

Duality and Stability in Complex Multiagent State-Dependent Network Dynamics

S. Rasoul Etesami*

Abstract

Despite significant progress on stability analysis of conventional multiagent networked systems with weakly coupled state-network dynamics, most of the existing results have shortcomings in addressing multiagent systems with highly coupled state-network dynamics. Motivated by numerous applications of such dynamics, in our previous work [1], we initiated a new direction for stability analysis of such systems that uses a sequential optimization framework. Building upon that, in this paper, we extend our results by providing another angle on multiagent network dynamics from a duality perspective, which allows us to view the network structure as dual variables of a constrained nonlinear program. Leveraging that idea, we show that the evolution of the coupled state-network multiagent dynamics can be viewed as iterates of a primal-dual algorithm for a static constrained optimization/saddle-point problem. This view bridges the Lyapunov stability of state-dependent network dynamics and frequently used optimization techniques such as block coordinated descent, mirror descent, the Newton method, and the subgradient method. As a result, we develop a systematic framework for analyzing the Lyapunov stability of state-dependent network dynamics using techniques from nonlinear optimization. Finally, we support our theoretical results through numerical simulations from social science.

Index Terms

Lyapunov stability; multiagent systems; state-dependent network dynamics; saddle-point dynamics; block coordinate descent; Newton method; nonlinear optimization.

I. INTRODUCTION

Many of the current challenges in science and engineering are related to complex networks, and distributed multiagent network systems are currently the focal point of many new applications. Such applications relate to the growing popularity of social networks, the analysis of large network data sets, and the problems that arise from interactions among agents in complex political, economic, and biological systems. These challenges may involve modeling of the interactions of agents in complex networks, the establishment of stability in the agents' interaction dynamics, and the design of efficient algorithms to obtain or approximate the equilibrium points. We can offer many motivating examples of relationships in political, social, and engineering applications that are governed by complex networks of heterogeneous agents. Agents may be strategic, or the networks can be dynamic in the sense that they can vary over time, depending on the agents' states or decisions. The following are a few examples.

– *Network security*: A basic task in network security is that of providing a mechanism for securing the operation of a set of networked heterogeneous agents (e.g., service providers, computers, or data centers) despite external malicious attacks (Figure 1). One way of doing that is to incentivize the agents to invest in their security (e.g., by installing antivirus software) [2]. However, since the agents are interconnected, the compromise of one agent may affect its neighbors, and such a failure can cascade over the entire network. As a result, the decision made by each agent on how much to invest in its security level will indirectly affect all the others, and hence the connectivity structure of the network. Thus we face a highly dynamic network of heterogeneous agents in which the agents' states/decisions and the network structure are highly influenced by each other.

*Department of Industrial and Enterprise Systems Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, 61801. Email: etesami1@illinois.edu

This material is based upon work supported by the National Science Foundation under Grant No. EPCN-1944403.

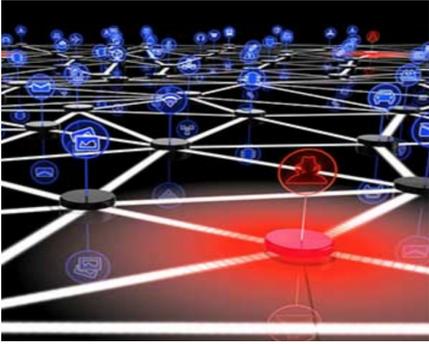


Fig. 1. Compromise of an agent changes the network structure and hence the security decisions of all other agents.



Fig. 2. Aircraft must keep a certain formation while the communication network among them is subject to change.

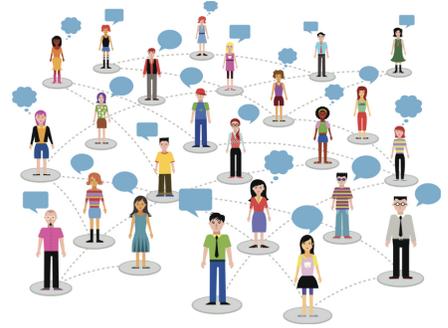


Fig. 3. The social network affects opinions, and that in turn creates new friendships and hence a new social network.

– *Formation control*: A goal in formation control is to design a distributed protocol such that a set of agents (e.g., the aircraft in Figure 2) collectively form a specific structure and eventually accomplish a task [3]. Agents may have different communication capabilities and can communicate only with those in their local neighborhoods. Consequently, depending on the agents’ states (e.g., remaining power or relative positions), the communication network they share is subject to change. As a result, the agents’ states and the communication network are highly coupled and dynamically evolve based on each other.

– *Social networks*: In social networks, there are often apparent affinities among people based on heterogeneous political or cultural beliefs that define an interaction network among them. However, on specific issues, alliances form among people from different groups. Almost every congressional vote provides an example of this phenomenon, wherein some representatives break away from their respective parties to vote with the other party [4] (Figure 3).

– *Stability of smart grids*: In the emerging smart grid, a significant amount of energy stems from renewable sources, electric vehicles, and storage units, many of which may be owned by consumers rather than utility companies. That phenomenon is turning every grid component into a *prosumer*: a joint producer and consumer of energy. Prosumers (agents) in the smart grid strategically interact with each other subject to power network constraints. In particular, depending on their states (e.g., energy consumption/production decisions), they may decide to buy/sell energy to different agents. Thus, a significant challenge is that of providing decentralized algorithms to stabilize the demand and response given that the structure of the agents’ interactions is a function of their own and their neighbors’ states/decisions.

Motivated by the above, and many other real applications such as robotics [5], [6], genomic [7], robustification [8], and consensus seeking [9], [10], our objective in this paper is to provide a systematic approach for analyzing the stability and convergence of agents interacting over a rich dynamic network that may evolve or vary based on agents’ states. To that end, we provide new connections between analysis of multiagent network systems and developed techniques in the mature field of nonlinear programming. Utilizing such connections, we show how Lyapunov stability of seemingly complex multiagent network dynamics can be analyzed using iterative optimization algorithms for finding a minimum or saddle-point of nonlinear functions.

A. Related Work

A general multiagent network problem involves a set of $[n] = \{1, 2, \dots, n\}$ agents (social individuals, grid prosumers, unmanned vehicles, etc.) At each time instance $k = 0, 1, 2, \dots$, there is an underlying network $\mathcal{G}_k = ([n], \mathcal{E}_k)$ that determines the communication network shared by the agents. Here, \mathcal{E}_k denotes the set of edges of the network at time k ; the edges can be undirected or directed. The state of agent $i \in [n]$ at time k is given by a vector x_i^k and evolves based on i ’s interaction with its neighbors.

In particular, the overall state of the system at time $k + 1$, denoted by \mathbf{x}^{k+1} , can be obtained by using a general update rule $\mathbf{x}^{k+1} = f_k(\mathbf{x}^k, \mathcal{G}_k)$, $k = 0, 1, 2, \dots$, where $f_k(\cdot)$ can be a general time-varying function, depending on the problem setup, and captures the interaction laws among the agents. Therefore, the main goal here is to understand whether the generated sequence of states $\{\mathbf{x}^k\}_{k=0}^\infty$ will converge (stabilize) to any equilibrium. That has been the subject of much research effort, including work in distributed control and computation [11]. Unfortunately, despite enormous efforts in the area, the stability problem for such dynamics in its full generality is still far from being solved. However, partial solutions to the problem under certain simplifying assumptions are known. For instance, there has been a rich body of literature on the analysis of multiagent network systems, mainly from the static point of view, in which a set of agents iteratively interact over a *fixed* network to achieve a certain goal, such as consensus or optimization of an objective function. The classical models of DeGroot [12] and Friedkin-Johnsen [13] in the social sciences are two special types of such systems [11], [14]. Below is a sample result in this area [15], [16]:

Theorem 1: Given a fixed and undirected connected network $\mathcal{G} = ([n], \mathcal{E})$, at any time $k = 0, 1, \dots$, let every agent i take the average of its own state and those of its neighbors, i.e., $x_i^{k+1} = \sum_{j \in N_i \cup \{i\}} a_{ij} x_j^k$, $\forall i \in [n]$, where $a_{ij} > 0$ are positive constant weights such that $\sum_{j=1}^n a_{ij} = 1$, $\forall i$. Then the generated averaging dynamics are Lyapunov stable and converge to an equilibrium point.

By comparing Theorem 1 with the aforementioned general dynamics, one can identify several simplifying assumptions that have been made in most of the existing results on multiagent network systems. 1) The underlying networks are fixed as $\mathcal{G}_k = \mathcal{G}$, $\forall k$. 2) The underlying networks are connected and undirected. 3) The underlying networks \mathcal{G}_k do not depend on the agents' states \mathbf{x}^k . Therefore, a large body of literature has been developed to establish the stability of the above general dynamics under less restrictive assumptions. The results in the static case can often be generalized to time-varying networks through the assumption of a certain "independency" between the network process and the state dynamics. For instance, one of the commonly used assumptions is that the network dynamics are governed by an exogenous process that is uncoupled from the state dynamics [17]–[23]. Here is an extension from [24]:

Theorem 2: Consider a sequence of time-varying directed graphs $\mathcal{G}_k = ([n], \mathcal{E}_k)$ with the weight of edge (i, j) at time k being $a_{ij}(k)$. Assume that the sequence of graphs is B -strongly connected, meaning that for any $k \geq 0$, the graph $\mathcal{G} = ([n], \cup_{s=k}^{k+B} \mathcal{E}_s)$ is strongly connected. Moreover, given a positive constant $\alpha \in (0, 1)$, assume $a_{ij}(k) \in [\alpha, 1] \cup \{0\}$, $\forall i, j, k$, and $a_{ii}(k) \geq \alpha$, $\sum_j a_{ij}(k) = 1$, $\forall i, k$. Then the averaging dynamics $x_i^{k+1} = \sum_{j \in N_i} a_{ij}(k) x_j^k$, $i \in [n]$, will converge to a consensus point.

While Theorem 2 relaxes the static communication network to time-varying networks, it still has shortcomings in addressing many realistic multiagent systems. For instance, the network connectivity must be preserved over any time window of length B , and it is hard to check whether that is happening (especially if the networks are generated endogenously based on the agents' states). Second, the assumptions on the weight matrices are somewhat restrictive as, in real situations, the weights can approach 0 and then increase again to 1. Besides, the theorem uses an implicit assumption on the symmetry of the networks by imposing strong connectivity. Finally, in realistic situations, the evolution of the network itself depends on the evolution of agents' states. In contrast, in the above theorem, the network dynamics are driven by an exogenous process that is independent of how the states evolve. A generalization of Theorem 2 is to allow weak coupling between the state and network dynamics, with certain network connectivity/symmetry assumptions [15], [25], [26]. We refer to [27], [28] for other extensions of such results that use a backward product of stochastic matrices. We mention here that our work is also related to dynamic clustering, for which the goal is to provide a theoretical justification for cluster synchronization in multiagent systems by using saddle-point dynamics [29], [30]. However, the network structure in that application is fixed and captured by a set of linear constraints. In contrast, in our work, the network dynamically evolves as a complex function of the state variables.

While the existing results can address a large class of multiagent network systems, there are still many examples that do not fit into any of the categories mentioned above or for which the application of the above techniques provides poor results on the behavior of the agents. Our work is fundamentally different from the earlier literature and offers a new perspective for the averaging dynamics by capturing the internal co-evolution of state and network dynamics. Thus, in this paper, we depart from conventional methods for stability analysis of multiagent dynamics, such as Markov chains or products of stochastic matrices. Our approach allows us to relax some of the common assumptions, such as global knowledge on the network connectivity throughout the dynamics. We believe that this new approach, together with the current results on the multiagent averaging dynamics, can be used to analyze a broader class of complex state-dependent network dynamics.

B. Contributions and Organization

Inspired by the above shortcomings and building upon our previous work [1], in this paper, we provide a principled framework from an optimization perspective to study the Lyapunov stability of multiagent state-dependent network dynamics. We show that despite the challenges due to state-network coupling, it is still possible to capture the co-evolution of network and state dynamics for a broad class of multiagent systems, even under an asymmetric or nonlinear environment. More precisely, we show that often the network structure among the agents can be viewed as *dual variables* of a constrained optimization problem, where the existence of an edge is related to the tightness of the corresponding constraint. As a result, we can view multiagent network dynamics as an iterative primal-dual algorithm for a static constrained optimization problem where the primal updates correspond to state updates of the dynamics, and the dual updates correspond to the network evolution. The KKT optimality conditions also guide the coupling between the network and state dynamics. This approach allows us to view the constrained Lagrangian of the underlying static problem as a Lyapunov or “semi-Lyapunov” function for the multiagent dynamics. Therefore, we obtain a principled way to establish the stability of multiagent network dynamics in terms of the asymptotic convergence of an iterative optimization algorithm. That means that a variety of iterative optimization methods can be used to study the stability of multiagent state-dependent network dynamics.

In Section II, we first provide our problem formulation, which involves modeling of a large class of state-dependent network dynamics. In Section III, we apply a sequential optimization framework based on the block coordinate descent method to establish Lyapunov stability for a large class of state-dependent network dynamics. We consider that method under both symmetric and asymmetric network structures and use the change of variables to generate other types of state-dependent network dynamics. In Section IV, we use a saddle-point model to extend our results to a case in which there is a conflict between the network structure and the state evolution. In Section V, we consider continuous-time dynamics for which the edge emergence between agents is no longer a binary event, but rather a continuous weight process. In Section VI, we provide numerical results, and we conclude the paper by identifying some future directions of research in Section VII.

C. Notation

For a positive integer $n \in \mathbb{Z}^+$ we set $[n] := \{1, 2, \dots, n\}$. We use bold symbols for vectors. Given a vector $\mathbf{v} \in \mathbb{R}^n$ we use \mathbf{v}^T to denote its transpose, $\|\mathbf{v}\|^2 = \mathbf{v}^T \mathbf{v}$ to denote its Euclidean norm, and $\|\mathbf{v}\|_1 = \sum_{i=1}^n |v_i|$ to denote its l_1 -norm. We let $\text{diag}(\mathbf{v})$ be a diagonal matrix with vector \mathbf{v} as its diagonal elements and zero elsewhere. Given a positive-definite matrix Q , we let $\|\mathbf{v}\|_Q^2 = \mathbf{v}^T Q \mathbf{v}$. We use \mathcal{C} to denote the class of real-valued differentiable functions with a finite global minimum. Similarly, we let \mathcal{C}^2 be the class of twice differentiable functions with a finite global minimum. Given a strictly convex function $\Psi \in \mathcal{C}$, we use $D_\Psi(\mathbf{x}, \mathbf{y}) = \Psi(\mathbf{x}) - \Psi(\mathbf{y}) - \nabla \Psi(\mathbf{y})^T (\mathbf{x} - \mathbf{y})$ to denote the Bregman divergence with respect to Ψ . Finally, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is called *L-Lipschitz continuous* if there exists $L > 0$ such that $|f(\mathbf{x}) - f(\mathbf{y})| \leq L \|\mathbf{x} - \mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

II. PROBLEM FORMULATION

Let us consider a multiagent network system consisting of $[n]$ agents. At any given time $k = 0, 1, 2, \dots$, we use $x_i^k \in \mathbb{R}$ to denote the state of agent i , and $\mathbf{x}^k = (x_1^k, \dots, x_n^k)^T$ to denote the state of the entire system at that time.¹ Moreover, we assume that each agent $i \in [n]$ has $n - 1$ measurement functions $g_{ij}(x_i, x_j) : \mathbb{R}^2 \rightarrow \mathbb{R}$, one for every other agent $j \in [n] \setminus \{i\}$. For most of this paper, we assume that the measurement functions g_{ij} are twice differentiable and belong to \mathcal{C}^2 . Given any state \mathbf{x} , we assume that the set of neighbors of an agent $i \in [n]$ is determined by the logic constraints $g_{ij}(x_i, x_j) \leq 0$, $j \in [n] \setminus \{i\}$. In other words, for a given state \mathbf{x} , agent i is influenced by agent j (or j is a neighbor of i) if and only if $g_{ij}(x_i, x_j) \leq 0$. In particular, we use $N_i(\mathbf{x}) := \{j : g_{ij}(x_i, x_j) \leq 0\}$ to denote the set of neighbors of agent i at a state \mathbf{x} . At any time instance k , each agent $i \in [n]$ interacts with its neighbors and updates its state at the next time step to

$$x_i^k = \phi_i(\mathbf{x}^k, N_i(\mathbf{x}^k)), \quad i \in [n], \quad (1)$$

where $\phi_i(\cdot)$ is an agent-specific update rule, which is a function of the states of agent i 's neighbors. Note that the above discrete-time dynamics contain a broad class of state-dependent network dynamics for which the network at time k is given by $\mathcal{G}_k = ([n], \{(i, j) : j \in N_i(\mathbf{x}^k)\})$. It is evident that the network structure at time k depends on the agents' states at that time, and the state at the next time step $k + 1$ is a function of the network structure at the current time k .

