t} \in \Re^n$ the quantified opinion of agent $i \in \mathcal{V}$ at time $t \in \{0, 1, \ldots\}$. Suppose that agent i arrives at a revision $x_{i,t+1}$ by taking the quantified opinion $x_{j,t}$ of the other agents into account with certain weights $a_{ij} = \phi(\varrho(x_i, x_j))$ where $0 < \phi < 1$ is the scaled *influence function* which acts on the "difference of opinions" $\varrho(x_i, x_j)$ [39], [40]. The metric $\varrho(., .)$ needs to be properly interpreted, adapted to the specific context of the problem at hand [40]. If the distance (in metric ϱ) between the quantified opinions of agents i and j, namely x_i and x_j, is large enough, then the opinion of agent j may *not* influence the revision of the agent i.

As seen in social networks, agents weigh their beliefs differently depending on how close they are to other agents. Indeed, an agent does *not* have *any* choice to weigh other agents' beliefs. So one can ask the relevant question, can we achieve distributed optimization with social agents whose interactions depend on their proximity?

Moreover, in other practical networks such as flocking of birds [38], mobile robotic networks [48], [54], or wireless networks [49], the weight of the interaction link between agents i and j, namely $w_{ij} : \|x_i - x_j\| \longrightarrow [0, 1]$, is assumed to depend on Euclidean distance of the positions of agents i and j, namely x_i and x_j, respectively. As a direct consequence, the entries of the weighted graph matrix of the network depend on states of the agents, resulting in *state-dependent* networks. We stress that existence of consensus in a multi-robot network has been investigated in [54] as a special case of distributed optimization problems (see also [48]). The Cucker-Smale weight [38], which depends on distance between two agents i and j, is of the form

$$\mathcal{W}_{ij}(x_i, x_j) = \frac{Q}{(\sigma^2 + \|x_i - x_j\|_2)^\beta} \tag{1}$$

where $Q, \sigma >$, and $\beta \geq 0$. In some networks, the weight of the link between two agents may not depend on the Euclidean distance of the agents and may depend on the *received data* [58]. A general state-dependent network has been considered in [59] where the weight $\mathcal{W}_{ij}(x_i, x_j)$ that agent i assigns to the information x_j received from agent j is a *general* function of x_i and x_j.

Therefore, a relevant question that arises is whether we are able to achieve distributed optimization over multi-agent networks whose interactions depend on their states? The above considerations motivate the needs for studying distributed optimization in light of *state-dependent* interactions that we formulate in this paper.

In a vehicular platoon, driver's attention is not continuous but can be modeled as *switching*, while the driver changes his/her attention as closer the cars in front and behind get. Mathematically, vehicular platoons are modeled as distance-dependent networks in [66], while with this realistic scenario we require to consider *time-varying* links as well. This motivates also to consider time-varying networks *combined* with state-dependent networks in our distributed optimization problem.

In our consideration, we have two networks: 1) the first is the network induced by the *state weights*; 2) the second network on top of the state-dependent network is a *time-varying* network; the time-varying network identifies the neighbors whose state-dependent weights are used to compute. Precisely, we formulate the combined *state-dependent* & *time-varying* network as follows.

A network of $m \in \mathbb{N}$ nodes labeled by the set $\mathcal{V} = \{1, 2, \ldots, m\}$ is considered. The topology of the network is represented by $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$ at time $t \in \mathbb{N} \cup \{0\}$ with the ordered edge set $\mathcal{E}_t \subseteq \mathcal{V} \times \mathcal{V}$. Let consider the set $G = \{\mathcal{G}_t : t \in \mathbb{N} \cup \{0\}\}$. Since $m \in \mathbb{N}$, the cardinality of G, namely $|G| = \bar{N}$, is finite. We write $\mathcal{N}_i^{in}(\mathcal{G})/\mathcal{N}_i^{out}(\mathcal{G})$ for the labels of agent i's in/out neighbors at graph $\mathcal{G} \in G$ so that there is an arc in $\mathcal{G}$ from vertex j/i to vertex i/j only if agent i receives/sends information from/to agent j. We write $\mathcal{N}_i(\mathcal{G})$ when $\mathcal{N}_i^{in}(\mathcal{G}) = \mathcal{N}_i^{out}(\mathcal{G})$. In this formulation, the *in* and *out* neighbors of each agent $i \in \mathcal{V}$ at each graph $\mathcal{G} \in G$ are *fixed*, while we will consider that the weights of links are possibly *state-dependent*. We assume that there is no self-looped arc in the communication graph. For better explanation of our formulation, consider a network of social agents in which only people who are friends communicate with each other where their decisions depend on difference among opinions (as stated above), while they pay attention to their friends at different *time*. We need to mention that in our formulation, the state-dependent weight $\mathcal{W}_{ij}(x_i, x_j)$ between any two agents i and j is general and may be a function of distance or received data.

In large part of literature on distributed consensus and optimization the graph topology (neighbors) are predefined. This makes sense in distributed computing systems where we are allowed to select the communication topology, as far as it is strongly connected.[4] In natural systems or also mobile robotics applications the neighbors are function of the distance or difference of opinions/states among nodes. Weights like Cuker-Smale allow to avoid mathematical complications of finite radius neighbors while providing effective but implicit notion of neighbors. So the effective graph topology of interactions changes with the state of the nodes. Finally with the addition of arbitrarily switching links (as far as the union of the sub-graphs is strongly connected and occur infinitely often), it is possible to model and analyze practical situations where effective neighbors are arbitrarily sampled.

Now, for each node $i \in \mathcal{V}$, we associate a convex cost function $f_i : \Re^n \longrightarrow \Re$ which is merely observed by node i. The objective of each agent is to seek the solution of the following optimization problem:

$$\min_s \sum_{i=1}^m f_i(s)$$

where $s \in \Re^n$. Since each node i knows only its own f_i, the nodes cannot individually compute the optimal solution and, therefore, must collaborate to do so. We assume that there are no communication delay and no communication noise in delivering a message from agent j to agent i.

The above problem can be formulated as

$$\begin{aligned} &\min_x \qquad f(x) := \sum_{i=1}^m f_i(x_i) \\ &\text{subject to} \quad x_1 = x_2 = \ldots = x_m \end{aligned} \tag{2}$$

where $x = [x_1^T, x_2^T, \ldots, x_m^T]^T, x_i \in \Re^n, i \in \mathcal{V}$, and the constraint set is achieved through state-dependent interactions and time-varying topologies. The set

$$\mathcal{C} := \{x \in \Re^{mn} | x_i = x_j, 1 \leq i, j \leq m, x_i \in \Re^n\} \tag{3}$$

is known as *consensus subspace* which is a convex set. Note that the Hilbert space considered in this paper is $X = (\Re^{mn}, \|.\|)$.

In this paper, we represent $W(x, \mathcal{G}) := \mathcal{W}(x, \mathcal{G}) \otimes I_n$ and $\mathcal{W}(x, \mathcal{G}) = [\mathcal{W}_{ij}(x_i, x_j, \mathcal{G})]$ for the *state-dependent* weighted matrix of the fixed graph $\mathcal{G} \in G$ in a *switching* network having all possible communication topologies in the set G. We represent $\mathcal{W}$, or W, to represent *state-independent* weighted matrix of the

[4]Random switching links are allowed under various assumptions, but the neighbors are predefined.

graph for non-switching networks. For example, if $\mathcal{W}(\hat{x}, \mathcal{G}_{\tilde{k}}) = I_m$, for some $\hat{x} \in \Re^{mn}$ and some $1 \leq \tilde{k} \leq \bar{N}$, implies that there are no edges in $\mathcal{G}_k$, or/and all nodes are not activated for communication updates in asynchronous protocol or both. Therefore, our *formulation* of distributed optimization problems here (and of course primarily in [61]) *unifies* switching networks and asynchrony in a better understandable formulation.

Now we impose Assumptions 1 and 2 below on $\mathcal{W}(x, \mathcal{G})$.

Assumption 1: For each fixed $\mathcal{G} \in G$, the weights $\mathcal{W}_{ij}(x_i, x_j, \mathcal{G}) : \Re^n \times \Re^n \longrightarrow [0, 1]$ are continuous, and the state-dependent weighted matrix of the graph is doubly stochastic, i.e.,

i) $\sum_{j \in \mathcal{N}_i^{in}(\mathcal{G}) \cup \{i\}} \mathcal{W}_{ij}(x_i, x_j, \mathcal{G}) = 1, i = 1, 2, \ldots, m,$

ii) $\sum_{j \in \mathcal{N}_i^{out}(\mathcal{G}) \cup \{i\}} \mathcal{W}_{ij}(x_i, x_j, \mathcal{G}) = 1, i = 1, 2, \ldots, m.$

Assumption 1 allows us to remove the couple of information exchange with the analysis of our proposed algorithm and to consider time-varying graphs as well (see Remark 5 below). The state-dependent weight $\mathcal{W}_{ij}(x_i, x_j)$ between any two agents i and j in Assumption 1 is general and may be a function of distance or received data.

Remark 1: Assumption 1 when applied to state-dependent weights would require undirected graphs, and directed graphs are allowed in the special case of state-independent weights. Note that any network with undirected links satisfies Assumption 1.

Assumption 2: The union of the graphs in G is strongly connected for all $x \in \Re^{mn}$, i.e.,

$$Re\left[\lambda_2\left(\sum_{\mathcal{G} \in G}(I_m - \mathcal{W}(x, \mathcal{G}))\right)\right] > 0, \quad \forall x \in \Re^{mn}. \tag{4}$$

Assumption 2 ensures that the union of graphs in the set G is strongly connected and that the information sent from each node is eventually obtained by every other node. Note that the statement of Assumption 2 is a necessary and sufficient condition (see Appendix A). Assumption 2 allows *state-dependent disconnectivity* of networks, namely $\mathcal{W}_{ij}(\tilde{x}_i, \tilde{x}_j, \tilde{\mathcal{G}}) = 0$ for some $\tilde{x}_i, \tilde{x}_j \in \Re^n$, and $\tilde{\mathcal{G}} \in G$, while (4) needs to be satisfied.