Definition 1: The measurement functions $g_{ij}(\cdot)$ are called *symmetric* if for all $i \neq j$ we have $g_{ij}(x_i, x_j) = g_{ji}(x_j, x_i)$. Note that for symmetric measurement functions, the communication network \mathcal{G}_k at any time k is an undirected graph.

One of our main objectives in this paper is to provide a general class of update rules $\phi_i(\cdot)$ such that the state-dependent network dynamics (1) converge to some equilibrium point or are Lyapunov stable in the following sense:

Definition 2: A function $V : \mathbb{R}^n \rightarrow \mathbb{R}$ is called a *Lyapunov function* for the discrete time dynamical system $\mathbf{z}^{k+1} = h_k(\mathbf{z}^k)$, $k = 0, 1, 2, \dots$, if it is decreasing along the trajectories of the dynamics, i.e., $V(\mathbf{z}^{k+1}) < V(\mathbf{z}^k)$, $\forall k$. We refer to a dynamical system that admits a Lyapunov function as *Lyapunov stable*.

Remark 1: While a Lyapunov function $V(\cdot)$ is typically defined to be a nonnegative function with $V(0) = 0$, as we shall see, all the Lyapunov functions in this paper are bounded below by some global constant M . Therefore, if those functions are shifted by a constant $|M|$, it is ensured that $V(\cdot) + |M|$ is nonnegative and strictly decreasing along the trajectory of the dynamics. Moreover, we do not require $V(0) = 0$, as the origin is not necessarily an equilibrium point of the dynamics.

To illustrate the generality of the above model, let us consider the well-known homogeneous Hegselmann-Krause (HK) model from social science [4]. In the homogeneous HK model, there is a set of $[n]$ agents, and it is assumed that at each time instance $k = 0, 1, 2, \dots$, the opinion (state) of agent $i \in [n]$ can be represented by a scalar $x_i^k \in \mathbb{R}$. Each agent i updates its state at time $k + 1$ by taking the arithmetic average of its state and those of others that are in its ϵ -neighborhood at time k , i.e.,

$$x_i^{k+1} = \frac{x_i^k + \sum_{j \in N_i(\mathbf{x}^k)} x_j^k}{1 + |N_i(\mathbf{x}^k)|}, \quad i \in [n].$$

Here $\epsilon > 0$ is a constant parameter, and $N_i(\mathbf{x}^k) = \{j \in [n] \setminus \{i\} : |x_i^k - x_j^k| \leq \epsilon\}$ denotes the set of neighbors of agent i at time k . In fact, it is known that such dynamics are Lyapunov stable and converge to an equilibrium point [4], [31]. It is easy to see that homogeneous HK dynamics are a very special case of the state-dependent network dynamics (1) for which the symmetric measurements are $g_{ij}(x_i, x_j) = \frac{(x_i - x_j)^2}{2} - \frac{\epsilon^2}{2}$, and the update rule is given by $\phi_i(\mathbf{x}, N_i(\mathbf{x})) = \frac{x_i + \sum_{j \in N_i(\mathbf{x})} x_j}{1 + |N_i(\mathbf{x})|}$.

¹For simplicity of presentation, we assume that agents' states are scalar real numbers. However, most of the results can be naturally extended to the case in which agents' states are vectors in \mathbb{R}^d .

III. LYAPUNOV STABILITY USING BLOCK COORDINATE DESCENT

A popular approach to solving optimization problems is the so-called *block coordinate descent* (BCD) method, which is also known as the *Gauss-Seidel method*. At each iteration of this method, the objective function is minimized with respect to a single block of variables while the rest of the blocks are held *fixed*. More specifically, consider this optimization problem: $\min\{F(\mathbf{y}_1, \dots, \mathbf{y}_n), \mathbf{y}_i \in Y_i, \forall i\}$, where $Y_i \subseteq \mathbb{R}^{m_i}$ is a closed convex set, and $F : \prod_{i=1}^n Y_i \rightarrow \mathbb{R}$ is a continuous function. At iteration $t = 0, 1, \dots$ of the BCD method, the block variable \mathbf{y}_i is updated through solving of the subproblem $\mathbf{y}_i^t = \arg \min_{\mathbf{z}_i \in Y_i} F(\mathbf{y}_1^t, \dots, \mathbf{y}_{i-1}^t, \mathbf{z}_i, \mathbf{y}_{i+1}^t, \dots, \mathbf{y}_n^t)$, $i \in [n]$. Since, in practice, finding the exact minimum in each iteration might be difficult, one can consider an *inexact* BCD method, whereby a smooth regularizer is added to the objective function or is approximated by a simpler convex function. In either case, and under some mild assumptions, it can be shown that the inexact BCD method will converge to a stationary point of the objective function $F(\cdot)$.

Now let us consider the following constrained nonlinear program:

$$\begin{aligned} \min \quad & f(\mathbf{x}) := \sum_{i=1}^n f_i(x_i) \\ \text{s.t.} \quad & g_{ij}(x_i, x_j) \leq 0, \quad \forall i \neq j, \\ & \mathbf{x} \in \mathbb{R}^n, \end{aligned}$$

where $f_i \in \mathcal{C}^2, \forall i \in [n]$ and $g_{ij} \in \mathcal{C}^2, \forall i \neq j$ are differentiable measurement functions between the two members of each pair of agents. In other words, each agent $i \in [n]$ has a private function $f_i(x_i)$, and the agents collectively want to choose their states to minimize the global objective function $f(\mathbf{x}) := \sum_{i=1}^n f_i(x_i)$ while they all remain connected. If we dualize the constraints by using dual variables $\lambda_{ij} \geq 0$, and form the Lagrangian function, we have,

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{i \neq j} \lambda_{ij} g_{ij}(x_i, x_j),$$

which is a function consisting of two block variables, namely a *state* block variable $\mathbf{x} := (x_1, \dots, x_n) \in \mathbb{R}^n$, and a nonnegative *network* block variable $\boldsymbol{\lambda} := (\lambda_{ij}, i \neq j)$. Now let us minimize the Lagrangian function by using the BCD method subject to the box constraints $\lambda_{ij} \in [0, 1], \forall i, j$. As $L(\mathbf{x}, \boldsymbol{\lambda})$ is a linear function of $\boldsymbol{\lambda}$, if we fix the state block variable and minimize $L(\mathbf{x}, \boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda} \in [0, 1]^{n(n-1)}$, we get $\lambda_{ij} = 1$ if $g_{ij}(x_i, x_j) \leq 0$ (i.e., there is a directed edge from agent i to agent j), and $\lambda_{ij} = 0$ if $g_{ij}(x_i, x_j) > 0$ (i.e., no such an edge exists). Thus, fixing the state variable and minimizing the Lagrangian with respect to $\boldsymbol{\lambda} \in [0, 1]^{n(n-1)}$, the dual variables precisely capture the network structure among the agents for that state.

Motivated by many applications of distributed averaging over networks, such as for consensus [15], opinion dynamics [4], [12], distributed optimization [18], [19], and formation control [3], in this paper, we provide several classes of distributed averaging dynamics over complex state-dependent networks. As we shall see, these dynamics can not only recover several types of well-known linear averaging dynamics from the physical and social sciences, but also be extended to *nonlinear* averaging dynamics over state-dependent network topologies. The following theorem provides our first class of nonlinear averaging dynamics whose specification to quadratic measurements can recover several linear averaging dynamics.

Theorem 3: Let $f_i(x_i) \in \mathcal{C}^2$ and $g_{ij}(x_i, x_j) \in \mathcal{C}^2$ be symmetric functions with $|\frac{\partial^2 g_{ij}}{\partial x_i \partial x_j}| \leq m$, $|\frac{\partial^2 f_i}{\partial x_i^2}| \leq$

$m, \forall i, j$.² Then the state-dependent network dynamics

$$x_i^{k+1} = x_i^k - \frac{\frac{\partial}{\partial x_i} f_i(x_i^k) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k)}{2m(|N_i(\mathbf{x}^k)| + 1)}, \quad i \in [n] \quad (2)$$

admit a Lyapunov function $V(\mathbf{x}) := \sum_i f_i(x_i) + \frac{1}{2} \sum_{i,j} \min\{g_{ij}(x_i, x_j), 0\}$ such that $V(\mathbf{x}^{k+1}) \leq V(\mathbf{x}^k) - m\|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2$. If, in addition, $g_{ij}(\cdot)$ are convex and $f_i(\cdot)$ are strictly convex functions, then the dynamics (2) will converge to an equilibrium point.

Proof: Let us consider the following Lagrangian function

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{i=1}^n f_i(\mathbf{x}_i) + \frac{1}{2} \sum_{i \neq j} \lambda_{ij} g_{ij}(x_i, x_j),$$

and consider the BCD method applied to this function when $\boldsymbol{\lambda} \in [0, 1]^{n(n-1)}$ and $\mathbf{x} \in \mathbb{R}^n$. If the state variable is fixed to \mathbf{x}^k , and we set $\boldsymbol{\lambda}^k := \operatorname{argmin}_{\boldsymbol{\lambda} \in [0, 1]^{n(n-1)}} L(\mathbf{x}^k, \boldsymbol{\lambda})$, it is easy to see that $\boldsymbol{\lambda}^k$ precisely captures the network structure among the agents at the current state \mathbf{x}^k . Next, let us fix the network variable to $\boldsymbol{\lambda}^k$, and consider

$$\begin{aligned} L_k(\mathbf{x}) &:= L(\mathbf{x}, \boldsymbol{\lambda}^k) = \sum_{i=1}^n f_i(\mathbf{x}_i) + \frac{1}{2} \sum_{i \neq j} \lambda_{ij}^k g_{ij}(x_i, x_j) \\ &= \sum_{i=1}^n f_i(\mathbf{x}_i) + \frac{1}{2} \sum_i \sum_{j \in N_i(\mathbf{x}^k)} g_{ij}(x_i, x_j). \end{aligned}$$

Ideally, we want to set the state at the next time step \mathbf{x}^{k+1} to a global minimizer of $L_k(\mathbf{x})$. However, since it might be difficult to solve the minimization problem $\min_{\mathbf{x} \in \mathbb{R}^n} L_k(\mathbf{x})$, we use an inexact BCD method, with which a quadratic upper approximation of $L_k(\mathbf{x})$ is minimized. More precisely, consider the quadratic approximation of $L_k(\mathbf{x})$ at the current point \mathbf{x}^k :

$$L_k(\mathbf{x}) \simeq L(\mathbf{x}^k) + (\mathbf{x} - \mathbf{x}^k)^T \nabla L_k(\mathbf{x}^k) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^k)^T \nabla^2 L_k(\mathbf{x}^k) (\mathbf{x} - \mathbf{x}^k), \quad (3)$$

where $\nabla L_k(\mathbf{x}^k)$ is the gradient of $L_k(\mathbf{x})$ at \mathbf{x}^k , whose i th component is given by

$$[\nabla L_k(\mathbf{x}^k)]_i = \frac{\partial}{\partial x_i} f_i(x_i) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k).$$

Moreover, $\nabla^2 L_k(\mathbf{x}^k)$ is the Hessian of $L_k(\mathbf{x})$ at \mathbf{x}^k , with the Hessian matrix function

$$[\nabla^2 L_k(\mathbf{x})]_{ij} = \begin{cases} \frac{\partial^2}{\partial x_i^2} f_i(x_i) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial^2 g_{ij}(x_i, x_j)}{\partial x_i^2} & \text{if } j = i \\ \frac{\partial^2 g_{ij}(x_i, x_j)}{\partial x_i \partial x_j} & \text{if } j \in N_i(\mathbf{x}^k) \\ 0 & \text{if } j \notin N_i(\mathbf{x}^k). \end{cases}$$

Now, if we assume $|\frac{\partial^2 g_{ij}}{\partial x_i \partial x_j}| \leq m, |\frac{\partial^2 f_i}{\partial x_i^2}| \leq m, \forall i, j$, and use the Gershgorin Circle Theorem, one can see that for any \mathbf{x} , the Hessian $\nabla^2 L_k(\mathbf{x})$ is dominated by the diagonal matrix $Q_k := 2m \cdot \operatorname{diag}(|N_1(\mathbf{x}^k)| + 1, \dots, |N_n(\mathbf{x}^k)| + 1)$. Via the Tailor expansion, for every $\mathbf{x} \in \mathbb{R}^n$ there exists an $\boldsymbol{\zeta}_x \in \mathbb{R}^n$ such that,

$$\begin{aligned} L_k(\mathbf{x}) &= L(\mathbf{x}^k) + (\mathbf{x} - \mathbf{x}^k)^T \nabla L_k(\mathbf{x}^k) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^k)^T \nabla^2 L_k(\boldsymbol{\zeta}_x) (\mathbf{x} - \mathbf{x}^k) \\ &\leq L(\mathbf{x}^k) + (\mathbf{x} - \mathbf{x}^k)^T \nabla L_k(\mathbf{x}^k) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^k)^T Q_k (\mathbf{x} - \mathbf{x}^k) := u_k(\mathbf{x}). \end{aligned}$$

²For instance, any m -smooth function possesses this property.

Therefore, $u_k(\mathbf{x})$ is a quadratic upper approximation for $L_k(\mathbf{x})$ for any \mathbf{x} . Let

$$\mathbf{x}^{k+1} = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} u_k(\mathbf{x}) = \mathbf{x}^k - Q_k^{-1} \nabla L_k(\mathbf{x}^k). \quad (4)$$

We can write,

$$L_k(\mathbf{x}^{k+1}) \leq u_k(\mathbf{x}^{k+1}) \leq u_k(\mathbf{x}^k) = L_k(\mathbf{x}^k).$$

That shows that the state-dependent network dynamics,

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k - \frac{\frac{\partial}{\partial x_i} f_i(x_i^k) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k)}{2m(|N_i(\mathbf{x}^k)| + 1)},$$

are Lyapunov stable (i.e., $L(\cdot)$ decreases regardless of the state or network updates), and the function

$$V(\mathbf{x}) := \min_{\boldsymbol{\lambda} \in [0,1]^{n(n-1)}} L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_i f_i(x_i) + \frac{1}{2} \sum_{i,j} \min\{g_{ij}(\mathbf{x}), 0\}$$

serves as a Lyapunov function. Moreover, the drift of this Lyapunov is bounded by

$$\begin{aligned} V(\mathbf{x}^k) - V(\mathbf{x}^{k+1}) &= L(\mathbf{x}^k, \boldsymbol{\lambda}^k) - L(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^{k+1}) \geq L(\mathbf{x}^k, \boldsymbol{\lambda}^k) - L(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^k) \\ &= L_k(\mathbf{x}^k) - L_k(\mathbf{x}^{k+1}) = u_k(\mathbf{x}^k) - L_k(\mathbf{x}^{k+1}) \\ &\geq u_k(\mathbf{x}^k) - u_k(\mathbf{x}^{k+1}) = \frac{1}{2} (\nabla L_k(\mathbf{x}^k))^T Q_k^{-1} \nabla L_k(\mathbf{x}^k) \\ &= \frac{1}{2} \|Q_k^{-1} \nabla L_k(\mathbf{x}^k)\|_{Q_k}^2 = \frac{1}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{Q_k}^2 \geq m \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2, \end{aligned}$$

where the last equality follows from (4), and the last inequality holds as all the diagonal entries of Q_k are greater than $2m$. Therefore, $V(\mathbf{x}^{k+1}) \leq V(\mathbf{x}^k) - m \|\mathbf{x}^k - \mathbf{x}^{k+1}\|^2$. Since $V(\cdot)$ is lower bounded by a finite value, we get $\lim_{k \rightarrow \infty} \|\mathbf{x}^{k+1} - \mathbf{x}^k\| = 0$. That, in view of (4) and the fact that diagonal entries of Q_k^{-1} are lower bounded by $\frac{1}{2mn}$, implies $\lim_{k \rightarrow \infty} \nabla L_k(\mathbf{x}^k) = \mathbf{0}$.

Next, to show the convergence of the dynamics (2) in the case of convex measurements $g_{ij}(\cdot)$ and strictly convex functions $f_i(\cdot)$, we note that for any k , $L_k(\mathbf{x})$ belongs to the following *finite* family of strictly convex functions

$$\mathcal{H} := \left\{ \sum_i f_i(x_i) + \frac{1}{2} \sum_{i,j} \lambda_{ij} g_{ij}(x_i, x_j) : \lambda_{ij} \in \{0, 1\}, \forall i, j \right\},$$

which contains at most $O(2^{n^2})$ functions. The reason is that $L_k(\mathbf{x}) = L(\mathbf{x}, \boldsymbol{\lambda}^k)$, where $\boldsymbol{\lambda}^k$ is the solution of the linear program $\min_{\boldsymbol{\lambda} \in [0,1]^{n(n-1)}} L(\mathbf{x}^k, \boldsymbol{\lambda})$, and must be an extreme point of $[0, 1]^{n(n-1)}$. Now, given any $h(\mathbf{x}) \in \mathcal{H}$, let $h_1 < h_2 < \dots$, be all the indices k for which $L_k(\mathbf{x}) = h(\mathbf{x})$. Then we can partition the sequence $\{\mathbf{x}^k\}$ into at most $|\mathcal{H}|$ subsequences $\{\{\mathbf{x}^{h_\ell}\}_{\ell \geq 1}, h \in \mathcal{H}\}$. Since $\lim_{k \rightarrow \infty} \nabla L_k(\mathbf{x}^k) = \mathbf{0}$, that means that for any subsequence $\{\mathbf{x}^{h_\ell}\}$ we have, $\lim_{\ell \rightarrow \infty} \nabla h(\mathbf{x}^{h_\ell}) = \lim_{\ell \rightarrow \infty} \nabla L_{h_\ell}(\mathbf{x}^{h_\ell}) = \mathbf{0}$. As $h(\cdot)$ is a strictly convex function with a finite global minimum, the subsequence $\{\mathbf{x}^{h_\ell}\}_{\ell \geq 1}$ must converge to the unique minimizer of $h(\cdot)$, denoted by \mathbf{x}_h . Since there are a finite number of such subsequences, for any $\epsilon > 0$, there exists K_ϵ such that $\|\mathbf{x}^{h_\ell} - \mathbf{x}_h\| < \epsilon, \forall h \in \mathcal{H}, \ell > K_\epsilon$. Let $\mathcal{X} = \{\mathbf{x}_h = \operatorname{argmin} h(\mathbf{x}) : h \in \mathcal{H}\}$ be the finite set of minimizers of all the functions in \mathcal{H} , and choose $\epsilon := \frac{1}{3} \min_{\mathbf{x}_p \neq \mathbf{x}_q \in \mathcal{X}} \|\mathbf{x}_p - \mathbf{x}_q\|$. Then for $\ell > K_\epsilon$, each subsequence $\{\mathbf{x}^{h_\ell}\}_{\ell \geq 1}$ lies in an ϵ -neighborhood of its limit point \mathbf{x}_h , and, moreover, there is no jump of the iterates between two distinct ϵ -neighborhoods (Otherwise, $\|\mathbf{x}^{k+1} - \mathbf{x}^k\| > \frac{\epsilon}{3}$ for some k , contradicting the fact that $\lim_{k \rightarrow \infty} \|\mathbf{x}^{k+1} - \mathbf{x}^k\| = 0$.) Thus, for $\ell > K_\epsilon$, all the subsequences $\{\{\mathbf{x}^{h_\ell}\}_{\ell \geq 1}, h \in \mathcal{H}\}$ must lie in the same ϵ -neighborhood, and hence the sequence $\{\mathbf{x}^k\}$ converges to a limit point $\mathbf{x}^* \in \mathcal{X}$. \blacksquare

Example 1: Let $g_{ij}(x_i, x_j) = \frac{(x_i - x_j)^2}{2} - \frac{\epsilon^2}{2}$, $i \neq j$ be symmetric quadratic measurements and $f_i(x_i) = 0, \forall i \in [n]$. Clearly, these functions satisfy the statement of Theorem 3 with $m = 1$. By applying Theorem 3 directly to these functions, we obtain

$$x_i^{k+1} = x_i^k - \frac{\sum_{j \in N_i(\mathbf{x}^k)} (x_i^k - x_j^k)}{2(|N_i(\mathbf{x}^k)| + 1)} = \frac{|N_i| + 2}{2(|N_i| + 1)} x_i^k + \frac{1}{2} \frac{\sum_{j \in N_i(\mathbf{x}^k)} x_j^k}{|N_i(\mathbf{x}^k)| + 1}, \quad i \in [n],$$

which has been shown to be Lyapunov stable. However, the above dynamics are not exactly the homogeneous HK dynamics, but rather a “lazy” version of them wherein each agent puts a higher weight of (nearly) $\frac{1}{2}$ on its own state. The reason for losing a factor of $\frac{1}{2}$ for the *general* (nonquadratic) measurements is that the quadratic upper approximation in (3) may not be exact, and the cost of such approximation is reflected by an extra factor of $\frac{1}{2}$ in the underlying *nonlinear* dynamics (2). However, if the measurement functions are quadratic (as in the HK model), the quadratic approximation in (3) becomes exact, and one can skip the approximation step in (4) by directly computing the gradient and Hessian matrices in a closed form to show that $L_k(\mathbf{x}^{k+1}) \leq L_k(\mathbf{x}^k)$. More precisely, for quadratic measurements we have

$$L_k(\mathbf{x}) = L(\mathbf{x}^k) + (\mathbf{x} - \mathbf{x}^k)^T \nabla L_k(\mathbf{x}^k) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^k)^T \nabla^2 L_k(\mathbf{x}^k) (\mathbf{x} - \mathbf{x}^k),$$

with a closed-form gradient $\nabla L_k(\mathbf{x}^k) = (D_k - A_k)\mathbf{x}^k$ and Hessian $\nabla^2 L_k(\mathbf{x}^k) = D_k + A_k$, where A_k is the adjacency matrix of the communication network at state \mathbf{x}^k and $D_k = \text{diag}(|N_1(\mathbf{x}^k)|, \dots, |N_n(\mathbf{x}^k)|)$. If we take $Q_k = \text{diag}(|N_1(\mathbf{x}^k)| + 1, \dots, |N_n(\mathbf{x}^k)| + 1)$ (i.e., without an extra factor of 2) and note that $\mathbf{x}^k - \mathbf{x}^{k+1} = Q_k^{-1} \nabla L_k(\mathbf{x}^k)$, we get

$$\begin{aligned} L_k(\mathbf{x}^{k+1}) - L_k(\mathbf{x}^k) &= (\mathbf{x}^{k+1} - \mathbf{x}^k)^T \nabla L_k(\mathbf{x}^k) + \frac{1}{2} (\mathbf{x}^{k+1} - \mathbf{x}^k)^T (D_k + A_k) (\mathbf{x}^{k+1} - \mathbf{x}^k) \\ &= \frac{1}{2} (\mathbf{x}^{k+1} - \mathbf{x}^k)^T [D_k + A_k - 2Q_k] (\mathbf{x}^{k+1} - \mathbf{x}^k) \leq -\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2, \end{aligned}$$

where the last inequality holds because $D_k + A_k - 2Q_k = -2I - (D_k - A_k) \leq -2I$. Therefore, Lyapunov stability of the homogeneous HK dynamics

$$\mathbf{x}^{k+1} = \mathbf{x}^k - Q_k^{-1} \nabla L_k(\mathbf{x}^k) = Q_k^{-1} (I + A_k) \mathbf{x}^k$$

can be viewed as a special case of Theorem 3 for specific quadratic measurements with an associated Lyapunov function $V(\mathbf{x}) = \sum_{i,j} \min\{\frac{(x_i - x_j)^2}{2} - \frac{\epsilon^2}{2}, 0\}$.

Example 2: Let $\mathcal{G} = ([n], \mathcal{E})$ be a fixed undirected graph with a positive weight $a_{ij} = a_{ji} > 0$ on each edge $\{i, j\} \in \mathcal{E}$. (We set $a_{ij} = a_{ji} = 0$ if $\{i, j\} \notin \mathcal{E}$.) Assume that $\sum_{j \in N_i \cup \{i\}} a_{ij} = 1, \forall i$, where N_i denotes the fixed set of neighbors of agent i . Now let us define $f_i(x_i) = 0, \forall i \in [n]$. Moreover, let $K > 0$ be a very large constant. Consider the following symmetric measurements:

$$g_{ij}(x_i, x_j) = \begin{cases} \frac{a_{ij}}{2} (x_i - x_j)^2 - K, & \text{if } \{i, j\} \in \mathcal{E} \\ 1 & \text{if } \{i, j\} \notin \mathcal{E}. \end{cases}$$

Clearly, these functions satisfy the statement of Theorem 3. As K is chosen to be a very large number, regardless of the state \mathbf{x}^k of the dynamics (to be defined soon), we always have $N_i(\mathbf{x}^k) = \{j : g_{ij}(x_i^k, x_j^k) \leq 0\} = N_i$. If we leverage the quadratic structure of the measurement functions, and use the exact approach as in Example 1, we obtain $\nabla L_k(\mathbf{x}^k) = (I - D - A)\mathbf{x}^k$ and $\nabla^2 L_k(\mathbf{x}^k) = I - D + A$, where $A = (a_{ij})$ is the weighted adjacency matrix of the graph \mathcal{G} (with zero diagonal entries), and $D = \text{diag}(a_{11}, \dots, a_{nn})$ is the diagonal matrix of self-degrees. Therefore, $Q_k := I$ is a diagonal matrix

dominating the Hessian matrix (as $I - D + A \leq I$ by the Gershgorin Theorem and since $a_{ii} = 1 - \sum_{j \in N_i} a_{ij}, \forall i$). Thus, the dynamics

$$\mathbf{x}^{k+1} = \mathbf{x}^k - Q_k^{-1} \nabla L_k(\mathbf{x}^k) = \mathbf{x}^k - (I - D - A)\mathbf{x}^k = (D + A)\mathbf{x}^k$$

are Lyapunov stable with a Lyapunov function

$$V(x) = \frac{1}{2} \sum_{i,j} \min\{g_{ij}(x_i, x_j), 0\} = -|\mathcal{E}|K + \frac{1}{2} \sum_{\{i,j\} \in \mathcal{E}} a_{ij}(x_i - x_j)^2.$$

That is exactly the well-known Laplacian Lyapunov function for the conventional averaging dynamics $\mathbf{x}^{k+1} = (D + A)\mathbf{x}^k$, which is shifted by a constant $-|\mathcal{E}|K$. Hence, we recover the result of Theorem 1.

Example 3: In the proof of Theorem 3, we restricted our attention to quadratic upper approximations. However, motivated by the mirror descent algorithm from convex optimization [32], we can use any smooth convex mirror map $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}$ to construct an upper approximation for $L_k(\mathbf{x})$ at the point \mathbf{x}^k . By doing so, we obtain alternative state-dependent network dynamics whose Lyapunov stability and convergence can be established via an approach similar to that in Theorem 3. More precisely, Let $\Psi \in \mathcal{C}$ be a strictly convex function with $\nabla^2 \Psi(\mathbf{x}) \geq 2mnI$. Then, $u_k(\mathbf{x}) := L(\mathbf{x}^k) + (\mathbf{x} - \mathbf{x}^k)^T \nabla L_k(\mathbf{x}^k) + D_\Psi(\mathbf{x}, \mathbf{x}^k)$ serves as a convex upper approximation for $L_k(\mathbf{x})$; hence, if we update the state at the next time step to $\mathbf{x}^{k+1} = \arg\min_{\mathbf{x} \in \mathbb{R}^n} u_k(\mathbf{x})$, or equivalently to the solution of

$$\nabla \Psi(\mathbf{x}^{k+1}) = \nabla \Psi(\mathbf{x}^k) - \nabla L_k(\mathbf{x}^k), \quad (5)$$

that will guarantee a decrease in the Lyapunov $V(\mathbf{x}) = \sum_i f_i(x_i) + \frac{1}{2} \sum_{i,j} \min\{g_{ij}(\mathbf{x}), 0\}$. For instance, if we choose the mirror map to be the negative entropy function, i.e., $\Psi(\mathbf{x}) := \sum_{i=1}^n x_i \ln x_i$, and use (5), we obtain the following Lyapunov stable multiplicative dynamics:

$$x_i^{k+1} = x_i^k \cdot \exp\left(-\frac{\partial f(x_i)}{\partial x_i} - \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial g_{ij}(x_i^k, x_j^k)}{\partial x_i}\right).$$

A. Asymmetric State-Dependent Network Dynamics

Asymmetric (directed) interconnections among agents often introduce a significant challenge in the analysis of multiagent network dynamics. Unfortunately, the gradient operator is ‘‘symmetric,’’ meaning that fixing the network variable in the BCD method and updating the state variable in the negative direction of the gradient will always generate a symmetric class of averaging dynamics. However, one way to tackle that issue using sequential optimization is to introduce an independent copy of the state variable while making sure that the two copies remain close to each other. In other words, we capture the asymmetry between the agents by introducing an extra block variable into the BCD method and adding a penalty term (possibly asymmetric) to the objective function. That ensures that the two copies of the state variables remain close to each other. Here, the choice of the penalty function can be very problem-specific, resulting in different asymmetric state-dependent network dynamics. However, one natural choice for the penalty function is the Bregman divergence between the two copies of the state variables, as shown in the following theorem.