We show in Lemma 5 (in Appendix A) and Appendix B that the set

$$\mathcal{A} := \{x | W(x, \mathcal{G})x = x, \forall \mathcal{G} \in G, \text{with Assumptions 1 and 2}\}$$

is a convex set, and furthermore the set $\mathcal{C}$ defined in (3) can be achieved from the set $\mathcal{A}$, namely $\mathcal{A} = \mathcal{C}$. Consequently, under Assumptions 1 and 2, the constraint set of Problem (2) (i.e., $x_1 = \ldots = x_m$) is obtained by $\mathcal{A}$, which allows us to reformulate (2) as the following problem:

$$\begin{aligned} \min_x \quad & f(x) := \sum_{i=1}^m f_i(x_i) \\ \text{subject to} \quad & W(x, \mathcal{G})x = x \quad \forall \mathcal{G} \in G, \end{aligned} \tag{5}$$

so a solution of Problem (2) can be obtained by solving Problem (5) under Assumptions 1 and 2.

Now we introduce a new framework for modeling multi-agent optimization problems under *state-dependent* interactions and *time-varying* topologies. Problem stated by (5) can be reformulated as

$$\begin{aligned} \min_x \quad & f(x) := \sum_{i=1}^m f_i(x_i) \\ \text{subject to} \quad & x \in \bigcap_{\mathcal{G} \in G} Fix(T(x, \mathcal{G})) \end{aligned} \tag{6}$$

where $T(x, \mathcal{G}) := W(x, \mathcal{G})x$.

Definition 3 [61]: The operator $T(x) := Wx$ is called *weighted operator of the graph*.

Based on Definition 3, we give the following definition.

Definition 4: We call $T(x, \mathcal{G}) := W(x, \mathcal{G})x$ *state-dependent weighted operator of the graph* $\mathcal{G}$. If the graph is non-switching, we show $T(x) = W(x)x$.

Remark 2: The consensus subspace $\mathcal{C}$ (see (3)) is the intersection of fixed point sets of state-dependent weighted operator of the graphs with Assumptions 1 and 2 (see Appendices A and B for proofs), i.e.,

$$\mathcal{C} = \bigcap_{\mathcal{G} \in G} Fix(T(x, \mathcal{G})).$$

In the following two paragraphs, we explain why our problem *cannot* be formulated upon existing results on Fast Lipschitz optimization and convex minimization over fixed point set of nonexpansive mappings.

Recently, Fast Lipschitz Optimization has been introduced as a powerful method to capture the unique solution of convex or non-convex optimization problems [67]–[69]. If $\mathcal{W}$ does not depend on the states, i.e., the case of state-independent weighted graphs, the condition $\|W\| < 1$ is not satisfied for our problem because $\|W\| = 1$; in fact, the condition makes the operator $T(x) := Wx$ a contraction, and the feasible set will be a unique point (see Lemma 3 and Appendix C) instead of the set $\mathcal{C}$ defined in (3); for instance, if $W = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$, then $\|W\| = 1$ which implies that the set $\{y \in \Re^2 | y = Wy\}$ is the consensus subspace, but if $W = \begin{pmatrix} 0.25 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$, then $\|W\| < 1$ which implies that $\{y \in \Re^2 | y = Wy\} = \{\begin{pmatrix} 0 \\ 0 \end{pmatrix}\}$ which is a unique point. Furthermore, $W(x, \mathcal{G})$, for some $\mathcal{G} \in G$ in this paper, may not be differentiable. Therefore, the results given in [67]–[69] cannot not be applied here.

Convex minimization over fixed point set of a nonexpansive mapping has been studied in [70] and references therein. This method has been usefully applied to signal processing, inverse problem, network bandwidth allocation and so on [71]–[75]. If W does not depend on the states, the operator $T(x) := Wx$ is a nonexpansive mapping with Assumption 1 (see Appendix C). However, for state-dependent networks, the operator $T(x) := W(x)x$ may not be a nonexpansive mapping. Also, to implement the algorithms given by [70] and references therein, the underlying communication graph should be static (non-switching), and weighted matrix of the graph should be state-independent.

Therefore, the results given in [70] and references therein cannot be applied here. In [76], an incremental fixed point optimization algorithm is proposed for distributed optimization. In the algorithm, the graph topology among m agents is a B-connected directed cycle graph. In the algorithms given in [77], one agent is connected to other agents with undirected links. In [78], the graph topology among agents is an undirected cycle graph. In [76]–[78], each agent gets the exact state of each neighbor at each iteration, that is a very restrictive consideration. Therefore, the results given in [76]–[78] cannot not be applied for the problem considered in this paper.

IV. Main Results

We would like to solve (2), but we may have agents that are social with *state-dependent* interactions. So we cannot just use the usual methods (see Surveys [1]–[4] and references therein). Indeed, we require to develop algorithms to have a form based on a local term and a social term. We will show in this section that Algorithm (7) below solves (2) under some assumptions.

Based on the work of [61], we propose the following distributed algorithm for solving the problem stated by (6):

$$\begin{aligned} x_{i,t+1} &= \alpha_t(x_{i,t} - \beta\nabla f_i(x_{i,t})) + (1-\alpha_t)((1-\eta)x_{i,t} \\ &\quad + \eta \sum_{j\in\mathcal{N}_i^{in}(\mathcal{G}_t)\cup\{i\}} \mathcal{W}_{ij}(x_{i,t}, x_{j,t}, \mathcal{G}_t)x_{j,t}), \quad i\in\mathcal{V}, \end{aligned} \tag{7}$$

where $\alpha_t \in [0,1], \forall t \in \mathbb{N}\cup\{0\}$, $\eta\in(0,1)$, and β is a parameter to be chosen. The compact form of Algorithm (7) is given in (8) below.

Novelty: While the proposed algorithm in [61] works for *state-independent* networks and *cannot* be directly applied for state-dependent networks, we show in this paper that a generalization of the algorithm in [61] can solve distributed optimization over *state-dependent* networks with *time-varying* topologies. The differences with the work of [61] are: 1) the general operator $T(x,\mathcal{G})$ in [61] is assumed to be nonexpansive, whereas for distributed optimization (6) the operator $T(x,\mathcal{G}) := W(x,\mathcal{G})x$ for each fixed $\mathcal{G}\in G$ may *not* be nonexpansive for state-dependent networks (see the last paragraph in the previous section where we showed that for a non-switching network, i.e., $T(x) = W(x)x$); 2) the weight of a link in [61] and other work such as [9], to cite a few, takes values in a set having finite cardinality, whereas here the weight takes values in a continuum set (see Assumption 1); 3) network connectivity is a priori unknown at each iteration and is only determined by the states of the agents at the iteration; and 4) the weight of each link, taking values in this continuum set, is determined *endogenously* by the states of the agents at each iteration so that no local property can be assumed or checked *a priori* at the iteration. In fact, *all possible network topologies and corresponding weights are a priori known at each iteration in [61], whereas in this paper not only the weights but also the network topology $\mathcal{G}_{\tilde{t}}\in G$ for some $\tilde{t}\in\mathbb{N}$ may change due to state-dependent weights, e.g., if $\mathcal{W}_{ij}(x_{i,\tilde{t}}, x_{j,\tilde{t}}, \mathcal{G}_{\tilde{t}}) = 0$ for some $x_{i,\tilde{t}}, x_{j,\tilde{t}}\in\Re^n$, then this implies disconnectivity of the link between agents i and j.*

Now we impose the following assumptions on local cost functions and network connectivity.

Assumption 3: $f_i(x_i)$ is ρ-strongly convex, and $\nabla f_i(x_i)$ is K-Lipschitz continuous.

Assumption 4: There exists a nonempty subset $\tilde{K}\subseteq G$ such that the union of all graphs in $\tilde{K}$ is strongly connected for all $x\in\Re^{mn}$, and each graph in $\tilde{K}$ occurs infinitely often.

Remark 3: Assumption 4 includes B-connectivity assumption as a special case.

Now we give the main theorem in this paper.

Theorem 1: Consider the problem stated by (6) with Assumptions 1-4. Suppose that $\beta\in(0,\frac{2}{K})$ and the sequence $\alpha_t\in[0,1], t\in\mathbb{N}\cup\{0\}$, satisfies

(a) $\lim_{t\longrightarrow\infty}\alpha_t = 0$,

(b) $\sum_{t=0}^{\infty}\alpha_t = \infty$.

Then the sequence generated by (7) or its compact form

$$\begin{aligned} x_{t+1} &= \alpha_t(x_t - \beta\nabla f(x_t)) + (1-\alpha_t)((1-\eta)x_t \\ &\quad + \eta W(x_t,\mathcal{G}_t)x_t), \end{aligned} \tag{8}$$

where $\eta\in(0,1)$, globally converges to the unique solution of the problem.

Note that an example of α_t satisfying *(a)* and *(b)* in Theorem 1 is $\alpha_t := \frac{1}{(1+t)^\zeta}$ where $\zeta\in(0,1]$. Before we give the proof of Theorem 1, we need to present some remarks on properties of Algorithm (8) and Lemma 4 below needed in the proof.

Remark 4: A key distinguishing feature of Algorithm (8), comparing with most existing algorithms, is the presence of α_n in the second term that results in convergence properties with Assumption 3 *independent* of a specific choice of weights and converges under *any* weight models satisfying Assumptions 1 and 2.

Remark 5: Assumption 1 with Lemma 1 implies $\|W(x,\mathcal{G})\|\leq 1, \mathcal{G}\in G$. Consequently, this property of weighted matrices of the graphs allows Algorithm (8) to work under time-varying communication graphs.

Remark 6: The synchronous discrete-time algorithms proposed in [56] can solve distributed consensus problems over nonswitching undirected graphs where the weight of a link between two agents is nonincreasing with respect to Euclidean distance between their states and the algorithms need a parameter to be chosen based on a global information. The synchronous discrete-time algorithm proposed in [58] can solve distributed consensus problems over time-varying graphs without B-connectivity assumption for a specific weight where $x_i\in\Re, i\in\mathcal{V}$. The totally asynchronous algorithm (8) is able to solve distributed consensus problems for weights satisfying Assumption 1 and $x_i\in\Re^n$, while it is restricted to diminishing step size. Continuous algorithms have been proposed in [56], [57], [59], [60] for solving distributed consensus problems over non-switching graphs with state-dependent interactions.

Lemma 4: Let

$$\hat{T}(x,\mathcal{G}) := (1-\eta)x + \eta T(x,\mathcal{G}), \quad \mathcal{G}\in G, x\in X, \tag{9}$$

where T is defined in (6), and $\eta\in(0,1]$. Then

(i) $Fix(T(x,\mathcal{G})) = Fix(\hat{T}(x,\mathcal{G}))$.

(ii) $\langle x-\hat{T}(x,\mathcal{G}), x-z\rangle \geq \frac{\eta}{2}\|x-T(x,\mathcal{G})\|^2, \forall z\in\mathcal{C}, \forall\mathcal{G}\in G$.

(iii) $\|\hat{T}(x,\mathcal{G})-z\| \leq \|x-z\|, \forall z\in\mathcal{C}, \forall x\in X, \forall\mathcal{G}\in G$.

Proof: (i)

Consider a $\hat{x}\in Fix(T(x,\mathcal{G}))$. Thus $\hat{x}=T(\hat{x},\mathcal{G})$. Hence

$$\hat{T}(\hat{x},\mathcal{G}) = (1-\eta)\hat{x}+\eta T(\hat{x},\mathcal{G}) = \hat{x},$$

which implies that $Fix(T(x,\mathcal{G}))\subseteq Fix(\hat{T}(x,\mathcal{G}))$. Conversely, consider a $\hat{x}\in Fix(\hat{T}(x,\mathcal{G}))$. Indeed, $\hat{x}=\hat{T}(\hat{x},\mathcal{G})$. Thus we have

$$\hat{x}=\hat{T}(\hat{x},\mathcal{G}) = (1-\eta)\hat{x}+\eta T(\hat{x},\mathcal{G}),$$

or $\hat{x}=T(\hat{x},\mathcal{G})$, which implies that $Fix(\hat{T}(x,\mathcal{G}))\subseteq Fix(T(x,\mathcal{G}))$. Therefore, we can conclude that $Fix(\hat{T}(x,\mathcal{G})) = Fix(T(x,\mathcal{G}))$. Thus the proof of *(i)* is complete.

(ii) Since $z\in\mathcal{C}$ and $W(x,\mathcal{G})$ is a stochastic matrix for all $\mathcal{G}\in G, x\in X$, we have $W(x,\mathcal{G})z=z$. Therefore, we obtain

$$\begin{aligned}\|T(x,\mathcal{G})-z\| &= \|W(x,\mathcal{G})x-W(x,\mathcal{G})z\|\\ &\leq \|W(x,\mathcal{G})\|\|x-z\|.\end{aligned}$$

Since by Assumption 1 $W(x,\mathcal{G})$ is doubly stochastic, we have from Lemma 1 that $\|W(x,\mathcal{G})\|\leq 1, \forall\mathcal{G}\in G$. Hence,

$$\|T(x,\mathcal{G})-z\| \leq \|W(x,\mathcal{G})\|\|x-z\| \leq \|x-z\| \tag{10}$$

or

$$\|T(x,\mathcal{G})-z\|^2 \leq \|x-z\|^2, \forall z\in\mathcal{C}, \forall\mathcal{G}\in G. \tag{11}$$

Also, using the fact that

$$\|u+v\|^2 = \|u\|^2+\|v\|^2+2\langle u,v\rangle, \quad \forall u,v\in X, \tag{12}$$

we obtain

$$\begin{aligned}\|T(x,\mathcal{G})-z\|^2 &= \|T(x,\mathcal{G})-x+x-z\|^2\\ &= \|T(x,\mathcal{G})-x\|^2+\|x-z\|^2\\ &\quad + 2\langle T(x,\mathcal{G})-x, x-z\rangle.\end{aligned} \tag{13}$$

Substituting (13) for (11) yields

$$2\langle x-T(x,\mathcal{G}), x-z\rangle \geq \|T(x,\mathcal{G})-x\|^2. \tag{14}$$

Substituting $\frac{x-\hat{T}(x,\mathcal{G})}{\eta}$ for $x-T(x,\mathcal{G})$ (see (9)) in the left hand side of the inequality (14) implies *(ii)*. Thus the proof of *(ii)* is complete.

(iii) We have from (10) and $z=(1-\eta)z+\eta z$ that

$$\begin{aligned}\|\hat{T}(x,\mathcal{G})-z\| &\overset{(9)}{=} \|(1-\eta)(x-z)+\eta(T(x,\mathcal{G})-z)\|\\ &\leq (1-\eta)\|x-z\|+\eta\|T(x,\mathcal{G})-z\|\\ &\overset{(10)}{\leq} (1-\eta)\|x-z\|+\eta\|x-z\|\\ &= \|x-z\|.\end{aligned}$$

Therefore, the proof of *(iii)* is complete.

Proof of Theorem 1:

We prove Theorem 1 in three steps:

Step 1: $\{x_t\}_{t=0}^{\infty}$ is bounded.

Step 2: $\{x_t\}_{t=0}^{\infty}$ converges to an element in the feasible set.

Step 3: $\{x_t\}_{t=0}^{\infty}$ converges to the optimal solution.

Now we prove each step with details.

Step 1: $\{x_t\}_{t=0}^{\infty}$ is bounded.

Algorithm (8) with $T(x,\mathcal{G})$ and $\hat{T}(x,\mathcal{G})$ defined in (6) and (9), respectively, can be written as

$$x_{t+1} = \alpha_t(x_t-\beta\nabla f(x_t))+(1-\alpha_t)\hat{T}(x_t,\mathcal{G}_t). \tag{15}$$

Since $f(x)$ is strongly convex and $\mathcal{C}$ is closed, the problem has a unique solution. Let x^* be the unique solution of the problem. We have $x^*=\alpha_t x^*+(1-\alpha_t)x^*$. Therefore, we have from (15) that

$$\begin{aligned}&\|x_{t+1}-x^*\|\\ &\overset{(15)}{=} \|\alpha_t(x_t-\beta\nabla f(x_t))+(1-\alpha_t)\hat{T}(x_t,\mathcal{G}_t)-x^*\|\\ &= \|\alpha_t(x_t-\beta\nabla f(x_t)-x^*)+(1-\alpha_t)(\hat{T}(x_t,\mathcal{G}_t)-x^*)\|.\end{aligned} \tag{16}$$

We obtain from part *(iii)* of Lemma 4 that

$$\begin{aligned}&\|x_{t+1}-x^*\| \qquad (17)\\ &\overset{(16)}{=} \|\alpha_t(x_t-\beta\nabla f(x_t)-x^*)+(1-\alpha_t)(\hat{T}(x_t,\mathcal{G}_t)-x^*)\|\\ &\leq \alpha_t\|x_t-\beta\nabla f(x_t)-x^*\|+(1-\alpha_t)\|\hat{T}(x_t,\mathcal{G}_t)-x^*\|\\ &\overset{(iii)}{\leq} \alpha_t\|x_t-\beta\nabla f(x_t)-x^*\|+(1-\alpha_t)\|x_t-x^*\|.\end{aligned} \tag{18}$$

If $\nabla f_i(x_i)$ is K-Lipschitz, then $f_i(x_i)$ is K-strongly smooth by Lemma 3.4 in [79]. Since $f_i(x_i)$ is ρ-strongly convex and K-strongly smooth, the operator $H(x):=x-\beta\nabla f(x)$ where $\beta\in(0,\frac{2}{K})$ is a contraction[5] (see [80] for details). As a matter of fact, there exists a $0<\gamma\leq 1$ such that

$$\|x-y-\beta(\nabla f(x)-\nabla f(y))\| \leq (1-\gamma)\|x-y\|, \forall x,y\in\Re^{mn}.$$

Consequently, we have

$$\|x_t-x^*-\beta(\nabla f(x_t)-\nabla f(x^*))\| \leq (1-\gamma)\|x_t-x^*\|. \tag{19}$$

Now we have that

$$\begin{aligned}&\|x_t-\beta\nabla f(x_t)-x^*\|\\ &= \|x_t-x^*-\beta(\nabla f(x_t)-\nabla f(x^*))-\beta\nabla f(x^*)\|\\ &\leq \|x_t-x^*-\beta(\nabla f(x_t)-\nabla f(x^*))\|+\beta\|\nabla f(x^*)\|\\ &\overset{(19)}{\leq} (1-\gamma)\|x_t-x^*\|+\beta\|\nabla f(x^*)\|.\end{aligned} \tag{20}$$

Substituting (20) for (18) yields

$$\begin{aligned}\|x_{t+1}-x^*\| &\leq (1-\gamma\alpha_t)\|x_t-x^*\|+\alpha_t\beta\|\nabla f(x^*)\|\\ &= (1-\gamma\alpha_t)\|x_t-x^*\|+\gamma\alpha_t\left(\frac{\beta\|\nabla f(x^*)\|}{\gamma}\right)\end{aligned}$$

[5]The range of β in Theorems 1 and 2 and Corollaries 1 and 2 in [61] can be enlarged from $\beta\in(0,\frac{2\rho}{K^2})$ to $\beta\in(0,\frac{2}{K})$. The proof is similar to that of *Step 1* in this paper and is omitted.

which by induction implies

$$\|x_{t+1} - x^*\| \leq max \left\{ \|x_0 - x^*\|, \frac{\beta\|\nabla f(x^*)\|}{\gamma} \right\}.$$

Thus $\{x_t\}_{t=0}^{\infty}$ is bounded. Consequently, due to Lipschitz continuity of $\nabla f(x)$ (see Assumption 3), the sequence $\{\nabla f(x_t)\}_{t=0}^{\infty}$ is bounded, too. Since $\{x_t\}_{t=0}^{\infty}$ is bounded, we obtain from part *(iii)* of Lemma 4 that the sequence $\{\hat{T}(x_t, \mathcal{G}_t)\}_{t=0}^{\infty}$ is bounded. Since both $\{x_t\}_{t=0}^{\infty}$ and $\{\hat{T}(x_t, \mathcal{G}_t)\}_{t=0}^{\infty}$ are bounded, we obtain from (9) that the sequence $\{T(x_t, \mathcal{G}_t)\}_{t=0}^{\infty}$ is bounded, too.

Step 2: $\{x_t\}_{t=0}^{\infty}$ converges to an element in the feasible set.