Theorem 4: Let $f_i(x_i) \in \mathcal{C}^2$ and $g_{ij}(x_i, x_j) \in \mathcal{C}^2$ be L -Lipschitz continuous functions such that $|\frac{\partial^2 g_{ij}}{\partial x_i \partial x_j}| \leq m, |\frac{\partial^2 f_i}{\partial x_i^2}| \leq m, \forall i, j$. If $\|\mathbf{y} - \mathbf{x}\|_1 \leq \frac{1}{nL} D_f(\mathbf{y}, \mathbf{x}), \forall \mathbf{x}, \mathbf{y}$, where $f(\mathbf{x}) = \sum_{i=1}^n f_i(x_i)$, then the dynamics

$$x_i^{k+1} = x_i^k - \frac{\sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k)}{m(|N_i(\mathbf{x}^k)| + 1)}, \quad i \in [n], \quad (6)$$

are Lyapunov stable with a Lyapunov function $V(\mathbf{x}) = \sum_{i,j} \min\{g_{ij}(x_i, x_j), 0\}$.

Proof: Let $c_i(\mathbf{x}, \boldsymbol{\lambda}_i) := \sum_{j \neq i} \lambda_{ij} g_{ij}(x_i, x_j)$ denote the cost of agent i with respect to its neighbors, and let \mathbf{y} be an independent copy of the state variable \mathbf{x} . Consider the following function with three independent block variables $\boldsymbol{\lambda} \in [0, 1]^{n(n-1)}$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$L(\mathbf{y}, \mathbf{x}, \boldsymbol{\lambda}) := \sum_{i=1}^n c_i(y_i, \mathbf{x}_{-i}, \boldsymbol{\lambda}_i) + D_f(\mathbf{y}, \mathbf{x}) = \sum_{i=1}^n \sum_{j \neq i} \lambda_{ij} g_{ij}(y_i, x_j) + D_f(\mathbf{y}, \mathbf{x}),$$

where $D_f(\mathbf{y}, \mathbf{x}) := f(\mathbf{y}) - f(\mathbf{x}) - (\mathbf{y} - \mathbf{x})^T \nabla f(\mathbf{x})$. Note that here we no longer require the symmetry assumption, so that in general, $g_{ij}(x_i, x_j) \neq g_{ji}(x_j, x_i)$. The reason for introducing the Bregman distance $D_f(\mathbf{y}, \mathbf{x})$ into the objective function is that it would be ideal if the two copies of the state variables coincide. But instead of adding the hard constraint $\mathbf{y} = \mathbf{x}$ into our optimization problem, we relax this constraint by adding a soft penalty term to the objective function. Now, let us apply the BCD method to the following minimization:

$$\min_{\boldsymbol{\lambda} \in [0, 1]^{n(n-1)}} \min_{\mathbf{x} \in \mathbb{R}^n} \min_{\mathbf{y} \in \mathbb{R}^n} \left\{ \sum_{i=1}^n \sum_{j \neq i} \lambda_{ij} g_{ij}(y_i, x_j) + D_f(\mathbf{x}, \mathbf{y}) \right\}. \quad (7)$$

First, assume that both state variables are fixed to $\mathbf{y} = \mathbf{x} = \mathbf{x}^k$. Then, if we minimize the objective function (7) with respect to $\boldsymbol{\lambda} \in [0, 1]^{n(n-1)}$, we precisely capture the asymmetric network structure $\boldsymbol{\lambda}^k$ associated with the state \mathbf{x}^k . (I.e., $\lambda_{ij}^k = 1$ if and only if $g_{ij}(x_i^k, x_j^k) \leq 0$.) Next, let us fix $\boldsymbol{\lambda} = \boldsymbol{\lambda}^k$ and $\mathbf{x} = \mathbf{x}^k$, and consider minimizing (7) with respect to the \mathbf{y} variable. However, to obtain a closed form for the optimal solution, instead of solving the minimization exactly, we minimize its quadratic upper approximation at the current state \mathbf{x}^k , given by

$$z_k(\mathbf{y}) := L(\mathbf{x}^k, \mathbf{x}^k, \boldsymbol{\lambda}^k) + (\mathbf{y} - \mathbf{x}^k)^T \nabla_{\mathbf{y}} L(\mathbf{x}^k, \mathbf{x}^k, \boldsymbol{\lambda}^k) + \frac{1}{2} (\mathbf{y} - \mathbf{x}^k)^T P_k (\mathbf{y} - \mathbf{x}^k),$$

where P_k is a diagonal matrix whose i th diagonal entry is $m(|N_i(\mathbf{x}^k)| + 1)$. To see why $L(\mathbf{y}, \mathbf{x}^k, \boldsymbol{\lambda}^k) \leq z_k(\mathbf{y})$, $\forall \mathbf{y}$, we note that

$$[\nabla_{\mathbf{y}} L(\mathbf{y}, \mathbf{x}^k, \boldsymbol{\lambda}^k)]_i = \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial y_i} g_{ij}(y_i, x_j^k) + \left(\frac{\partial}{\partial y_i} f_i(y_i) - \frac{\partial}{\partial x_i} f_i(x_i^k) \right),$$

which implies that $\nabla^2 L(\mathbf{y}, \mathbf{x}^k, \boldsymbol{\lambda}^k)$ is a *diagonal* matrix whose diagonal entries are

$$[\nabla^2 L(\mathbf{y}, \mathbf{x}^k, \boldsymbol{\lambda}^k)]_{ii} = \frac{\partial^2}{\partial y_i^2} f_i(y_i) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial g_{ij}(y_i, x_j^k)}{\partial^2 y_i}.$$

As $|\frac{\partial^2 f_i(y_i)}{\partial y_i^2}| \leq m$ and $|\frac{\partial g_{ij}(y_i, x_j^k)}{\partial y_i^2}| \leq m$, for all $i \in [n]$, the Hessian matrix is dominated by P_k , and the result follows from the Taylor expansion. Thus, the optimal solution to $\min_{\mathbf{y} \in \mathbb{R}^n} z_k(\mathbf{y})$ is given by $\mathbf{x}^k - P_k^{-1} \nabla_{\mathbf{y}} L(\mathbf{x}^k, \mathbf{x}^k, \boldsymbol{\lambda}^k)$, which is precisely the next state of the dynamics (6). Therefore, if the block variable \mathbf{y} is updated to \mathbf{x}^{k+1} , while the other variables are fixed to $\boldsymbol{\lambda} = \boldsymbol{\lambda}^k$, $\mathbf{x} = \mathbf{x}^k$, that will decrease the objective function because

$$L(\mathbf{x}^{k+1}, \mathbf{x}^k, \boldsymbol{\lambda}^k) \leq z_k(\mathbf{x}^{k+1}) = \min_{\mathbf{y} \in \mathbb{R}^n} z_k(\mathbf{y}) \leq z_k(\mathbf{x}^k) = L(\mathbf{x}^k, \mathbf{x}^k, \boldsymbol{\lambda}^k).$$

Finally, let us fix $\mathbf{y} = \mathbf{x}^{k+1}$, $\boldsymbol{\lambda} = \boldsymbol{\lambda}^k$ (which are the solutions to their corresponding sub-optimizations), and consider $\min_{\mathbf{x} \in \mathbb{R}^n} L(\mathbf{x}^{k+1}, \mathbf{x}, \boldsymbol{\lambda}^k)$. In particular, if we show that $L(\mathbf{x}^{k+1}, \mathbf{x}^{k+1}, \boldsymbol{\lambda}^k) \leq L(\mathbf{x}^{k+1}, \mathbf{x}^k, \boldsymbol{\lambda}^k)$, that will complete the BCD loop and imply that $L(\mathbf{y}, \mathbf{x}, \boldsymbol{\lambda})$ is decreasing along the trajectory of the

asymmetric dynamics (6). Here, the role of the penalty term in the objective function comes into play. More precisely, using the Lipschitz continuity and the fact that $D_f(\mathbf{x}^{k+1}, \mathbf{x}^{k+1}) = 0$, we can write,

$$\begin{aligned}
L(\mathbf{x}^{k+1}, \mathbf{x}^{k+1}, \boldsymbol{\lambda}^k) - L(\mathbf{x}^{k+1}, \mathbf{x}^k, \boldsymbol{\lambda}^k) &= -D_f(\mathbf{x}^{k+1}, \mathbf{x}^k) \\
&+ \sum_{i=1}^n \sum_{j \in N_i(\mathbf{x}^k)} \left(g_{ij}(x_i^{k+1}, x_j^{k+1}) - g_{ij}(x_i^{k+1}, x_j^k) \right) \\
&\leq -D_f(\mathbf{x}^{k+1}, \mathbf{x}^k) + \sum_{i,j} |g_{ij}(x_i^{k+1}, x_j^{k+1}) - g_{ij}(x_i^{k+1}, x_j^k)| \\
&\leq -D_f(\mathbf{x}^{k+1}, \mathbf{x}^k) + nL \sum_j |x_j^{k+1} - x_j^k| \\
&= -D_f(\mathbf{x}^{k+1}, \mathbf{x}^k) + nL \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_1 < 0,
\end{aligned}$$

where the last inequality follows from the assumption on the choice of the Bregman map $f(\cdot)$. Therefore, $V(\mathbf{x}) := \min_{\boldsymbol{\lambda} \in [0,1]^{n(n-1)}} L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{i,j} \min\{g_{ij}(x_i, x_j), 0\}$ is a decreasing function along the trajectories of the asymmetric dynamics (6). \blacksquare

B. BCD Method with Change of Variables

In this section, we show how a suitable change of block variables in the BCD method can generate new state-dependent network dynamics, whose Lyapunov stability can be established using the same approach as before. The change of variable can be applied to either the state or the network variable. However, in this section, we focus on a more compelling case, in which the change of variable is applied on the network variable; we only illustrate the idea of a change of variable for the state through the following simple example.

Example 4: Let us recall the homogeneous HK model in which the state of agent i at the next time step is updated to $x_i^{k+1} = \frac{x_i^k + \sum_{j \in N_i(\mathbf{x}^k)} x_j^k}{1 + |N_i(\mathbf{x}^k)|}$, $i \in [n]$, where the neighborhood set of agent i is given by $N_i(\mathbf{x}^k) = \{j \in [n] \setminus \{i\} : |x_i^k - x_j^k| \leq \epsilon\}$. As the dynamics are invariant with respect to translation of the states, without loss of generality, we may assume that $x_i^k > 1, \forall i, k$. Now let us define a new state variable by setting $y_i := \ln(x_i), i \in [n]$. If we apply the HK dynamics on those new logarithmic states, we obtain $y_i^{k+1} = \frac{y_i^k + \sum_{j \in N_i(\mathbf{y}^k)} y_j^k}{1 + |N_i(\mathbf{y}^k)|}$, with $N_i(\mathbf{y}^k) = \{j \in [n] \setminus \{i\} : |y_i^k - y_j^k| \leq \epsilon\}$, which are Lyapunov stable and converge to an equilibrium. If we rewrite the dynamics in terms of x variables, we obtain a new class of state-dependent geometric averaging dynamics $x_i^{k+1} = \left(\prod_{j \in \bar{N}_i(\mathbf{x}^k)} x_j^k \right)^{\frac{1}{|\bar{N}_i(\mathbf{x}^k)|}}$, with a new definition of a neighborhood set $\bar{N}_i(\mathbf{x}^k) = \{j \in [n] : e^{-\epsilon} \leq \frac{x_i^k}{x_j^k} \leq e^\epsilon\}$, that are also Lyapunov stable and converge.

Next, we turn our attention to the case of changing the network variables. So far, the dual variable λ_{ij} in the Lagrangian function $L(\mathbf{x}, \boldsymbol{\lambda})$ has been used to capture the existence of an edge from agent i to agent j . In particular, we saw that restricting λ_{ij} to the unit interval $[0, 1]$ and minimizing $L(\mathbf{x}, \boldsymbol{\lambda})$ for the network variable in the BCD method would automatically force λ_{ij} to take binary values in $\{1, 0\}$ (and hence capture the switching behavior of an edge formation between i and j). In particular, if we fix the state block variable in $L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{i,j} \lambda_{ij} g_{ij}(x_i, x_j)$ and minimize it with respect to $\boldsymbol{\lambda} \in [0, 1]^{n(n-1)}$, that will give us the network structure for that state, i.e., $\lambda_{ij}^* = 1 - \text{sgn}(g_{ij}(x_i, x_j))$, where $\text{sgn}(\cdot)$ is the sign function. This relation will give us a complicated characterization of λ_{ij}^* that is a combination of the sign function and the measurement function. An alternative way to recover the same network structure would start by transforming the original network variables from λ_{ij} to $f_{ij}(\lambda_{ij}) := 1 - \text{sgn}(\lambda_{ij})$, which would cause the transformed Lagrangian function to become $\hat{L}(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{i,j} f_{ij}(\lambda_{ij}) g_{ij}(x_i, x_j)$. If we then apply the BCD method to $\hat{L}(\mathbf{x}, \boldsymbol{\lambda})$ by fixing the state variable and optimizing $\hat{L}(\mathbf{x}, \boldsymbol{\lambda})$ with

respect to the *unconstrained* network variable $\lambda_{ij} \in \mathbb{R}$, we obtain a simpler optimal network variable $\hat{\lambda}_{ij} = g_{ij}(x_i, x_j)$. Note that

$$\hat{L}(\mathbf{x}, \hat{\boldsymbol{\lambda}}) = \min_{\boldsymbol{\lambda} \in \mathbb{R}^{n(n-1)}} \hat{L}(\mathbf{x}, \boldsymbol{\lambda}) = \min_{\boldsymbol{\lambda} \in [0,1]^{n(n-1)}} L(\mathbf{x}, \boldsymbol{\lambda}) = L(\mathbf{x}, \boldsymbol{\lambda}^*).$$

Such a transformation of the network variables has three advantages: i) it removes the box constraints on the network variables which cause switching and absorbs them into the structure of the transformation function; ii) the optimal network variable in the BCD method has a simpler form after the change of variable; and iii) by choosing different transfer functions, one can obtain different types of state-dependent network dynamics. The following theorem provides a sample result using the idea of a change of network variables.

Theorem 5: Let $g_{ij}(x_i, x_j) \in \mathcal{C}^2$ be symmetric functions such that $|\frac{\partial^2 g_{ij}}{\partial x_i \partial x_j}| \leq m, \forall i, j$. Moreover, let $f_{ij}(\lambda) \in \mathcal{C}$ be symmetric and nonnegative decreasing functions. Then the state-dependent network dynamics

$$x_i^{k+1} = x_i^k - \frac{\sum_j f_{ij}(g_{ij}(x_i^k, x_j^k)) \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k)}{2m \sum_j f_{ij}(g_{ij}(x_i^k, x_j^k))}, \quad i \in [n] \quad (8)$$

admit a Lyapunov function $V(\mathbf{x}) = \sum_{i \neq j} \int_0^{g_{ij}(x_i, x_j)} f_{ij}(\lambda) d\lambda$. Moreover, for each k there exists a positive-definite matrix Q_k such that $V(\mathbf{x}^k) - V(\mathbf{x}^{k+1}) \geq \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{Q_k}^2$, and $\lim_{k \rightarrow \infty} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{Q_k} = 0$.

Proof: Let us consider the BCD method applied to the transformed Lagrangian

$$\hat{L}(\mathbf{x}, \boldsymbol{\lambda}) := \sum_{i \neq j} \left(f_{ij}(\lambda_{ij}) g_{ij}(x_i, x_j) - h_{ij}(\lambda_{ij}) \right),$$

with $\mathbf{x} \in \mathbb{R}^n$ and $\boldsymbol{\lambda} \in \mathbb{R}^{n(n-1)}$, for some real-valued functions $h_{ij}(\lambda_{ij})$ to be determined later. If for any fixed state \mathbf{x} , the function $\hat{L}(\mathbf{x}, \boldsymbol{\lambda})$ has a unique minimum with respect to $\boldsymbol{\lambda} \in \mathbb{R}^{n(n-1)}$, we can apply the BCD method and be assured that the function will decrease because of the network updates. Now let us first fix the state variable to \mathbf{x}^k . Assuming that the functions f_{ij}, h_{ij} are differentiable, to find $\operatorname{argmin}_{\boldsymbol{\lambda} \in \mathbb{R}^{n(n-1)}} \hat{L}(\mathbf{x}^k, \boldsymbol{\lambda})$, we set

$$\frac{\partial}{\partial \lambda_{ij}} \hat{L}(\mathbf{x}^k, \boldsymbol{\lambda}) = f'_{ij}(\lambda_{ij}) g_{ij}(x_i^k, x_j^k) - h'_{ij}(\lambda_{ij}) = 0, \quad (9)$$

which implies $\frac{h'_{ij}(\lambda_{ij})}{f'_{ij}(\lambda_{ij})} = g_{ij}(x_i^k, x_j^k)$. Therefore, if we define $h'_{ij}(\lambda) := \lambda f'_{ij}(\lambda)$, or, equivalently, $h_{ij}(\lambda) := \int_0^\lambda s f'_{ij}(s) ds$, equation (9) has a unique solution $\lambda_{ij}^* = g_{ij}(x_i^k, x_j^k)$. To show that this solution is the minimizer of $\hat{L}(\mathbf{x}^k, \boldsymbol{\lambda})$, we note that

$$\frac{\partial}{\partial \lambda_{ij}} \hat{L}(\mathbf{x}^k, \boldsymbol{\lambda}) = f'_{ij}(\lambda_{ij}) \left(g_{ij}(x_i^k, x_j^k) - \lambda_{ij} \right).$$

Since $f_{ij}(\cdot)$ is a decreasing function, $f'_{ij}(\lambda_{ij}) < 0$. Thus, for $\lambda_{ij} \leq g_{ij}(x_i^k, x_j^k)$, the function $\hat{L}(\mathbf{x}^k, \boldsymbol{\lambda})$ is decreasing with respect to λ_{ij} , and for $\lambda_{ij} \geq g_{ij}(x_i^k, x_j^k)$, it is increasing. (Note that $\hat{L}(\mathbf{x}^k, \boldsymbol{\lambda})$ is splittable over its $\boldsymbol{\lambda}$ -components, so we can analyze each of its summands separately.) Thus, given a fixed state \mathbf{x}^k , the unique global minimum of $\hat{L}(\mathbf{x}^k, \boldsymbol{\lambda})$ is obtained at $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$, where $\lambda_{ij}^* = g_{ij}(x_i^k, x_j^k)$.

The rest of the proof follows along the same analysis used in Theorem 3, and we only sketch it here. Let us fix the network variable to $\boldsymbol{\lambda}^*$, and consider $\min_{\mathbf{x} \in \mathbb{R}^n} \hat{L}(\mathbf{x}, \boldsymbol{\lambda}^*)$. To find a minimizer we use an

inexact method by using a quadratic upper approximation of $\hat{L}(\mathbf{x}, \boldsymbol{\lambda}^*)$ at \mathbf{x}^k . The i th component of the gradient of $\hat{L}(\mathbf{x}, \boldsymbol{\lambda}^*)$ at \mathbf{x}^k is given by

$$\begin{aligned} [\nabla_{\mathbf{x}} \hat{L}(\mathbf{x}^k, \boldsymbol{\lambda}^*)]_i &= \sum_j \left(f_{ij}(\lambda_{ij}^*) \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k) + f_{ji}(\lambda_{ji}^*) \frac{\partial}{\partial x_i} g_{ji}(x_j^k, x_i^k) \right) \\ &= 2 \sum_j f_{ij}(\lambda_{ij}^*) \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k) \\ &= 2 \sum_j f_{ij}(g_{ij}(x_i^k, x_j^k)) \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k), \end{aligned}$$

where the second equality is by the symmetry of f_{ij}, g_{ij} . The Hessian matrix equals

$$[\nabla^2 \hat{L}(\mathbf{x}, \boldsymbol{\lambda}^*)]_{ij} = \begin{cases} 2 \sum_j f_{ij}(g_{ij}(x_i^k, x_j^k)) \frac{\partial^2 g_{ij}(x_i, x_j)}{\partial x_i^2} & \text{if } j = i \\ 2 f_{ij}(g_{ij}(x_i^k, x_j^k)) \frac{\partial^2 g_{ij}(x_i, x_j)}{\partial x_i \partial x_j} & \text{if } j \neq i, \end{cases}$$

which is dominated by a diagonal matrix Q_k whose i th diagonal entry is given by $[Q_k]_{ii} = 2 + 4m \sum_j f_{ij}(g_{ij}(x_i^k, x_j^k))$. Thus, the optimal solution to the quadratic upper approximation of $\hat{L}(\mathbf{x}, \boldsymbol{\lambda}^*)$ at the point $\mathbf{x} = \mathbf{x}^k$ is given by $\mathbf{x}^k - Q_k^{-1} \nabla_{\mathbf{x}} \hat{L}(\mathbf{x}^k, \boldsymbol{\lambda}^*)$, which in view of (8) equals to \mathbf{x}^{k+1} . As a result, $\hat{L}(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^*) < \hat{L}(\mathbf{x}^k, \boldsymbol{\lambda}^*)$, and

$$\begin{aligned} V(\mathbf{x}) &= \min_{\boldsymbol{\lambda} \in \mathbb{R}^{n(n-1)}} \hat{L}(\mathbf{x}, \boldsymbol{\lambda}) = \sum_{i \neq j} \left(f_{ij}(g_{ij}(x_i, x_j)) g_{ij}(x_i, x_j) - \int_0^{g_{ij}(x_i, x_j)} \lambda f'_{ij}(\lambda) d\lambda \right) \\ &= \sum_{i \neq j} \int_0^{g_{ij}(x_i, x_j)} f_{ij}(\lambda) d\lambda \end{aligned}$$

serves as a Lyapunov function for the dynamics (8), and the last equality is obtained using integration by parts. Now, as in Theorem 3, we can bound the drift of this Lyapunov function at time k as

$$\begin{aligned} V(\mathbf{x}^k) - V(\mathbf{x}^{k+1}) &\geq \hat{L}(\mathbf{x}^k, \boldsymbol{\lambda}^*) - \hat{L}(\mathbf{x}^{k+1}, \boldsymbol{\lambda}^*) \\ &\geq \frac{1}{2} (\nabla_{\mathbf{x}} \hat{L}(\mathbf{x}^k, \boldsymbol{\lambda}^*))^T Q_k^{-1} \nabla_{\mathbf{x}} \hat{L}(\mathbf{x}^k, \boldsymbol{\lambda}^*) \\ &= \frac{1}{2} \|Q_k^{-1} \nabla_{\mathbf{x}} \hat{L}(\mathbf{x}^k, \boldsymbol{\lambda}^*)\|_{Q_k}^2 \\ &= \frac{1}{2} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{Q_k}^2. \end{aligned}$$

Finally, $V(\cdot)$ is a nonnegative function due to the nonnegativity of $f_{ij}(\cdot)$. Therefore, if we sum the above relations for $k = 1, 2, \dots$, we get $\sum_{k=1}^{\infty} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{Q_k}^2 \leq V(\mathbf{x}^0) < \infty$. This implies that $\lim_{k \rightarrow \infty} \|\mathbf{x}^k - \mathbf{x}^{k+1}\|_{Q_k} = 0$. \blacksquare

Remark 2: In fact, the differentiability of the transfer functions f_{ij} in the above theorem can be further relaxed to any nonnegative decreasing symmetric functions f_{ij} . In particular, Theorem 5 is a heterogeneous extension of [33, Corollary 1] (see also [34]) when $f_{ij}(\lambda) = f(\sqrt{\lambda}), \forall i, j$ and $g_{ij}(x_i, x_j) = (x_i - x_j)^2$.

IV. STABILITY USING DISCRETE-TIME SADDLE-POINT DYNAMICS

In the previous section, we considered the stability of state-dependent network dynamics when the network structure and agents' states are aligned with each other. More precisely, in the application of the BCD method on the Lagrangian function $L(\mathbf{x}, \boldsymbol{\lambda})$, we considered a double minimization problem

$\min_{\mathbf{x}} \min_{\boldsymbol{\lambda}} L(\mathbf{x}, \boldsymbol{\lambda})$, which essentially means that the network coordinator (viewed as a *network player*) breaks/adds the links in favor of the agents' states (viewed as a *state player*). In essence, that means that there is no conflict between the network and state players, as they are both minimizing the same Lagrangian function. But what if the network and state players have conflicting objectives? In that case, we have a 2-player zero-sum game between the network and the state with the payoff function $L(\mathbf{x}, \boldsymbol{\lambda})$, so that the network player aims to maximize the payoff function while the state player aims to minimize it, i.e., $\min_{\mathbf{x}} \max_{\boldsymbol{\lambda}} L(\mathbf{x}, \boldsymbol{\lambda})$.

To model such conflicting behavior, we assume that each agent $i \in [n]$ holds $n-1$ *convex* measurement functions $g_{ij}(x_i, x_j), j \in [n] \setminus \{i\}$. In this section, we restrict our attention to symmetric measurement functions; however, for asymmetric measurement functions, results similar to those in Theorem 4 can be obtained. For a given state \mathbf{x} , two agents i and j become each other's neighbors if $g_{ij}(x_i, x_j) \geq 0$. (Note that the logic constraint is now the reverse of what it was in the previous section.) Intuitively, an edge is formed between two agents i and j if and only if their states are far from each other. Now let us consider the following convex program:

$$\begin{aligned} \min \quad & f(\mathbf{x}) := \sum_{i=1}^n f_i(x_i) \\ \text{s.t.} \quad & \frac{1}{2}g_{ij}(x_i, x_j) \leq 0, \quad \forall i \neq j, \\ & \mathbf{x} \in \mathbb{R}^n, \end{aligned} \tag{10}$$

where $f_i(x_i), i \in [n]$ are agents' private convex functions.³ To solve this problem, one can form the Lagrangian function $L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \frac{1}{2} \sum_{i,j} \lambda_{ij} g_{ij}(x_i, x_j)$, and solve the following saddle-point problem: $\min_{\mathbf{x} \in \mathbb{R}^n} \max_{\boldsymbol{\lambda} \geq 0} L(\mathbf{x}, \boldsymbol{\lambda})$.⁴

By using KKT optimality conditions, we know that if the constraint $g_{ij}(x_i, x_j) \leq 0$ is satisfied but not *tight* (i.e., $g_{ij}(x_i, x_j) < 0$), then the corresponding optimal dual variable must be zero, i.e., $\lambda_{ij} = 0$. If the dual variables are viewed as network variables, that means that there is no edge between the agents i and j , and that is consistent with the logical condition of not having an edge between i and j . On the other hand, if the constraint $g_{ij}(x_i, x_j) \leq 0$ is not satisfied (i.e., $g_{ij}(x_i, x_j) > 0$), then one must set the corresponding dual variable to $\lambda_{ij} = \infty$ to maximize $\max_{\boldsymbol{\lambda} \geq 0} L(\mathbf{x}, \boldsymbol{\lambda})$. However, if the dual variables are upper-bounded by 1, to achieve the maximum value in $\max_{\boldsymbol{\lambda} \in [0,1]^{n(n-1)}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$, we must set λ_{ij} to its upper bound, i.e., $\lambda_{ij} = 1$. That is again consistent with the logical condition of having an edge between i and j . These facts together suggest that the network switches that may occur during the update process of state-dependent network dynamics are merely the KKT optimality conditions that guide the iterates to the optimal solution of (10), assuming that there is a budget constraint on the dual variables. In other words, if the dual variables could have been freely chosen from $[0, \infty)$, then the iterates of the dynamics would converge to the optimal solution of (10). However, the budget constraints on the dual variables do not allow us to penalize the violated constraints arbitrarily large and force them to be feasible. Therefore, the solutions that are obtained from state-network updates may not necessarily generate a feasible solution to (10). Nevertheless, this approach allows us to view the state-network dynamics as an iterative primal-dual algorithm guided by KKT optimality conditions for solving a saddle-point problem with box constraints on the dual variables. Alternatively, the state-dependent network dynamics can be viewed as Nash dynamics in a zero-sum game between a network player and a state player, with budget constraints on the action set of the network player. In the following, we use these observations to develop Lyapunov stable and convergent state-dependent network dynamics via static saddle-point problems.

³Here, each constraint is scaled by $\frac{1}{2}$ without causing changes to the feasible set.

⁴To do so, one must find a solution $(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})$ such that $L(\bar{\mathbf{x}}, \boldsymbol{\lambda}) \leq L(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}}) \leq L(\mathbf{x}, \bar{\boldsymbol{\lambda}}), \forall \mathbf{x}, \forall \boldsymbol{\lambda} \geq \mathbf{0}$.

Theorem 6: Let $g_{ij}(x_i, x_j) \in \mathcal{C}$ be a symmetric convex function and $f_i(x_i) \in \mathcal{C}$ be a convex function. Assume that agent $i \in [n]$ updates its state as

$$x_i^{k+1} = x_i^k - \alpha^k \left(\frac{\partial}{\partial x_i} f_i(x_i^k) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k) \right), \quad (11)$$

where $N_i(\mathbf{x}^k) = \{j : g_{ij}(x_i^k, x_j^k) > 0\}$ denotes the set of neighbors of agent i at time k . Then, for any positive sequence $\alpha^k = \gamma_k [\sum_i (\frac{\partial f_i(x_i^k)}{\partial x_i} + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial g_{ij}(x_i^k, x_j^k)}{\partial x_i})^2]^{-\frac{1}{2}}$, with $\lim_k \gamma_k = 0$ and $\sum_k \gamma_k = \infty$, the dynamics (11) converge to an equilibrium \mathbf{x}^* . Moreover, for sufficiently small α^k , $V(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}^*\|^2$ serves as a Lyapunov function.