From (15) and $x_t = \alpha_t x_t + (1-\alpha_t)x_t$, we have

$$x_{t+1} - x_t + \alpha_t\beta\nabla f(x_t) = (1-\alpha_t)(\hat{T}(x_t, \mathcal{G}_t) - x_t). \quad (21)$$

Hence, we obtain

$$\begin{aligned}&\langle x_{t+1} - x_t + \alpha_t\beta\nabla f(x_t), x_t - x^*\rangle \\ &\overset{(21)}{=} -(1-\alpha_t)\langle x_t - \hat{T}(x_t, \mathcal{G}_t), x_t - x^*\rangle. \end{aligned} \quad (22)$$

Since $x^* \in \mathcal{C}$, we have from part *(ii)* of Lemma 4 that

$$\langle x_t - \hat{T}(x_t, \mathcal{G}_t), x_t - x^*\rangle \geq \frac{\eta}{2}\|x_t - T(x_t, \mathcal{G}_t)\|^2. \quad (23)$$

From (22) and (23), we obtain

$$\begin{aligned}&\langle x_{t+1} - x_t + \alpha_t\beta\nabla f(x_t), x_t - x^*\rangle \\ &\leq -\frac{\eta}{2}(1-\alpha_t)\|x_t - T(x_t, \mathcal{G}_t)\|^2 \end{aligned} \quad (24)$$

or equivalently

$$\begin{aligned}&-\langle x_t - x_{t+1}, x_t - x^*\rangle \leq -\alpha_t\langle\beta\nabla f(x_t), x_t - x^*\rangle \\ &-\frac{\eta}{2}(1-\alpha_t)\|x_t - T(x_t, \mathcal{G}_t)\|^2. \end{aligned} \quad (25)$$

For any $u, v \in X$ we have

$$\langle u, v\rangle = -\frac{1}{2}\|u-v\|^2 + \frac{1}{2}\|u\|^2 + \frac{1}{2}\|v\|^2. \quad (26)$$

From (26) we obtain

$$\langle x_t - x_{t+1}, x_t - x^*\rangle = -C_{t+1} + C_t + \frac{1}{2}\|x_t - x_{t+1}\|^2, \quad (27)$$

where $C_t := \frac{1}{2}\|x_t - x^*\|^2$. From (25) and (27) we obtain

$$\begin{aligned}&C_{t+1} - C_t - \frac{1}{2}\|x_t - x_{t+1}\|^2 \leq -\alpha_t\langle\beta\nabla f(x_t), x_t - x^*\rangle \\ &-\frac{\eta}{2}(1-\alpha_n)\|x_t - T(x_t, \mathcal{G}_t)\|^2. \end{aligned} \quad (28)$$

From (21) and (12) we have

$$\begin{aligned}&\|x_{t+1} - x_t\|^2 \\ &\overset{(21)}{=} \|-\alpha_t\beta\nabla f(x_t) + (1-\alpha_t)(\hat{T}(x_t, \mathcal{G}_t) - x_t)\|^2 \\ &\overset{(12)}{=} \alpha_t^2\|\beta\nabla f(x_t)\|^2 \\ &+ (1-\alpha_t)^2\|\hat{T}(x_t, \mathcal{G}_t) - x_t\|^2 \\ &- 2\alpha_t(1-\alpha_t)\langle\beta\nabla f(x_t), \hat{T}(\mathcal{G}_t, x_t) - x_t\rangle. \end{aligned} \quad (29)$$

We have from (9) that

$$\|\hat{T}(x_t, \mathcal{G}_t) - x_t\| = \eta\|x_t - T(x_t, \mathcal{G}_t)\|. \quad (30)$$

Since $\alpha_t \in [0, 1]$, we have also that

$$(1-\alpha_t)^2 \leq (1-\alpha_t). \quad (31)$$

Substituting (30) for (29) and using (31) yield

$$\begin{aligned}&\frac{1}{2}\|x_{t+1} - x_t\|^2 \\ &\overset{(29)-(30)}{=} \frac{1}{2}\alpha_t^2\|\beta\nabla f(x_t)\|^2 \\ &+ \frac{1}{2}(1-\alpha_t)^2\eta^2\|T(x_t, \mathcal{G}_t) - x_t\|^2 \\ &- \alpha_t(1-\alpha_t)\langle\beta\nabla f(x_t), \hat{T}(x_t, \mathcal{G}_t) - x_t\rangle \\ &\overset{(31)}{\leq} \frac{1}{2}\alpha_t^2\|\beta\nabla f(x_t)\|^2 \\ &+ \frac{1}{2}(1-\alpha_t)\eta^2\|T(x_t, \mathcal{G}_t) - x_t\|^2 \\ &- \alpha_t(1-\alpha_t)\langle\beta\nabla f(x_t), \hat{T}(x_t, \mathcal{G}_t) - x_t\rangle. \end{aligned} \quad (32)$$

From (28) and (32), we obtain

$$\begin{aligned}C_{t+1} - C_t &\overset{(28)}{\leq} \frac{1}{2}\|x_{t+1} - x_t\|^2 \\ &- \alpha_t\langle\beta\nabla f(x_t), x_t - x^*\rangle \\ &- \frac{\eta}{2}(1-\alpha_t)\|x_t - T(x_t, \mathcal{G}_t)\|^2 \\ &\overset{(32)}{\leq} -\left(\frac{1}{2} - \frac{\eta}{2}\right)\eta(1-\alpha_t)\|x_t - T(x_t, \mathcal{G}_t)\|^2 \\ &+ \alpha_t\left(\frac{1}{2}\alpha_t\|\beta\nabla f(x_t)\|^2\right. \\ &- \langle\beta\nabla f(x_t), x_t - x^*\rangle \\ &\left.-(1-\alpha_t)\langle\beta\nabla f(x_t), \hat{T}(x_t, \mathcal{G}_t) - x_t\rangle\right). \end{aligned} \quad (33)$$

Now we claim that there exists a $t_0 \in \mathbb{N}$ such that the sequence $\{C_t\}_{t=0}^{\infty}$ is non-increasing for $t \geq t_0$. Assume by contradiction that this is not true. Then there exists a subsequence $\{C_{t_j}\}_{j=0}^{\infty}$ such that

$$C_{t_j+1} - C_{t_j} > 0$$

which together with (33) yields

$$\begin{aligned}&0 < C_{t_j+1} - C_{t_j} \\ &\overset{(33)}{\leq} -\left(\frac{1}{2} - \frac{\eta}{2}\right)\eta(1-\alpha_{t_j})\|x_{t_j} - T(x_{t_j}, \mathcal{G}_{t_j})\|^2 \\ &+ \alpha_{t_j}\left(\frac{1}{2}\alpha_{t_j}\beta^2\|\nabla f(x_{t_j})\|^2 - \langle\beta\nabla f(x_{t_j}), x_{t_j} - x^*\rangle\right. \\ &\left.-(1-\alpha_{t_j})\langle\beta\nabla f(x_{t_j}), \hat{T}(x_{t_j}, \mathcal{G}_{t_j}) - x_{t_j}\rangle\right). \end{aligned} \quad (34)$$

Since $\{x_t\}_{t=0}^{\infty}$, $\{\nabla f(x_t)\}_{t=0}^{\infty}$, $\{T(x_t, \mathcal{G}_t)\}_{t=0}^{\infty}$, and $\{\hat{T}(x_t, \mathcal{G}_t)\}_{t=0}^{\infty}$ are bounded (see *Step 1*), we obtain from

(34), where $\eta \in (0,1)$, by Theorem 1 *(a)* that

$$0 < \liminf_{j \longrightarrow \infty} \left(-\left(\frac{1}{2} - \frac{\eta}{2}\right) \eta(1-\alpha_{t_j}) \|x_{t_j} - T(x_{t_j}, \mathcal{G}_{t_j})\|^2 + \alpha_{t_j} \left(\frac{1}{2}\alpha_{t_j} \|\beta \nabla f(x_{t_j})\|^2 - \langle \beta \nabla f(x_{t_j}), x_{t_j} - x^* \rangle - (1-\alpha_{t_j}) \langle \beta \nabla f(x_{t_j}), \hat{T}(\omega^*_{t_j}, x_{t_j}) - x_{t_j} \rangle \right) \right) \leq 0 \tag{35}$$

which is a contradiction. Therefore, there exists a $t_0 \in \mathbb{N}$ such that the sequence $\{C_t\}$ is non-increasing for $t \geq t_0$. Since $\{C_t\}$ is bounded below, it converges.

Taking the limit of both sides of (33) and using the convergence of $\{C_t\}$, boundedness of $\{x_t\}_{t=0}^{\infty}, \{\nabla f(x_t)\}_{t=0}^{\infty}, \{T(x_t, \mathcal{G}_t)\}_{t=0}^{\infty}, \{\hat{T}(x_t, \mathcal{G}_t)\}_{t=0}^{\infty}, \eta \in (0,1)$, and Theorem 1 *(a)* yield

$$\lim_{t \longrightarrow \infty} \|x_t - T(x_t, \mathcal{G}_t)\| = 0 \tag{36}$$

or from (30) yields

$$\lim_{t \longrightarrow \infty} \|x_t - \hat{T}(x_t, \mathcal{G}_t)\| = 0. \tag{37}$$

We have from (15) and *Step 1* that

$$\lim_{n \longrightarrow \infty} \|x_{n+1} - \hat{T}(x_n, \mathcal{G}_n)\| = \lim_{n \longrightarrow \infty} \alpha_n \|x_n - \beta \nabla f(x_n) - \hat{T}(x_n, \mathcal{G}_n)\| = 0. \tag{38}$$

(37) and (38) together implies

$$\lim_{n \longrightarrow \infty} \|x_{n+1} - x_n\| \leq \lim_{n \longrightarrow \infty} \|x_{n+1} - \hat{T}(x_n, \mathcal{G}_n)\| + \lim_{n \longrightarrow \infty} \|x_n - \hat{T}(x_n, \mathcal{G}_n)\| = 0$$

which together with *Step 1* implies that the sequence $\{x_n\}_{n=0}^{\infty}$ is convergent. From this fact, we obtain by Assumption 4, part *(i)* of Lemma 4, and (37) that $\{x_n\}_{n=0}^{\infty}$ converges to an element in the feasible set.

Comment 1: One can also conclude from (36) and Assumption 4 that $\{x_n\}_{n=0}^{\infty}$ converges to an element in the feasible set.

Step 3: $\{x_t\}_{t=0}^{\infty}$ converges to the optimal solution.