Proof: Let us consider the Lagrangian function

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \frac{1}{2} \sum_{i,j} \lambda_{ij} g_{ij}(x_i, x_j),$$

where $f(\mathbf{x}) = \sum_{i=1}^n f_i(x_i)$, and suppose that we want to solve the following saddle-point problem with box constraints on the dual variables:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^n} \max_{\boldsymbol{\lambda} \in [0,1]^{n(n-1)}} L(\mathbf{x}, \boldsymbol{\lambda}) &= \min_{\mathbf{x} \in \mathbb{R}^n} \max_{\boldsymbol{\lambda} \in [0,1]^{n(n-1)}} \left\{ f(\mathbf{x}) + \frac{1}{2} \sum_{i,j} \lambda_{ij} g_{ij}(x_i, x_j) \right\} \\ &= \min_{\mathbf{x} \in \mathbb{R}^n} \left\{ f(\mathbf{x}) + \frac{1}{2} \sum_{i,j} \max\{g_{ij}(x_i, x_j), 0\} \right\}. \end{aligned}$$

If we define $\Phi(\mathbf{x}) := f(\mathbf{x}) + \frac{1}{2} \sum_{i,j} \max\{g_{ij}(x_i, x_j), 0\}$ and note that for any i, j , $\max\{g_{ij}(x_i, x_j), 0\}$ is a convex function, we can easily see that $\Phi(\mathbf{x})$ is a convex function of \mathbf{x} . Therefore, if a subgradient algorithm is applied to the unconstrained convex problem $\min_{\mathbf{x} \in \mathbb{R}^n} \Phi(\mathbf{x})$ with an appropriate choice of step sizes $\alpha^k, k = 1, 2, \dots$, the generated sequence will converge to a minimizer of $\Phi(\mathbf{x})$ denoted by \mathbf{x}^* . More precisely, let us use g^k to denote the subgradient of $\Phi(\mathbf{x})$ at \mathbf{x}^k . Then, it is known that the discrete time dynamics

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha_k g^k, \quad (12)$$

with diminishing step length $\alpha^k = \frac{\gamma_k}{\|g^k\|}$ with $\lim_k \gamma_k = 0$ and $\sum_k \gamma_k = \infty$, will converge to \mathbf{x}^* [35]. Now, let $J = \{(r, s) : g_{rs}(x_r^k, x_s^k) > 0\}$ and $\bar{J} = \{(r, s) : g_{rs}(x_r^k, x_s^k) \leq 0\}$. Then, for every $(r, s) \in J$, the function $\max\{g_{rs}(x_r, x_s), 0\}$ has a unique subgradient at \mathbf{x}^k , which is $\nabla g_{rs}(x_r^k, x_s^k)$. Moreover, for every $(r, s) \in \bar{J}$, the minimum of the convex function $\max\{g_{rs}(x_r, x_s), 0\}$ equals 0 and is achieved at \mathbf{x}^k . Thus, $\mathbf{0}$ is a subgradient of $\max\{g_{rs}(x_r, x_s), 0\}$ at \mathbf{x}^k for every $(i, j) \in \bar{J}$. Through use of the additivity rule of the subgradient, we conclude that $g^k = \nabla f(\mathbf{x}) + \frac{1}{2} \sum_{(r,s) \in J} \nabla g_{rs}(x_r^k, x_s^k)$ is a subgradient for $\Phi(\cdot)$ at \mathbf{x}^k . In particular, the i th component of g^k is given by

$$\begin{aligned} g_i^k &= \frac{\partial}{\partial x_i} f(\mathbf{x}^k) + \frac{1}{2} \sum_j \left(\mathbf{1}_{\{g_{ij}(x_i^k, x_j^k) > 0\}} \frac{\partial g_{ij}(x_i^k, x_j^k)}{\partial x_i} + \mathbf{1}_{\{g_{ji}(x_j^k, x_i^k) > 0\}} \frac{\partial g_{ji}(x_j^k, x_i^k)}{\partial x_i} \right) \\ &= \frac{\partial}{\partial x_i} f_i(x_i^k) + \sum_j \left(\mathbf{1}_{\{g_{ij}(x_i^k, x_j^k) > 0\}} \cdot \frac{\partial g_{ij}(x_i^k, x_j^k)}{\partial x_i} \right) \\ &= \frac{\partial}{\partial x_i} f_i(x_i^k) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial g_{ij}(x_i^k, x_j^k)}{\partial x_i}, \end{aligned} \quad (13)$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function. In the above relations, the second equality holds by symmetry of the functions $g_{ij}(x_i, x_j) = g_{ji}(x_j, x_i)$, and the last equality is due to the definition of an edge emergence between two nodes i and j . By substituting (13) into (12), we obtain the desired dynamics (11).

Finally, we can use the definition of the subgradient to write

$$\begin{aligned}\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 &= \|\mathbf{x}^k - \mathbf{x}^* - \alpha^k g^k\|^2 \\ &= \|\mathbf{x}^k - \mathbf{x}^*\|^2 + (\alpha^k)^2 \|g^k\|^2 - 2\alpha^k (g^k)^T (\mathbf{x}^k - \mathbf{x}^*) \\ &\leq \|\mathbf{x}^k - \mathbf{x}^*\|^2 + (\alpha^k)^2 \|g^k\|^2 - 2\alpha^k (\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*)).\end{aligned}$$

Therefore, for any $\alpha^k \in [0, \frac{2(\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*))}{\|g^k\|^2}]$, we have $\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^k - \mathbf{x}^*\|^2$. \blacksquare

Remark 3: Let X be the set of minimizers of $\min_{\mathbf{x} \in \mathbb{R}^n} \Phi(\mathbf{x})$, which is a nonempty closed convex set. Moreover, let $d(\mathbf{x}, X) = \|\mathbf{x} - \Pi_X[\mathbf{x}]\|$ be the minimum distance of the point \mathbf{x} from the set X , where $\Pi_X[\mathbf{x}]$ is the projection of \mathbf{x} on the set X . As for any $\mathbf{x}^* \in X$, $\|\mathbf{x}^{k+1} - \mathbf{x}^*\|^2 \leq \|\mathbf{x}^k - \mathbf{x}^*\|^2$, by choosing $\mathbf{x}^* = \Pi_X[\mathbf{x}^k]$ we obtain

$$d(\mathbf{x}^{k+1}, X) = \|\mathbf{x}^{k+1} - \Pi_X[\mathbf{x}^{k+1}]\| \leq \|\mathbf{x}^{k+1} - \Pi_X[\mathbf{x}^k]\| \leq \|\mathbf{x}^k - \Pi_X[\mathbf{x}^k]\| = d(\mathbf{x}^k, X).$$

Thus, for a sufficiently small step size $\alpha^k \in [0, \frac{2(\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*))}{\|g^k\|^2}]$, the distance from the iterates (11) to the optimal set X also serves as a Lyapunov function.

Example 5: An interesting special case of Theorem 6 is when $f_i(x_i) = 0, \forall i$, and the set of constraints $\{g_{ij}(x_i, x_j) \leq 0, \forall i, j\}$ is feasible. In this case the set of minimizers of $\Phi(\mathbf{x}) = \sum_{i,j} \max\{\frac{g_{ij}(x_i, x_j)}{2}, 0\}$ is precisely the feasible set $\{\mathbf{x} \in \mathbb{R}^n : g_{ij}(x_i, x_j) \leq 0, \forall i, j\}$. In particular, the minimum value of $\Phi(\cdot)$ is zero which is obtained at any feasible point $\mathbf{x}^* \in \{\mathbf{x} \in \mathbb{R}^n : g_{ij}(x_i, x_j) \leq 0, \forall i, j\}$. Now if the norm of the gradient of each measurement function g_{ij} is bounded above by a constant G , we can write,

$$\begin{aligned}\frac{2(\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*))}{\|g^k\|^2} &= \frac{2\Phi(\mathbf{x}^k)}{\|g^k\|^2} = \frac{\sum_i \sum_{j \in N_i(\mathbf{x}^k)} g_{ij}(x_i^k, x_j^k)}{\sum_i (\sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k))^2} \\ &\geq \frac{\sum_i \sum_{j \in N_i(\mathbf{x}^k)} g_{ij}(x_i^k, x_j^k)}{n \sum_{i,j} (\frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k))^2} \geq \frac{\max_{i,j} \{g_{ij}(x_i^k, x_j^k)\}}{n^2 G^2}.\end{aligned}\quad (14)$$

Let us define the ϵ -equilibrium set as the set of all the points for which each constraint $g_{ij}(x_i, x_j) \leq 0$ is violated by at most ϵ , i.e., $X_\epsilon = \{\mathbf{x} \in \mathbb{R}^n : g_{ij}(x_i, x_j) \leq \epsilon, \forall i, j\}$, and consider the dynamics (11) with the constant step size $\alpha_k = \frac{\epsilon}{n^2 G^2}$. Then, if $\mathbf{x}^k \notin X_\epsilon$, we have $\max_{i,j} \{g_{ij}(x_i^k, x_j^k)\} > \epsilon$, which, in view of (14), implies that $\alpha^k \in [0, \frac{2(\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^*))}{\|g^k\|^2}]$. Thus, as long as $\mathbf{x}^k \notin X_\epsilon$, $d(\mathbf{x}, X)$ serves as a Lyapunov function, and we can write

$$\begin{aligned}d(\mathbf{x}^{k+1}, X) &\leq d(\mathbf{x}^k, X) + (\alpha^k)^2 \|g^k\|^2 - 2\alpha^k \Phi(\mathbf{x}^k) \\ &= d(\mathbf{x}^k, X) + \frac{\epsilon^2}{n^4 G^4} \|g^k\|^2 - \frac{2\epsilon}{n^2 G^2} \Phi(\mathbf{x}^k) \\ &\leq d(\mathbf{x}^k, X) + \frac{\epsilon^2}{n^4 G^4} n^2 G^2 - \frac{2\epsilon}{n^2 G^2} \epsilon = d(\mathbf{x}^k, X) - \frac{\epsilon^2}{n^2 G^2},\end{aligned}$$

where in the last inequality we have used the fact that $\|g^k\|^2 \leq n^2 G^2$ and $\Phi(\mathbf{x}^k) \geq \epsilon$ (as $\mathbf{x}^k \notin X_\epsilon$). Since $d(\mathbf{x}^k, X) \geq 0, \forall k$, we conclude that after at most $\frac{d(\mathbf{x}^0, X) n^2 G^2}{\epsilon^2}$ iterations, the state-dependent network dynamics $x_i^{k+1} = x_i^k - \frac{\epsilon}{n^2 G^2} \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k)$ will reach an ϵ -neighborhood of its equilibrium set X_ϵ .

A. Saddle-Point Dynamics with Heterogeneous Step Size

The subgradient method is not the only algorithm for minimizing a convex function, and one can consider other algorithms that can result in different state-dependent network dynamics. The following theorem provides another multiagent network dynamics motivated by the fact that different agents often have different scaling parameters in their update rules. These dynamics can be viewed as the *quasi-Newton* method [35] in the context of multiagent network dynamics.

Definition 3: A function $V : \mathbb{R}^n \rightarrow \mathbb{R}$ is called a *semi-Lyapunov function* for the discrete-time dynamics $\mathbf{z}^{k+1} = h(\mathbf{z}^k)$, $k = 0, 1, 2, \dots$, if $V(\mathbf{z}^{k+1}) < V(\mathbf{z}^k)$ for any $\mathbf{z}^k \in \mathbb{R}^m \setminus D$, where D is a measure-zero subset of \mathbb{R}^n .

Theorem 7: Let $f \in \mathcal{C}$ be a convex function and $g_{ij} \in \mathcal{C}^2$ be a symmetric convex function whose zeros form a measure-zero subset D . Let $\Phi(\mathbf{x}) = f(\mathbf{x}) + \frac{1}{2} \sum_{i,j} \max\{g_{ij}(x_i, x_j), 0\}$, whose level set and subgradient at point \mathbf{x} are given by $L_{\mathbf{x}} = \{\mathbf{y} : \Phi(\mathbf{y}) \leq \Phi(\mathbf{x})\}$ and $g^{\mathbf{x}}$, respectively. If for any $\mathbf{x} \notin D$, there exists a positive-definite diagonal matrix $G^{\mathbf{x}}$ such that $\Phi(\mathbf{y}) \leq \Phi(\mathbf{x}) + (\mathbf{y} - \mathbf{x})^T g^{\mathbf{x}} + \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_{G^{\mathbf{x}}}^2$, $\forall \mathbf{y} \in L_{\mathbf{x}}$, then the dynamics

$$x_i^{k+1} = x_i^k - \frac{1}{G_{ii}^k} \left(\frac{\partial}{\partial x_i} f(\mathbf{x}^k) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k) \right), \quad i \in [n], \quad (15)$$

admit the semi-Lyapunov function $\Phi(\mathbf{x})$. Moreover, if $G^k \leq mI, \forall k$, and $\mathbf{x}^k \in D$ for at most finitely many iterates k , then the dynamics (15) will converge.

Proof: Consider the convex function $\Phi(\mathbf{x}) = f(\mathbf{x}) + \frac{1}{2} \sum_{i,j} \max\{g_{ij}(x_i, x_j), 0\}$, which is differentiable at any point except on a measure-zero subset $D := \{\mathbf{x} \in \mathbb{R}^n : g_{ij}(x_i, x_j) = 0 \text{ for some } i, j\}$. As before, we know that the i th component of the gradient of $\Phi(\mathbf{x})$ at \mathbf{x}^k (or the subgradient, if $\mathbf{x}^k \in D$) is given by $g_i^k = \frac{\partial}{\partial x_i} f(\mathbf{x}^k) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k)$. From the assumption, we know that there is a positive-definite diagonal matrix G^k such that

$$u_k(\mathbf{y}) := \Phi(\mathbf{x}^k) + (\mathbf{y} - \mathbf{x}^k)^T g^k + \frac{1}{2} \|\mathbf{y} - \mathbf{x}^k\|_{G^k}^2$$

forms a quadratic upper approximation for $\Phi(\mathbf{y}), \forall \mathbf{y} \in L_{\mathbf{x}^k}$, and $u_k(\mathbf{x}^k) = \Phi(\mathbf{x}^k)$. On the other hand, it is easy to see that

$$\mathbf{x}^{k+1} = \mathbf{x}^k - (G^k)^{-1} g^k = \operatorname{argmin}_{\mathbf{y} \in \mathbb{R}^n} u_k(\mathbf{y}).$$