It remains to prove that $\lim_{t \longrightarrow \infty} \|x_t - x^*\| = 0$. From *Step 2*, we have that $\{x_t\}$ converges to $\bar{x} \in \mathcal{C}$. Since $x^* \in \mathcal{C}$ is the optimal solution, we have from optimality condition for convex optimization that

$$\langle \bar{x} - x^*, \nabla f(x^*) \rangle \geq 0, \forall \bar{x} \in \mathcal{C}. \tag{39}$$

We have from (12) that

$$\begin{aligned} &\|x_{t+1} - x^*\|^2 \\ &= \|x_{t+1} - x^* + \alpha_t \beta \nabla f(x^*) - \alpha_t \beta \nabla f(x^*)\|^2 \\ &\overset{(12)}{=} \|x_{t+1} - x^* + \alpha_t \beta \nabla f(x^*)\|^2 + \alpha_t^2 \|\beta \nabla f(x^*)\|^2 \\ &\quad - 2\alpha_t \langle \beta \nabla f(x^*), x_{t+1} - x^* + \alpha_t \beta \nabla f(x^*) \rangle. \end{aligned} \tag{40}$$

Using $x^* = \alpha_t x^* + (1-\alpha_t)x^*, \forall t \in \mathbb{N} \cup \{0\}$ and (15), we obtain

$$\begin{aligned} \|x_{t+1} - x^* + \alpha_t \beta \nabla f(x^*)\|^2 = \|\alpha_t [x_t - x^* \\ - \beta(\nabla f(x_t) - \nabla f(x^*))] \\ + (1-\alpha_t)[\hat{T}(x_t, \mathcal{G}_t) - x^*]\|^2. \end{aligned} \tag{41}$$

Moreover, we have

$$\begin{aligned} &\langle \beta \nabla f(x^*), x_{t+1} - x^* + \alpha_t \beta \nabla f(x^*) \rangle \\ &= \langle \beta \nabla f(x^*), x_{t+1} - x^* \rangle + \alpha_t \langle \beta \nabla f(x^*), \beta \nabla f(x^*) \rangle \\ &= \langle \beta \nabla f(x^*), x_{t+1} - x^* \rangle + \alpha_t \|\beta \nabla f(x^*)\|^2. \end{aligned} \tag{42}$$

Substituting (41) and (42) for (40) yields

$$\begin{aligned} &\|x_{t+1} - x^*\|^2 \\ &\overset{(40)}{=} \|x_{t+1} - x^* + \alpha_t \beta \nabla f(x^*)\|^2 \\ &\quad + \alpha_t^2 \|\beta \nabla f(x^*)\|^2 \\ &\quad - 2\alpha_t \langle \beta \nabla f(x^*), x_{t+1} - x^* + \alpha_t \beta \nabla f(x^*) \rangle \\ &\overset{(41)}{=} \|\alpha_t [x_t - x^* - \beta(\nabla f(x_t) - \nabla f(x^*))] \\ &\quad + (1-\alpha_t)[\hat{T}(x_t, \mathcal{G}_t) - x^*]\|^2 \\ &\quad + \alpha_t^2 \|\beta \nabla f(x^*)\|^2 \\ &\quad - 2\alpha_t \langle \beta \nabla f(x^*), x_{t+1} - x^* + \alpha_t \beta \nabla f(x^*) \rangle \\ &\overset{(42)}{=} \|\alpha_t [x_t - x^* - \beta(\nabla f(x_t) - \nabla f(x^*))] \\ &\quad + (1-\alpha_t)[\hat{T}(x_t, \mathcal{G}_t) - x^*]\|^2 \\ &\quad - 2\alpha_t \langle \beta \nabla f(x^*), x_{t+1} - x^* \rangle - \alpha_t^2 \|\beta \nabla f(x^*)\|^2 \\ &\overset{(12)}{=} \alpha_t^2 \|x_t - x^* - \beta(\nabla f(x_t) - \nabla f(x^*))\|^2 \\ &\quad + (1-\alpha_t)^2 \|\hat{T}(x_t, \mathcal{G}_t) - x^*\|^2 \\ &\quad + 2\alpha_t(1-\alpha_t) \langle x_t - x^* \\ &\quad - \beta(\nabla f(x_t) - \nabla f(x^*)), \hat{T}(x_t, \mathcal{G}_t) - x^* \rangle \\ &\quad - 2\alpha_t \langle \beta \nabla f(x^*), x_{t+1} - x^* \rangle - \alpha_t^2 \|\beta \nabla f(x^*)\|^2. \end{aligned} \tag{43}$$

From Cauchy–Schwarz inequality, part *(iii)* of Lemma 4, and (19), we obtain

$$\begin{aligned} &\langle x_t - x^* - \beta(\nabla f(x_t) - \nabla f(x^*)), \hat{T}(x_t, \mathcal{G}_t) - x^* \rangle \\ &\leq \|x_t - x^* - \beta(\nabla f(x_t) - \nabla f(x^*))\| \|\hat{T}(x_t, \mathcal{G}_t) - x^*\| \\ &\overset{(iii)}{\leq} \|x_t - x^* - \beta(\nabla f(x_t) - \nabla f(x^*))\| \|x_t - x^*\| \\ &\overset{(19)}{\leq} (1-\gamma)\|x_t - x^*\|^2. \end{aligned} \tag{44}$$

From (19), we also obtain

$$\|x_t - x^* - \beta(\nabla f(x_t) - \nabla f(x^*))\|^2 \leq (1-\gamma)^2 \|x_t - x^*\|^2. \tag{45}$$

Therefore, from part *(iii)* of Lemma 4, (44), and (45), we have

$$\begin{aligned}
&\|x_{t+1}-x^*\|^2\\
&\overset{(43)}{=} \alpha_t^2\|x_t-x^*-\beta(\nabla f(x_t)-\nabla f(x^*))\|^2\\
&\quad+(1-\alpha_t)^2\|\hat{T}(x_t,\mathcal{G}_t)-x^*\|^2\\
&\quad+2\alpha_t(1-\alpha_t)\langle x_t-x^*-\beta(\nabla f(x_t)-\nabla f(x^*)),\\
&\quad\hat{T}(x_t,\mathcal{G}_t)-x^*\rangle\\
&\quad-2\alpha_t\langle\beta\nabla f(x^*),x_{t+1}-x^*\rangle-\alpha_t^2\|\beta\nabla f(x^*)\|^2\\
&\overset{(iii)}{\le} \alpha_t^2\|x_t-x^*-\beta(\nabla f(x_t)-\nabla f(x^*))\|^2\\
&\quad+(1-\alpha_t)^2\|x_t-x^*\|^2\\
&\quad+2\alpha_t(1-\alpha_t)\langle x_t-x^*-\beta(\nabla f(x_t)-\nabla f(x^*)),\\
&\quad\hat{T}(x_t,\mathcal{G}_t)-x^*\rangle\\
&\quad-2\alpha_t\langle\beta\nabla f(x^*),x_{t+1}-x^*\rangle-\alpha_t^2\|\beta\nabla f(x^*)\|^2\\
&\overset{(44)-(45)}{\le} (1-2\gamma\alpha_t)\|x_t-x^*\|^2\\
&\quad+\alpha_t(\gamma^2\alpha_t\|x_t-x^*\|^2-2\langle\beta\nabla f(x^*),x_{t+1}-x^*\rangle)\\
&=(1-\gamma\alpha_t)\|x_t-x^*\|^2-\gamma\alpha_t\|x_t-x^*\|^2\\
&\quad+\alpha_t(\gamma^2\alpha_t\|x_t-x^*\|^2-2\langle\beta\nabla f(x^*),x_{t+1}-x^*\rangle). \qquad (46)
\end{aligned}$$

Since $\gamma\alpha_t\|x_t-x^*\|^2\geq 0$, we have

$$\begin{aligned}
&(1-\gamma\alpha_t)\|x_t-x^*\|^2\\
&\quad-\gamma\alpha_t\|x_t-x^*\|^2\\
&\quad+\alpha_t(\gamma^2\alpha_t\|x_t-x^*\|^2-2\langle\beta\nabla f(x^*),x_{t+1}-x^*\rangle)\\
&\le(1-\gamma\alpha_t)\|x_t-x^*\|^2\\
&\quad+\alpha_t(\gamma^2\alpha_t\|x_t-x^*\|^2-2\langle\beta\nabla f(x^*),x_{t+1}-x^*\rangle). \qquad (47)
\end{aligned}$$

Substituting (47) for (46) yields

$$\begin{aligned}
&\|x_{t+1}-x^*\|^2\le(1-\gamma\alpha_t)\|x_t-x^*\|^2\\
&\quad+\gamma\alpha_t\left(\frac{\gamma^2\alpha_t\|x_t-x^*\|^2-2\langle\beta\nabla f(x^*),x_{t+1}-x^*\rangle}{\gamma}\right).
\end{aligned} \qquad (48)$$

From *Step 1*, *Step 2*, (39), and Theorem 1 *(a)*, we obtain

$$\lim_{t\longrightarrow\infty}(\gamma^2\alpha_t\|x_t-x^*\|^2-2\beta\langle\nabla f(x^*),x_{t+1}-x^*\rangle)\le 0. \quad (49)$$

According to Lemma 2 by setting

$$\begin{aligned}
a_t&=\|x_t-x^*\|^2,\\
b_t&=\gamma\alpha_t,\\
h_t&=\left(\frac{\gamma^2\alpha_t\|x_t-x^*\|^2-2\beta\langle\nabla f(x^*),x_{t+1}-x^*\rangle}{\gamma}\right),
\end{aligned}$$

we obtain from (48), (49), and Theorem 1 *(b)* that

$$\lim_{t\longrightarrow\infty}\|x_t-x^*\|^2=0.$$

Therefore, $\{x_t\}_{t=0}^{\infty}$ converges to x^* as $t\longrightarrow\infty$. Thus the proof of Theorem 1 is complete.

Remark 7: Algorithm (8), which is based on the algorithm proposed in [61], is a totally asynchronous algorithm for distributed optimization over *state-dependent* networks. We refer an interested reader for totally asynchronous algorithms *without* diminishing step size over *state-independent* networks to [81] for distributed consensus and to [82], [83] for linear algebraic equations (see [84] for more details).

Remark 8: Since Algorithm (8) is totally asynchronous, convergence rate of the algorithm in general cannot be established. Nevertheless, determining convergence rate based on suitable assumptions remains for future work.

V. Numerical Example

We give a practical example of distributed optimization with state-dependent interactions of Cucker-Smale form [38] and time-varying topologies.

Example 1: (Distributed Optimization for an Automated Warehouse):

Consider a warehouse which has m robots on the shop floor. In warehousing environment, assigning tasks to robotic agents can be modeled as optimization problems [85]. Typically, such optimizations are solved at a central entity, that are neither scalable nor can handle autonomous entities [85]. As it is a large warehouse, no centralized control is possible, and the robots must handle tasks in collaborative manner [85]. Since, nowadays, the interactions among robots are usually performed via a wireless network, the signal power at a receiver is inversely proportional to the some power of the distance between transmitter and receiver [86]. Hence, the weight of links between robots are state-dependent, considering the state as a position.

Now, consider $m=20$ robots on the shop floor of a warehouse. They need to bring some loads from different initial places to a place for delivery. To minimize the cost, we desire robots to put the loads in a place whose sum of squared distances to the initial places is minimized, i.e., we have the following optimization:

$$\min_s \sum_{i=1}^{20}\|s-d_i\|^2 \qquad (50)$$

where $s\in\Re^2$ is the decision variable, and d_i is the position of the initial place of the load i on the two-dimensional shop floor. This problem can be reformulated as the following distributed optimization problem:

$$\begin{aligned}
&\min_x \qquad f(x):=\sum_{i=1}^{20}0.5\|x_i-d_i\|^2\\
&\text{subject to}\quad x_1=x_2=\ldots=x_{20}
\end{aligned}$$

where $x_i=[x_i^1,x_i^2]^T$, and the constraint set is achieved via *distance-dependent* network with *time-varying* topologies.

We consider the initial position of agent i as $x_{i,0}=[10cos(\frac{(i-1)2\pi}{22}),10sin(\frac{(i-1)2\pi}{22})]^T$. Consequently, based on nearest neighbor behavioral approach of initial positions of

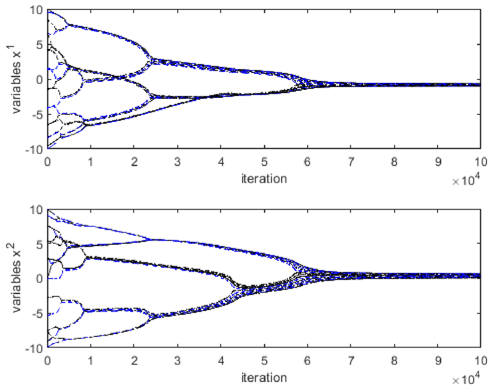


Fig. 1. Variables x_i^1 and $x_i^1, x_i^2, i = 1, \ldots, 20$, of the robotic agents. This figure shows that the variables are getting consensus when the robots communicate over the *switching* network.

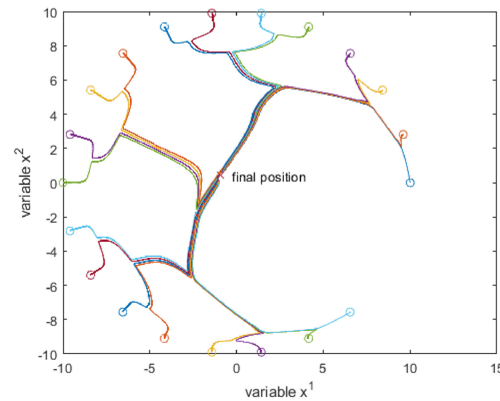


Fig. 2. The 2D plot. Initial positions and the final position are shown with 'o' and 'x,' respectively.

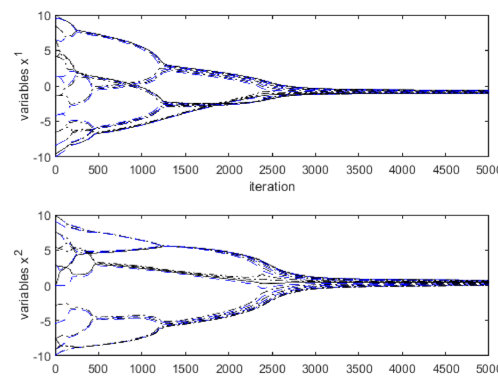


Fig. 3. Variables x_i^1 and $x_i^1, x_i^2, i = 1, \ldots, 20$, of the robotic agents. This figure shows that the variables are getting consensus when the robots communicate over the *non-switching* network.

robots, the topology of the underlying undirected graph is assumed to be a line graph, i.e., $1 \longleftrightarrow 2 \ldots \longleftrightarrow 20$. Hence, each agent only selects its neighbors based on its initial position and interacts only with them in the future. The weight of the link between robots i and j, which is assumed to depend inversely on the Euclidean distance of their positions, is considered of Cucker-Smale form [38]

$$\mathcal{W}_{ij}(x_i, x_j) = \frac{0.25}{1 + \|x_i - x_j\|^2}. \tag{51}$$

It is easy to see that the weight of a link determined by (51) takes values in a continuum set, and the weight of the link at each iteration in Algorithm (7) is only determined by the states of the agents for which no local property can be a priori assumed or determined in the future. One can see that $f_i(x_i) := 0.5\|x_i - d_i\|^2, i = 1, 2, \ldots, 20$, are 1-strongly convex, and $\nabla f_i(x_i)$ are 1-Lipschitz continuous.

To show that Algorithm (8) works under minimal connectivity conditions at each iteration, we consider that each link is activated periodically and works individually at each iteration. Thus, the union of the graphs which occur infinitely often is strongly connected for all $x \in \Re^{40}$. Therefore, Assumption 4 is fulfilled. Thus the conditions of Theorem 1 are satisfied.

We use $\eta = 0.7, \alpha_t = \frac{1}{1+t}, t \geq 0, \beta = \frac{1}{K} = \frac{1}{1}$, and randomly selected initial conditions for simulation. The results given by Algorithm (8) are shown in Figure 1. The two-dimensional (2D) plot is shown in Figure 2. Figures 1-2 show that the positions of robotic agents are approaching the solution of the optimization (50). We simulate this example when the underlying graph is non-switching, and the result is shown in Figure 3.

Comment 2: Now consider the non-switching directed graph $1 \longleftarrow 2 \ldots \longleftarrow 20 \longleftarrow 1$ where the state-independent weight of each link is assumed to be one. Consequently, the corresponding weighted matrix of the graph is *periodic and irreducible with zero diagonals*. One can simulate and see that Algorithm (8) still works in this case.

VI. Conclusion

In this paper, an unconstrained distributed convex optimization problem over graphs with state-dependent interactions and time-varying topologies is considered. A framework is proposed for modeling the problem in presence of state-dependent communications and time-varying topologies. A gradient-based discrete-time algorithm using diminishing step size is given for convergence to the unique solution of the problem under suitable assumptions. The proposed algorithm is also able to converge to average consensus of initial states of agents in presence of state-dependent communication and time-varying topologies. The algorithm is totally asynchronous without requiring B-connectivity assumption for convergence. The algorithm works even if the weighted matrix of the graph is periodic and irreducible for synchronous protocol. Finally, a case study on distributed optimization in an automated warehouse is given in order to show the results. Determining convergence rate of the algorithm based on suitable assumptions remains for future work.

Appendix A

We show in the following lemma that the statement of Assumption 2 is a necessary and sufficient condition.

Lemma 5: The union of all of the graphs in G is strongly connected for all $x \in \Re^{mn}$ if and only if $Re[\lambda_2(\sum_{\mathcal{G}\in G}(I_m - \mathcal{W}(x,\mathcal{G})))] > 0$ for all $x \in \Re^{mn}$.

Proof: If the union of all of the graphs is strongly connected for all $x \in \Re^{mn}$, the matrix $\sum_{\mathcal{G}\in G}\mathcal{W}(x,\mathcal{G})$ is irreducible for all $x \in \Re^{mn}$. Therefore, according to Perron-Frobenius theorem for irreducible matrices, it has a unique positive real largest eigenvalue for each $x \in \Re^{mn}$. Since, by Assumption 1, $\mathcal{W}(x,\mathcal{G}), \forall \mathcal{G} \in G, \forall x \in X$, is doubly stochastic, the unique largest eigenvalue of the matrix $\sum_{\mathcal{G}\in G}\mathcal{W}(x,\mathcal{G})$ is $\lambda^*(x) = \bar{N}$. Thus $Re[\lambda_2(\sum_{\mathcal{G}\in G}(I_m - \mathcal{W}(x,\mathcal{G})))] > 0$. Now we prove the opposite direction. We prove it by contradiction. Assume that the union of all of the graphs in G is not strongly connected for all $x \in \Re^{mn}$. It is well-known that there exists a permutation matrix P such that for some $\tilde{x} \in \Re^{mn}$

$$P\left(\sum_{\mathcal{G}\in G}\mathcal{W}(\tilde{x},\mathcal{G})\right)P^T = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix},$$

where A and C are square matrices. Therefore, $spec(\sum_{\mathcal{G}\in G}\mathcal{W}(\tilde{x},\mathcal{G})) = spec(A) \cup spec(C)$. From Assumption 1, all columns of A has summation equal to $\bar{N}$, and all rows of C has summation equal to $\bar{N}$. Therefore, the eigenvalue $\lambda(\tilde{x}) = \bar{N}$ has multiplicity 2 in $spec(\sum_{\mathcal{G}\in G}\mathcal{W}(\tilde{x},\mathcal{G}))$ which is a contradiction. Thus the proof is complete.