Let us consider an arbitrary $\mathbf{x}^k \notin D$. $g^k = \nabla \Phi(\mathbf{x}^k)$, and thus $-(G^k)^{-1} g^k$ is a descent direction for any positive-definite matrix $(G^k)^{-1}$. Therefore, for sufficiently small $\delta > 0$, $\Phi(\mathbf{x}^k - \delta(G^k)^{-1} g^k) < \Phi(\mathbf{x}^k)$, and hence $\mathbf{x}^k - \delta(G^k)^{-1} g^k \in L_{\mathbf{x}^k}$. Thus, the line segment $\{(1 - \alpha)\mathbf{x}^k + \alpha\mathbf{x}^{k+1}, \alpha \in [0, 1]\}$ intersects $L_{\mathbf{x}^k}$ in at least two different points (for $\alpha = 0$ and $\alpha = \delta$). Now, if $\mathbf{x}^{k+1} \notin L_{\mathbf{x}^k}$, that line segment must intersect with the boundary of $L_{\mathbf{x}^k}$ at another point $\bar{\mathbf{x}} := \bar{\alpha}\mathbf{x}^k + (1 - \bar{\alpha})\mathbf{x}^{k+1}$, for some $\bar{\alpha} \in (0, 1)$. (Note that the level set $L_{\mathbf{x}^k}$ is a closed convex set.) If we apply the continuity of $\Phi(\cdot)$, we conclude that $\Phi(\bar{\mathbf{x}}) = \Phi(\mathbf{x}^k)$, and thus

$$u_k(\mathbf{x}^k) = \Phi(\mathbf{x}^k) = \Phi(\bar{\mathbf{x}}) \leq u_k(\bar{\mathbf{x}}) \leq \bar{\alpha} u_k(\mathbf{x}^k) + (1 - \bar{\alpha}) u_k(\mathbf{x}^{k+1}) < u_k(\mathbf{x}^k),$$

where the first inequality results from $\Phi(\mathbf{y}) \leq u_k(\mathbf{y}), \forall \mathbf{y} \in L_{\mathbf{x}^k}$, and the second inequality results from the convexity of $u_k(\cdot)$. This contradiction shows that $\mathbf{x}^{k+1} \in L_{\mathbf{x}^k}$, which implies $\Phi(\mathbf{x}^{k+1}) < \Phi(\mathbf{x}^k)$. Therefore, $\Phi(\cdot)$ serves as a semi-Lyapunov function for the dynamics (15). In particular, the drift of

this Lyapunov function at $\mathbf{x}^k \notin D$ equals

$$\begin{aligned}
\Phi(\mathbf{x}^k) - \Phi(\mathbf{x}^{k+1}) &\geq \Phi(\mathbf{x}^k) - u_k(\mathbf{x}^{k+1}) \\
&= \Phi(\mathbf{x}^k) - \left(\Phi(\mathbf{x}^k) + (\mathbf{x}^{k+1} - \mathbf{x}^k)^T g^k + \frac{1}{2} \|\mathbf{x}^{k+1} - \mathbf{x}^k\|_{G^k}^2 \right) \\
&= \frac{1}{2} (g^k)^T (G^k)^{-1} g^k \\
&= \frac{1}{2} \|g^k\|_{(G^k)^{-1}}^2.
\end{aligned}$$

Summing the above inequality for $k = 0, \dots, K - 1$, we obtain

$$\Phi(\mathbf{x}^K) + \sum_{\{k: \mathbf{x}^k \in D\}} (\Phi(\mathbf{x}^{k+1}) - \Phi(\mathbf{x}^k)) \leq \Phi(\mathbf{x}^0) - \frac{1}{2} \sum_{\{k: \mathbf{x}^k \notin D\}} \|g^k\|_{(G^k)^{-1}}^2.$$

As this relation holds for any K , and $|\{k : \mathbf{x}^k \in D\}| < \infty$ by our assumption, we must have $\sum_{\{k: \mathbf{x}^k \notin D\}} \|g^k\|_{(G^k)^{-1}}^2 < \infty$, and hence $\lim_{k \rightarrow \infty} \|g^k\|_{(G^k)^{-1}}^2 = 0$. Thus, if there exists $m > 0$ such that $G^k \leq mI, \forall k$, we get $\lim_{k \rightarrow \infty} \|g^k\| = 0$. Since $\Phi(\cdot)$ is a convex function, we know that the set of minimizers of $\Phi(\cdot)$ is exactly the set of points that have $\mathbf{0}$ as their subgradient. Thus, $\{\mathbf{x}^k\}_{k=0}^\infty$ must converge to an equilibrium point that is a global minimum of the semi-Lyapunov function $\Phi(\cdot)$. ■

A natural choice for the matrices G^k in Theorem 7 is the Hessian matrix $\nabla^2 \Phi(\mathbf{x}^k)$, which is used in the Newton method for minimizing a smooth convex function. However, in practice, it is often easier to work with a sparse modification of $\nabla^2 \Phi(\mathbf{x}^k)$ given by a diagonal matrix containing only the diagonal entries of $\nabla^2 \Phi(\mathbf{x}^k)$. In particular, to assure positive definiteness, an identity matrix is added to such a diagonal matrix to form the quasi-Newton update rule. Using such a quasi-Newton method for minimizing $\Phi(\cdot)$, one obtains the following state-dependent network dynamics:

$$x_i^{k+1} = x_i^k - t_k \frac{\frac{\partial}{\partial x_i} f(\mathbf{x}^k) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial}{\partial x_i} g_{ij}(x_i^k, x_j^k)}{1 + \frac{\partial^2}{\partial x_i^2} f(\mathbf{x}^k) + \sum_{j \in N_i(\mathbf{x}^k)} \frac{\partial^2}{\partial x_i^2} g_{ij}(x_i^k, x_j^k)}, \quad (16)$$

where t_k is an appropriately chosen step size obtained using a line search or diminishing rule. In fact, it is known that for a sufficiently small neighborhood of the minimizers of $\Phi(\cdot)$, the Newton method with step size $t_k = 1$ will converge quadratically fast to the set of optimal points [35]. Therefore, we obtain a simple explanation for the convergence properties and equilibrium points of seemingly complex state-dependent network dynamics (16) by using the well-known quasi-Newton method. In particular, this view provides a rigorous explanation of why the trajectories of the state-dependent network dynamics of the form (16) (such as the HK model) converge exponentially fast as they get close to their equilibrium points.

Example 6: Let us consider a special case in which $f = 0$ and $g_{ij}(x_i, x_j) = \frac{1}{2}(x_i - x_j)^2 - \frac{\epsilon_{ij}^2}{2}$, where $\epsilon_{ij} = \epsilon_{ji} > 0$. Two agents i and j become each other's neighbors if their distance from each other is larger than ϵ_{ij} . Therefore, we obtain the complement of the original HK model. Here, $\Phi(\mathbf{x}) = \frac{1}{2} \sum_{ij} \max\{\frac{1}{2}(x_i - x_j)^2 - \frac{\epsilon_{ij}^2}{2}, 0\}$, and thus for $\mathbf{x}^k \notin D := \{\mathbf{x} : |x_i - x_j| = \epsilon_{ij}, \text{ for some } i, j\}$, we have

$$\begin{aligned}
\nabla_i \Phi(\mathbf{x}^k) &= \sum_{j \in N_i(\mathbf{x}^k)} (x_i^k - x_j^k) = |N_i(\mathbf{x}^k)| x_i^k - \sum_{j \in N_i(\mathbf{x}^k)} x_j^k, \\
\nabla_{ij}^2 \Phi(\mathbf{x}^k) &= \begin{cases} |N_i(\mathbf{x}^k)| & \text{if } i = j \\ -1 & \text{if } j \in N_i(\mathbf{x}^k) \\ 0 & \text{else.} \end{cases}
\end{aligned}$$

In other words, the Hessian matrix at \mathbf{x}^k is equal to the Laplacian of the connectivity network at state \mathbf{x}^k . As a result, the quasi-Newton dynamics (16) for minimizing the piecewise quadratic function $\Phi(\mathbf{x})$ become

$$x_i^{k+1} = x_i^k - t_k \frac{|N_i(\mathbf{x}^k)|x_i^k - \sum_{j \in N_i(\mathbf{x}^k)} x_j^k}{|N_i(\mathbf{x}^k)| + 1}.$$

In particular, for a sufficiently small choice of step size t_k , the function $\Phi(\mathbf{x})$ serves as a semi-Lyapunov function. Note that for unit step size $t_k = 1$, the above dynamics can be explicitly written as

$$\mathbf{x}_i^{k+1} = \frac{\sum_{j \in N_i(\mathbf{x}^k) \cup \{i\}} x_j^k}{|N_i(\mathbf{x}^k)| + 1}, \quad i \in [n]. \quad (17)$$

As a result, the dynamics of the complement HK model can be viewed as iterates of a quasi-Newton method with a unit step size to minimize $\Phi(\mathbf{x})$. Of course, for $t_k = 1$, there is no reason why $\Phi(\mathbf{x})$ should serve as a Lyapunov function, unless the initial point of the dynamics is sufficiently close to a minimizer of $\Phi(\cdot)$ (in which case the exponentially fast convergence of the quasi-Newton method with $t_k = 1$ is guaranteed). Nevertheless, the function $\Phi(\mathbf{x})$ is still very useful, as it globally guides the dynamics based on quasi-Newton iterates. In particular, the set of minimizers of $\Phi(\mathbf{x})$ characterize the equilibrium points of (17).⁵ This is because if $\lim_k \mathbf{x}^k = \mathbf{x}^*$, we must have $\mathbf{x}^* = \lim_k \mathbf{x}^{k+1} = \mathbf{x}^* - \lim_k (G^k)^{-1} \nabla \Phi(\mathbf{x}^k)$, where here $(G^k)^{-1} = \text{diag}(\frac{1}{|N_1(\mathbf{x}^k)|+1}, \dots, \frac{1}{|N_n(\mathbf{x}^k)|+1})$. This implies that $\lim_{k \rightarrow \infty} G_k^{-1} \nabla \Phi(\mathbf{x}^k) = 0$. As the entries of G_k^{-1} are uniformly bounded below by $\frac{1}{n+1}$, we must have $\lim_{k \rightarrow \infty} \nabla \Phi(\mathbf{x}^k) = 0$, and the result follows from the convexity of $\Phi(\cdot)$.

V. CONTINUOUS-TIME CONSTRAINED SADDLE-POINT DYNAMICS

In this section, we extend our discrete-time saddle-point dynamics to their continuous-time counterparts and show how they can be leveraged to establish Lyapunov stability of state-dependent network dynamics. Here, because of the continuity of the time index $t \in [0, \infty)$, an edge connectivity between a pair of agents (i, j) is no longer a binary event $\lambda_{ij} \in \{0, 1\}$, but rather a continuous weight function of time $\lambda_{ij}(t) \in [0, 1]$. Thus, $\lambda_{ij}(t)$ can be viewed as a connectivity strength between agents i and j at time t such that the maximum influence that the two agents can have on each other is 1 (i.e., they are fully connected) and the minimum influence is 0 (i.e., they are fully disconnected).

Motivated by the method of change of variables for discrete time dynamics in Section III-B, we state our results for continuous-time dynamics in a more general form wherein the agents' states are transformed from x_i to $p_i(x_i)$, and the network variables are transformed from λ_{ij} to $q_{ij}(\lambda_{ij})$. Here, we assume that $p_i(\cdot), q_{ij}(\cdot)$ are continuous and nondecreasing functions such that $p_i(0) = q_{ij}(0) = 0, \forall i, j$. In particular, we let the Lagrangian function have a more general form of $L(p(\mathbf{x}), q(\boldsymbol{\lambda}))$, as long as its partial derivatives exist, and be convex with respect to its first argument $p(\mathbf{x}) = (p_1(x_1), \dots, p_n(x_n))^T$ and concave with respect to its second argument $q(\boldsymbol{\lambda}) = (q_{ij}(\lambda_{ij}), i \neq j)^T$.

Remark 4: A special case of the above setting occurs when $p_i(x_i) = x_i, q_{ij}(\lambda_{ij}) = \lambda_{ij}$ are identity functions, and $L(\mathbf{x}, \boldsymbol{\lambda}) = \sum_i f_i(x_i) + \sum_{i \neq j} \lambda_{ij} g_{ij}(x_i, x_j), \boldsymbol{\lambda} \geq 0, \mathbf{x} \in \mathbb{R}^n$. It is clear that for the convex measurement functions g_{ij}, f_i , the standard Lagrangian function $L(\mathbf{x}, \boldsymbol{\lambda})$ is convex with respect to \mathbf{x} , and concave (i.e., linear) with respect to $\boldsymbol{\lambda}$.

To introduce a general class of continuous-time, state-dependent network dynamics, let us consider the following static constrained saddle-point problem:

$$\min_{\mathbf{x} \in \mathbb{R}^n} \max_{\boldsymbol{\lambda} \in [0, 1]^{n(n-1)}} L(p(\mathbf{x}), q(\boldsymbol{\lambda})). \quad (18)$$

⁵In fact, using a sorted vector Lyapunov function $V(\mathbf{x}) = \text{sort}(\{|x_i - x_j|, i \neq j\})$, one can show that the dynamics (17) do converge as $V(\mathbf{x})$ decreases lexicographically after each iteration.

To solve the above static saddle-point problem using continuous-time dynamics, we use the idea of a *gradient flow*, which was initially introduced in the seminal work of Arrow, Hurwicz, and Uzawa [36] and subsequently used in devising primal-dual algorithms for solving constrained optimization problems [37]. However, to adopt those dynamics for our more general setting (18), which has both lower- and upper-bound constraints on the dual variable λ , we introduce the following generalized gradient flow dynamics:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= -\nabla_{p(\mathbf{x})}L(p(\mathbf{x}), q(\boldsymbol{\lambda})) \\ \dot{\boldsymbol{\lambda}}(t) &= [\nabla_{q(\boldsymbol{\lambda})}L(p(\mathbf{x}), q(\boldsymbol{\lambda}))]_{\lambda}^{[0,1]},\end{aligned}\quad (19)$$

where $\nabla_{p(\mathbf{x})}L(p(\mathbf{x}), q(\boldsymbol{\lambda})) := \left(\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial p_1(x_1)}, \dots, \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial p_n(x_n)}\right)^T$ (the gradient vector $\nabla_{q(\boldsymbol{\lambda})}L(p(\mathbf{x}), q(\boldsymbol{\lambda}))$ is defined analogously), and $[a]_{\lambda}^{[0,1]}$ denotes the projection of the network dynamics to the unit interval, i.e.,

$$[a]_{\lambda}^{[0,1]} = \begin{cases} \min\{0, a\}, & \text{if } \lambda = 1 \\ a & \text{if } 0 < \lambda < 1 \\ \max\{0, a\} & \text{if } \lambda = 0. \end{cases}$$

When a is a vector rather than a scalar, the above projection is taken coordinatewise. The reason for introducing such a projection is that if for a pair of agents (i, j) we have $\lambda_{ij}(t) \in (0, 1)$, the edge variable $\lambda_{ij}(t)$ has not hit the boundary points $\{0, 1\}$, and it can freely increase or decrease without violating the box constraint $\lambda_{ij}(t) \in [0, 1]$. However, if $\lambda_{ij}(t) = 1$, then this edge variable is only allowed to decrease, and thus $\dot{\lambda}_{ij}(t) \leq 0$. Therefore, if $\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})} \geq 0$, we set $\dot{\lambda}_{ij}(t) = 0$ to block any further increase of $\lambda_{ij}(t)$. Similarly, if $\lambda_{ij}(t) = 0$ and $\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})} \leq 0$, we set $\dot{\lambda}_{ij}(t) = 0$ to block any further decrease of $\lambda_{ij}(t)$. Therefore, (19) provides a fairly general class of continuous-time, state-dependent network dynamics in which the strength of the edge connectivity changes dynamically as a function of the state variables.

Remark 5: It is worth noting that in the special setting of Remark 4, the network dynamics in (19) decompose to a simple form of $\dot{\lambda}_{ij}(t) = [g_{ij}(x_i(t), x_j(t))]_{\lambda_{ij}}^{[0,1]}$, $\forall i, j$. Thus, the more distant two agents i and j are from each other (i.e., the larger the measurement value $g_{ij}(x_i, x_j)$), the faster the edge connectivity between them grows and until it achieves its maximum connectivity at 1. That is analogous to the emergence of an edge between agents i and j if $g_{ij}(x_i, x_j) > 0$ in the discrete-time setting.