Appendix B

We show that $\mathcal{A} = \mathcal{C}$. We obtain from $W(x,\mathcal{G})x = x, \forall \mathcal{G} \in G$, that

$$(I_{mn} - W(x,\mathcal{G}))x = 0, \forall \mathcal{G} \in G,$$

which implies that

$$\sum_{\mathcal{G}\in G}(I_{mn} - W(x,\mathcal{G}))x = 0. \tag{52}$$

Now we have

$$\begin{aligned}\sum_{\mathcal{G}\in G}(I_{mn} - W(x,\mathcal{G})) &= \sum_{\mathcal{G}\in G}((I_m - \mathcal{W}(x,\mathcal{G})) \otimes I_n) \\ &= \Lambda(x) \otimes I_n,\end{aligned}$$

where

$$\Lambda(x) := \sum_{\mathcal{G}\in G}(I_m - \mathcal{W}(x,\mathcal{G})).$$

$\Lambda(x)$ has the following properties: the summation of all rows are equal to zero; the diagonal elements are non-negative; the off-diagonal elements are non-positive. Therefore, $\Lambda(x)$ has the Laplacian matrix structure. Since $Re[\lambda_2(\Lambda(x))] > 0, \forall x \in \Re^{mn}$, (see Appendix A) and $\lambda_1(\Lambda(x)) = 0$ with corresponding eigenvector $\mathbf{1}_m$, (52) implies that $x_1 = x_2 = \ldots = x_m$. Thus $\mathcal{A} \subseteq \mathcal{C}$. On the other hand, $\mathcal{C} \subseteq \mathcal{A}$ because for any $\tilde{y} \in \mathcal{C}$, we have that $W(\tilde{y},\mathcal{G}), \forall \mathcal{G} \in G$, is row stochastic (see Assumption 1), and thus $W(\tilde{y},\mathcal{G})\tilde{y} = \tilde{y}, \forall \mathcal{G} \in G$, and hence $\tilde{y} \in \mathcal{A}$. Therefore, $\mathcal{A} = \mathcal{C}$.

Appendix C

For arbitrary $x, y \in \Re^{mn}$, we have

$$\|T(x) - T(y)\| = \|Wx - Wy\| = \|W(x-y)\| \leq \|W\|\|x-y\|.$$

By Assumption 1 and Lemma 1, we obtain $\|W\| \leq 1$. Thus we have

$$\|T(x) - T(y)\| \leq \|W\|\|x-y\| \leq \|x-y\|$$

which implies that T is nonexpasnive. In the case where $\|W\| < 1$, we have

$$\|T(x) - T(y)\| < \|x-y\|$$

implying that T is a contraction.

Acknowledgment

The authors would like to thank the Associate Editor and anonymous reviewers for their constructive comments, and another proof for *Step 1*, that have improved the quality of this work.

References

[1] D. Jakovetić, D. Bajović, J. Xavier, and J. M. F. Moura, “Primal-dual methods for large-scale and distributed convex optimization and data analysis,” *Proc. IEEE*, vol. 108, no. 11, pp. 1923–1938, Nov. 2020.

[2] T. Yang *et al.*, “A survey of distributed optimization,” *Annu. Rev. Control*, vol. 47, pp. 278–305, 2019.

[3] D. K. Mazlum *et al.*, “A survey of distributed optimization and control algorithms for electric power systems,” *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 2941–2962, Nov. 2017.

[4] A. Nedić, “Distributed optimization,” *Encyclopedia of Systems and Control*, pp. 1–12, 2014, doi: 10.1007/978-1-4471-5102-9_219-1.

[5] S. S. Ram, A. Nedić, and V. V. Veeravalli, “A new class of distributed optimization algorithms: Application to regression of distributed data,” *Optim. Methods Softw.*, vol. 27, pp. 71–88, 2012.

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

[7] T. Chang, A. Nedić, and A. Scaglione, “Distributed constrained optimization by consensus-based primal-dual perturbation method,” *IEEE Trans. Autom. Control*, vol. 59, no. 6, pp. 1524–1538, Jun. 2014.

[8] I. Necoara and A. Nedich, “A fully distributed dual gradient method with linear convergence for large-scale separable convex problems,” in *Proc. Eur. Control Conf.*, Linz, Austria, 2015, pp. 304–309.

[9] A. Nedić and A. Olshevsky, “Distributed optimization over time-varying directed graphs,” *IEEE Trans. Autom. Control*, vol. 60, no. 3, pp. 601–615, Mar. 2015.

[10] I. Necoara and V. Nedelcu, “On linear convergence of a distributed dual gradient algorithm for linearly constrained separable convex problems,” *Automatica*, vol. 55, pp. 209–216, 2015.

[11] S. Sh Alaviani and N. Elia, “Distributed convex optimization with state-dependent interactions,” in *Proc. 25th Mediterranean Conf. Control Automat.*, Valletta, Malta, 2017, pp. 129–134.

[12] I. Lobel, A. Ozdaglar, and D. Feiger, “Distributed multi-agent optimization with state-dependent communication,” *Math. Program. Ser. B*, vol. 129, pp. 255–284, 2011.

[13] B. Zhou, X. Liao, T. Huang, and G. Chen, “Distributed multi-agent optimization with inequality constraints and random projections,” *Neurocomputing*, vol. 197, pp. 195–204, 2016.

[14] D. Varagnolo, F. Zanella, A. Cenedese, G. Pillonetto, and L. Schenato, “Newton-Raphson consensus for distributed convex optimization,” *IEEE Trans. Autom. Control*, vol. 61, no. 4, pp. 994–1009, Apr. 2016.

[15] G. Shi, K. H. Johansson, and Y. Hong, “Reaching an optimal consensus: Dynamical systems that compute intersections of convex sets,” *IEEE Trans. Autom. Control*, vol. 58, no. 3, pp. 610–622, Mar. 2013.

[16] J. Wang and N. Elia, “Control approach to distributed optimization,” in *Proc. Allerton Conf. Commun. Control Comput.*, Allerton House, Illinois, USA, 2010, pp. 557–561.

[17] W. Chen and W. Ren, "Event-triggered zero-gradient-sum distributed consensus optimization over directed networks," *Automatica*, vol. 65, pp. 90–97, 2016.

[18] J. Liu and W. Chen, "Distributed convex optimisation with event-triggered communication in networked systems," *Int. J. Syst. Sci.*, vol. 47, pp. 3876–3887, 2016.

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

[20] X. Wang, Y. Hong, and H. Ji, "Distributed optimization for a class of nonlinear multiagent systems with disturbance rejection," *IEEE Trans. Cybern.*, vol. 46, no. 7, pp. 1655–1666, Jul. 2016.

[21] T. H. Chang, M. Hong, and X. Wang, "Multi-agent distributed optimization via inexact consensus ADMM," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 482–497, Jan. 2015.

[22] T. Erseghe, "A distributed and scalable processing method based upon ADMM," *IEEE Signal Process. Lett.*, vol. 19, no. 9, pp. 563–566, Sep. 2012.

[23] J. F. C. Mota, J. M. F. Xavier, P. M. Q. Aguiar, and M. Püschel, "Distributed optimization with local domains: Applications in MPC and network flows," *IEEE Trans. Autom. Control*, vol. 60, no. 7, pp. 2004–2009, Jul. 2015.

[24] N. Bastianello, M. Todescato, R. Carli, and L. Schenato, "Distributed optimization over lossy networks via relaxed Peaceman-Rachford splitting: a robust ADMM approach," in *Proc. Eur. Control Conf.*, Limassol, Cyprus, 2018, pp. 477–482.

[25] N. Bastianello, R. Carli, L. Schenato, and M. Todescato, "A partition-based implementation of the relaxed ADMM for distributed convex optimization over lossy networks," in *Proc. IEEE Conf. Decis. Control*, Miami Beach, FL, USA, 2018, pp. 3379–3384.

[26] L. Majzoobi, F. Lahouti, and V. Shah-Mansouri, "Analysis of distributed ADMM algorithm for consensus optimization in presence of node error," *IEEE Trans. Signal Process.*, vol. 67, no. 7, pp. 1774–1784, Apr. 2019.

[27] Q. Ling, W. Shi, G. Wu, and A. Ribeiro, "DLM: Decentralized linearized alternating direction method of multipliers," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 4051–4064, Aug. 2015.

[28] J. N. Tsitsiklis, "Problems in decentralized decision making and computation," Ph.D. dissertation, Dept. Elect. Eng. Comp. Sci., MIT, Cambridge, MA, 1984.

[29] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ, China: Prentice Hall, 1989.

[30] M. Zhong and C. G. Cassandras, "Asynchronous distributed optimization with minimal communication and connectivity preservation," in *Proc. Joint 48th Conf. Decis. Control 28th Chin. Control Conf.*, Shanghai, China, 2009, pp. 5396–5401.

[31] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. AC-31, no. 9, pp. 803–812, Sep. 1986.

[32] A. Nedić and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Trans. Autom. Control*, vol. 54, pp. 48–61, Jan. 2009.

[33] S. Lee and A. Nedić, "Distributed random projection algorithm for convex optimization," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 2, pp. 221–229, Apr. 2013.

[34] M. Zhong and C. G. Cassandras, "Asynchronous distributed optimization with event-driven communication," *IEEE Trans. Autom. Control*, vol. 55, no. 12, pp. 2735–2750, Dec. 2010.

[35] M. Zhong and C. G. Cassandras, "Asynchronous distributed optimization with minimal communication," in *Proc. 47th Conf. Decis. Control*, Cancun, Mexico, 2008, pp. 363–368.

[36] W. Liu and Z. Hua, "Asynchronous algorithms for distributed optimisation and application to distributed regression with robustness to outliers," *IET Control Theory Appl.*, vol. 7, pp. 2084–2089, 2013.

[37] D. Liberzon, *Switching in Systems and Control*, Cambridge, MA, USA: Birkhäuser, 2003.

[38] F. Cucker and S. Smale, "Emergent behavior in flocks," *IEEE Trans. Autom. Control*, no. 5, vol. 52, pp. 852–862, May 2007.

[39] U. Krause, "A discrete nonlinear and non-autonomous model of consensus formation," in *Proc. Commun. Difference Equ., Gordon Breach*, Amsterdam, 2000, pp. 227–236.

[40] S. Totsch and E. Tadmor, "Heterophilious dynamics enhances consensus," *SIAM Rev.*, vol. 56, pp. 577–621, 2014.

[41] D. Acemoglu, A. Ozdaglar, and A. Parandehgheibi, "Spread of (mis)information in social networks," *Games Econ. Behav.*, vol. 70, pp. 194–227, 2010.

[42] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dyn. Games Appl.*, vol. 1, pp. 3–49, 2011.

[43] D. Acemoglu, G. Como, F. Fagnani, and A. Ozdaglar, "Opinion fluctuations and disagreement in social networks," *Math. Operations Res.*, vol. 38, pp. 1–27, 2013.

[44] D. Acemoglu, K. Bimpikis, and A. Ozdaglar, "Dynamics of information exchange in endogenous social networks," *Theor. Econ.*, vol. 9, pp. 41–97, 2014.

[45] R. Heyselmann and U. Krause, "Opinion dynamics and bounded confidence moels, analysis, and simulation," *J. Artif. Societies Social Simul.*, vol. 5, pp. 1–33, 2002.

[46] V. D. Blondel, J. M. Hendrickx, and J. N. Tsitsiklis, "On Krause's multi-agent consensus model with state-depednet connectivity," *IEEE Trans. Autom. Control*, vol. 54, no. 11, pp. 2586–2597, Nov. 2009.

[47] D. Acemoglu, A. Ozdaglar, and E. Yildiz, "Diffusion of innovations in social networks," in *Proc. 50th IEEE Conf. on Dec. Cont. Eur. Cont. Conf.*, Orlando, FL, USA, 2011, pp. 2329–2334.

[48] A. Simonetto, T. Kevicsky, and R. Babuška, "Constrained distributed algebraic connectivity maximization in robotic networks," *Automatica*, vol. 49, pp. 1348–1357, 2013.

[49] Y. Kim and M. Mesbahi, "On maximizing the second smallest eigenvalue of a state-dependent graph laplacian," *IEEE Trans. Autom. Control*, vol. 51, pp. 116–120, 2006.

[50] D. D. Siljak, " Dynamic graphs," *Nonlin. Anal.: Hybrid Syst.*, vol. 2, pp. 544–567, 2008.

[51] I. Rajapakse, M. Groudine, and M. Mesbahi, "Dynamics and control of state-dependent networks for probing genomic organization," in *Proc. Nation. Acad. Sci. USA*, vol. 108, pp. 17257–17262, 2011.

[52] L. Perea, P. Elosequi, and G. Gómez, "Extension of the cucker-smale control law to space flight formations," *J. Guid. Cont. Dyn.*, vol. 32, pp. 526–536, 2009.

[53] F. Paita, G. Gómez, and J. J. Masdemont, "A distributed attitude control law for formation flying based on the cucker-smale model," in *Proc. Int. Astronautical Congr.*, Toronto, Canada, 2014, pp. 1–11.

[54] V. Trianni, D. D. Simone, A. Reina, and A. Baronchelli, "Emergence of consensus in a multi-robot network: From abstract models to empirical validation," *IEEE Robot. Automat. Lett.*, vol. 1, no. 1, pp. 348–353, Jan. 2016.

[55] J. M. Hendrickx and J. N. Tsitsiklis, "Convergence of type-symmetric and cut-balanced consensus seeking systems," *IEEE Trans. Autom. Control*, vol. 58, no. 1, pp. 214–218, Jan. 2013.

[56] G. Jing, Y. Zheng, and L. Wang, "Consensus of multiagent systems with distance-dependent communication networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2712–2726, Nov. 2017.

[57] G. Jing and L. Wang, "Finite time coordination under state-dependent communication graphs with inherent links," *IEEE Trans. Circuits Syst., II, Exp. Briefs*, vol. 66, no. 6, pp. 968–972, Jun. 2019.

[58] O. Slučiak and M. Rupp, "Consensus algorithm with state-dependent weights," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1972–1985, Apr. 2016.

[59] A. Bogojeska, M. Mirchev, I. Mishkovski, and L. Kocarev, "Synchronization and consensus in state-dependent networks," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 2, pp. 522–529, Feb. 2014.

[60] A. Awad, A. Chapman, E. Schoof, A. Narang-Siddarth, and M. Mesbahi, "Time-scale separation on networks: Consensus, tracking, and state-dependent interactions," in *Proc. IEEE 54th Annu. Conf. Decis. Control*, Osaka, Japan, 2015, pp. 6172–6177.

[61] S. Sh Alaviani and N. Elia, "Distributed multi-agent convex optimization over random digraphs," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 986–998, Mar. 2020.

[62] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY, USA: Cambridge Univ. Press, 1985.

[63] H. K. Xu, "Iterative algorithms for nonlinear operators," *J. London Math. Soc.*, vol. 66, pp. 240–256, 2002.

[64] S. Banach, "Sur les opérations dans les ensembles abstraits et leurs applications aux équations integrals," *Fundam. math.*, vol. 3, pp. 133–181, 1922.

[65] M. de Condorcet, *Condorcet: Selected Writings*, K. M. Baker Ed. Indianapolis: Bobbs-Merrill, 1976.

[66] T. Zeng, O. Semiari, W. Saad, and M. Bennis, "Joint communication and control for wireless autonomous vehicular platoon systems," *IEEE Trans. Commun.*, vol. 67, pp. 7907–7922, 2019.

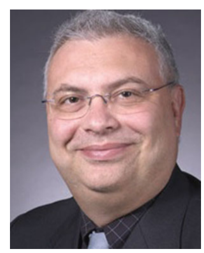
[67] C. Fischione, "Fast-lipschitz optimization with wireless sensor networks applications," *IEEE Trans. Autom. Control*, vol. 56, no. 10, pp. 2319–2331, Oct. 2011.

[68] M. Jakobsson and C. Fischione, "Extensions of fast-lipschitz optimization for convex and nonconvex problems," in *Proc. 3rd IFAC Workshop Distrib. Estimation Control Networked Syst.*, Santa Barbara, California, USA, 2012, pp. 162–167.

[69] M. Jakobsson, S. Magnusson, C. Fischione, and P. C. Weeraddana, "Extensions of fast-lipschitz optimization," *IEEE Trans. Autom. Control*, vol. 61, no. 4, pp. 861–876, Apr. 2016.
[70] H. Iiduka, "Acceleration method for convex optimization over the fixed point set of a nonexpansive mapping," *Math. Program., Ser. A*, vol. 149, pp. 131–165, 2015.
[71] H. Iiduka, "Convex optimization over fixed point sets of quasi-nonexpansive and nonexpansive mappings in utility-based bandwidth allocation problems with operational constraints," *J. Comput. Appl. Math.*, vol. 282, pp. 225–236, 2015.
[72] H. Iiduka, "Fixed point optimization algorithm and its applications to network bandwidth allocation," *J. Comput. Appl. Math.*, vol. 236, pp. 1733–1742, 2012.
[73] K. Slavakis, J. Yamada, and K. Sakaniwa, "Computation of symmetric positive definite Toeplitz matrices by the hybrid steepest descent method," *Signal Process.*, vol. 83, pp. 1135–1140, 2003.
[74] K. Slavakis and I. Yamada, "Robust wideband beamforming by the hybrid steepest descent method," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4511–4532, Sep. 2007.
[75] I. Yamada, N. Ogura, and N. Shirakawa, "A numerical robust hybrid steepest descent method for the convexly constrained generalized inverse problems," *Contemp. Math.*, vol. 313, pp. 269–305, 2002.
[76] H. Iiduka, "Fixed point optimization algorithms for distributed optimization in networked systems," *SIAM J. Optim.*, vol. 23, pp. 1–26, 2013.
[77] H. Iiduka, "Parallel optimization algorithm for smooth convex optimization over fixed point sets of quasi-nonexpansive mappings," *J. Operation Res. Soc. Jpn.*, vol. 58, pp. 330–352, 2015.
[78] H. Iiduka, "Convergence analysis of iterative methods for nonsmooth convex optimization over fixed point sets of quasi-nonexpansive mappings," *Math. Prog., Ser. A*, vol. 159, pp. 509–538, 2016.
[79] S. Bubeck, "Convex optimization: Algorithms and complexity," *Foundations Trends Mach. Learn.*, vol. 8, pp. 231–357, 2015.
[80] E. K. Ryu and S. Boyd, "A primer on monotone operator methods," *Appl. Comput. Math.*, vol. 15, pp. 3–43, 2016.
[81] S. S. Alaviani and N. Elia, "Distributed average consensus over random networks," in *Proc. Amer. Cont. Conf.*, Philadelphia, PA, USA, 2019, pp. 1854–1859.
[82] S. S. Alaviani and N. Elia, "A distributed algorithm for solving linear algebraic equations over random networks," in *Proc. IEEE Conf. Decis. Control*, Miami Beach, FL, USA, 2018, pp. 83–88.
[83] S. S. Alaviani and N. Elia, "A distributed algorithm for solving linear algebraic equations over random networks," *IEEE Trans. Autom. Control*, to be published, doi: 10.1109/TAC.2020.3010264.
[84] S. Sh Alaviani, "Applications of fixed point theory to distributed optimization, robust convex optimization, and stability of stochastic systems," Ph.D. dissertation, Dept. Elect. Comput. Eng., Iowa State Univ., Ames, Iowa, USA, 2019.
[85] A. Kattepur, H. K. Rath, A. Mukherjee, and A. Simha, "Distributed optimization framework for industry 4.0 automated warehouses," *EAI Endorsed Trans. Ind. Netw. Intell. Syst.*, vol. 5, pp. 1–10, 2018.
[86] K. Pahlavan and H. Levesque, *Wireless Information Networks*. New York, NY, USA: Wiley, 1995.

Seyyed Shaho Alaviani (Member, IEEE) received the M.E. and Ph.D. degrees in electrical engineering from Iowa State University, Ames, IA, USA, in 2018 and 2019, respectively. He is currently a Postdoctoral Fellow with the Department of Mechanical Engineering, Clemson University, Clemson, SC, USA.

His research interests include applied functional analysis, applied geometry, agents and agent-based systems, distributed optimization, robust convex and nonconvex optimization, machine learning, reinforcement learning, electric vehicles, stability and control of dynamical systems, game theory, spectral graph theory, time-delay systems, singular systems, two-dimensional systems, chaos, and applications of control theory in renewable energy systems.



Nicola Elia (Fellow, IEEE) received the Laurea degree in electrical engineering from the Politecnico of Turin, Turin, Italy, in 1987 and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1996.

From 1987 to 1990, he was a Control Engineer with Fiat Research Center, Turin, Italy. From 1996 to 1999, he was a Postdoctoral Associate with the Laboratory for Information and Decision Systems, MIT. From 1999 to 2018, he was with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, USA. He is currently a Professor of electrical and computer engineering with the University of Minnesota, Minneapolis, MN, USA. His research interests include computational methods for controller design, communication systems with access to feedback, control with communication constraints, and networked systems. He was the recipient of the National Science Foundation CAREER Award in 2001