To establish the Lyapunov stability of the continuous-time state-network dynamics (19), let $(\bar{\mathbf{x}}, \bar{\boldsymbol{\lambda}})$ be a saddle-point solution to (18). Note that by continuity and the convex-concave property of the Lagrangian function, the existence of a saddle-point in (18) is always guaranteed. Let us define $P_i(x_i) := \int_{\bar{x}_i}^{x_i} p_i(s) ds$ and $Q_{ij}(\lambda_{ij}) := \int_{\bar{\lambda}_{ij}}^{\lambda_{ij}} q_{ij}(s) ds$, where we note that by continuity and the monotonicity of p_i, q_{ij} , the functions P_i and Q_{ij} are differentiable convex functions. Now we are ready to state the main result of this section.

Theorem 8: Let $L(p(\mathbf{x}), q(\boldsymbol{\lambda}))$ be a convex function in $p(\mathbf{x})$ and a concave function in $q(\boldsymbol{\lambda})$. Then, the continuous-time state-dependent network dynamics (19) are Lyapunov stable. In particular,

$$V(\mathbf{x}, \boldsymbol{\lambda}) := \sum_{i=1}^n D_{P_i}(x_i, \bar{x}_i) + \sum_{i \neq j} D_{Q_{ij}}(\lambda_{ij}, \bar{\lambda}_{ij})$$

serves as a Lyapunov function for the dynamics (19), where $D_{\phi}(u, v) = \phi(u) - \phi(v) - \phi'(v)(u - v)$ denotes the Bregman divergence with respect to the convex function $\phi(\cdot)$.

Proof: Using the definition of the Bregman divergence, for every i and j we have

$$\begin{aligned}\dot{D}_{P_i}(x_i, \bar{x}_i) &= \frac{\partial D_{P_i}(x_i, \bar{x}_i)}{\partial x_i} \dot{x}_i = -(p_i(x_i) - p_i(\bar{x}_i)) \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial p_i(x_i)}, \\ \dot{D}_{Q_{ij}}(\lambda_{ij}, \bar{\lambda}_{ij}) &= \frac{\partial D_{Q_{ij}}(\lambda_{ij}, \bar{\lambda}_{ij})}{\partial \lambda_{ij}} \dot{\lambda}_{ij} = (q_{ij}(\lambda_{ij}) - q_{ij}(\bar{\lambda}_{ij})) \left[\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})} \right]_{\lambda_{ij}}^{[0,1]}.\end{aligned}$$

Now we can write

$$\begin{aligned}\dot{V}(\mathbf{x}, \boldsymbol{\lambda}) &= - \sum_i (p_i(x_i) - p_i(\bar{x}_i)) \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial p_i(x_i)} + \sum_{i \neq j} (q_{ij}(\lambda_{ij}) - q_{ij}(\bar{\lambda}_{ij})) \left[\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})} \right]_{\lambda_{ij}}^{[0,1]} \\ &\leq - \sum_i (p_i(x_i) - p_i(\bar{x}_i)) \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial p_i(x_i)} + \sum_{i \neq j} (q_{ij}(\lambda_{ij}) - q_{ij}(\bar{\lambda}_{ij})) \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})} \\ &= \left(\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial p(\mathbf{x})} \right)^T (p(\bar{\mathbf{x}}) - p(\mathbf{x})) + \left(\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q(\boldsymbol{\lambda})} \right)^T (q(\bar{\boldsymbol{\lambda}}) - q(\boldsymbol{\lambda})) \\ &\leq L(p(\bar{\mathbf{x}}), q(\boldsymbol{\lambda})) - L(p(\mathbf{x}), q(\boldsymbol{\lambda})) - (L(p(\mathbf{x}), q(\bar{\boldsymbol{\lambda}})) - L(p(\mathbf{x}), q(\boldsymbol{\lambda}))) \\ &= \left[L(p(\bar{\mathbf{x}}), q(\boldsymbol{\lambda})) - L(p(\bar{\mathbf{x}}), q(\bar{\boldsymbol{\lambda}})) \right] + \left[L(p(\bar{\mathbf{x}}), q(\bar{\boldsymbol{\lambda}})) - L(p(\mathbf{x}), q(\bar{\boldsymbol{\lambda}})) \right] \leq 0,\end{aligned}$$

where in the above derivations the last inequality is due to the definition of the saddle-point, and the second inequality follows from the convexity/concavity of $L(\cdot)$ with respect to its first/second argument. Finally, the first inequality is obtained by considering the following three cases:

- If $\lambda_{ij} = 0$, then $\left[\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})} \right]_{\lambda_{ij}}^{[0,1]} = \max\{0, \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})}\} \geq \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})}$, and $q_{ij}(\lambda_{ij}) - q_{ij}(\bar{\lambda}_{ij}) = q_{ij}(0) - q_{ij}(\bar{\lambda}_{ij}) \leq 0$.
- If $\lambda_{ij} \in (0, 1)$, then $\left[\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})} \right]_{\lambda_{ij}}^{[0,1]} = \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})}$.
- If $\lambda_{ij} = 1$, then $\left[\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})} \right]_{\lambda_{ij}}^{[0,1]} = \min\{0, \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})}\} \leq \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})}$, and $q_{ij}(\lambda_{ij}) - q_{ij}(\bar{\lambda}_{ij}) = q_{ij}(1) - q_{ij}(\bar{\lambda}_{ij}) \geq 0$.

Thus, in either of the above cases, we have

$$(q_{ij}(\lambda_{ij}) - q_{ij}(\bar{\lambda}_{ij})) \left[\frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})} \right]_{\lambda_{ij}}^{[0,1]} \leq (q_{ij}(\lambda_{ij}) - q_{ij}(\bar{\lambda}_{ij})) \frac{\partial L(p(\mathbf{x}), q(\boldsymbol{\lambda}))}{\partial q_{ij}(\lambda_{ij})},$$

and the result follows. ■

It is worth noting that Theorem 8 is a continuous-time counterpart of Theorem 6 in the sense that in both theorems, the Bregman distance of the iterates to a saddle-point serves as a Lyapunov function. However, because of the continuity of the network variables in the continuous-time model, the choice of the step size becomes irrelevant in Theorem 8, while for the discrete-time counterpart, the step sizes should be small enough to guarantee the convergence of the dynamics.

VI. SIMULATIONS

In this section, we describe several numerical experiments relating to social science to justify our theoretical results. In the first experiment, we considered a class of averaging dynamics with negative weights motivated by the presence of antagonistic relations in social groups such as Altafini's averaging dynamics [38]. The dynamics that we consider have the form of $x_i^{k+1} = \frac{|N_i(\mathbf{x}^k)|}{|N_i(\mathbf{x}^k)|+1} x_i^k - \frac{\sum_{j \in N_i(\mathbf{x}^k)} x_j^k}{|N_i(\mathbf{x}^k)|+1}$, $i \in [n]$, where $N_i(\mathbf{x}^k) = \{j \in [n] \setminus \{i\} : x_i^k x_j^k \leq 1\}$. In other words, two agents i and j are each other's neighbors if the product of their states is less than 1.⁶ On the left side of Figure 4, we depict the

⁶Similar results can be obtained if one replaces 1 in the definition of the agents' neighborhood with another constant, such as 0, in which case agents communicate only if their states have the opposite sign.

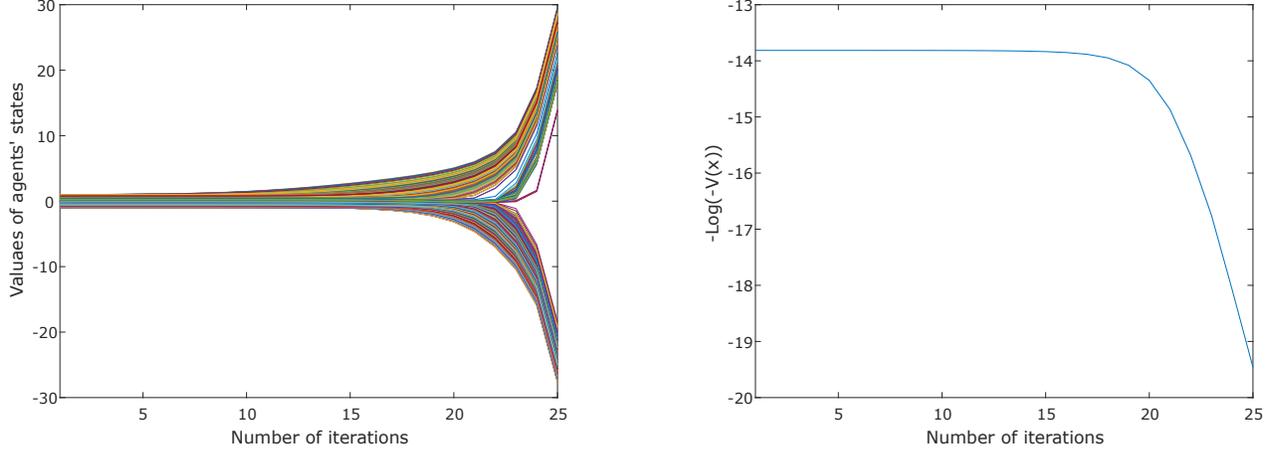


Fig. 4. A diverging averaging dynamic with negative weights and its associated energy function.

evolution of these dynamics for $k = 25$ iterations and $n = 1000$ agents, where the agents' initial states are distributed uniformly at random in the interval $[-1, 1]$. As can be seen, the agents eventually polarized to two groups and diverged to $+\infty$ and $-\infty$. On the other hand, it is easy to see that the above dynamics can be replicated with Theorem 3 if the symmetric measurements $g_{ij}(x_i, x_j) = x_i x_j - 1$ and $f_i(x_i) = \frac{x_i^2}{2}$ are chosen. The implication is that $V(x) = \sum_{i=1}^n \frac{x_i^2}{2} + \sum_{i,j} \min\{x_i x_j - 1, 0\}$ must decrease along the trajectories of the dynamics, as is shown on the right side of Figure 4. However, we note that the measurements $g_{ij}(x_i, x_j) = x_i x_j - 1$ are neither convex nor bounded below by a global constant. In fact, that is the main reason why the monotonically decreasing function $V(x)$ cannot guarantee the convergence of the dynamics. Therefore, the convexity and boundedness of measurements in Theorem 3 are necessary for the convergence of the trajectories.

In the second experiment, we considered the dynamics (17) in Example 6. Here, the symmetric measurements are $g_{ij}(x_i, x_j) = \frac{(x_i - x_j)^2}{2} - \frac{20^2}{2}$ (i.e., $\epsilon_{ij} = 20, \forall i \neq j$), and the set of neighbors of agent i at iteration k is given by $N_i(\mathbf{x}^k) = \{j : |x_i^k - x_j^k| \geq 20\}$. As was shown in Example 6, those dynamics are the complement of the homogeneous HK dynamics. The left side of Figure 5 shows the results of simulating the trajectories for $k = 120$ iterations and $n = 1000$ agents, with initial states distributed uniformly at random in the interval $[0, 100]$. As can be seen, although the dynamics eventually converge to several clusters, however, there is an oscillating pattern in the trajectories due to the conflicting objectives of the network structure and the agents' states. In particular, Theorem 7 suggests that the convex function $V(\mathbf{x}) = \frac{1}{2} \sum_{i,j} \max\{\frac{(x_i - x_j)^2}{2} - \frac{20^2}{2}, 0\}$ serves as a semi-Lyapunov function that almost always decreases along the trajectories. That phenomenon is shown on the right side of Figure 5, where $V(\mathbf{x})$ always decreases with a small jump at iteration $k = 3$. In particular, the dynamics converge to a minimizer of $V(\mathbf{x})$.

In our last experiment, we considered *heterogeneous* HK dynamics with associated dynamics $x_i^{k+1} = \frac{\sum_{j: |x_i^k - x_j^k| \leq \epsilon_i} x_j^k}{|\{j: |x_i^k - x_j^k| \leq \epsilon_i\}|}, i \in [n]$, whose Lyapunov stability and convergence have remained open for more than a decade [39], [40]. We can replicate the heterogeneous HK dynamics by considering the asymmetric measurements $g_{ij}(x_i, x_j) = \frac{(x_i - x_j)^2}{2} - \frac{\epsilon_i^2}{2}$ in the statement of Theorem 4. Although we do not know a choice of Bregman map $f(x)$ that can satisfy the assumption of Theorem 4 with respect to such asymmetric measurements, in Figure 6 we show that even a simple quadratic map performs quite well for the heterogeneous HK model. More precisely, based on the quadratic Bregman map $D_f(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|^2$, where $f(x) = \sum_{i=1}^n \frac{x_i^2}{2}$, Theorem 4 suggests the Lyapunov candidate $V(\mathbf{x}) = \sum_{i,j} \min\{\frac{(x_i - x_j)^2}{2} - \frac{\epsilon_i^2}{2}, 0\}$ for the heterogeneous HK dynamics. We simulated those dynamics for $k = 300$ iterations over a

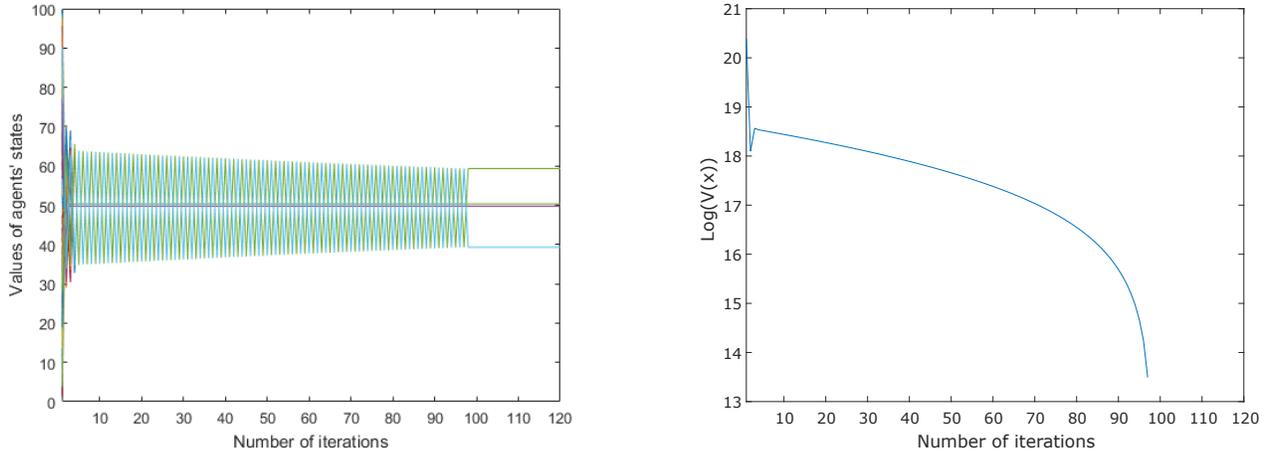


Fig. 5. Dynamics of the complement HK model and its associated semi-Lyapunov function.

set of $n = 1000$ agents whose initial states were distributed uniformly at random in the interval $[0, 100]$. In the top two graphs in Figure 6, the confidence bounds of agents $\epsilon_i, i \in [n]$ had been selected uniformly at random from the interval $[0, 10]$. In the bottom two graphs in Figure 6, the confidence bounds had been generated uniformly at random from the larger interval $[0, 50]$. In both cases, one can see that the proposed function $V(x)$ performed quite well and almost always decreased along the trajectories of the heterogeneous HK. That suggests that addition of a small correction term to the proposed $V(x)$ might turn this function into a valid Lyapunov function for the heterogeneous HK dynamics.

VII. CONCLUSIONS

In this paper, we developed a new framework for the stability analysis of multiagent state-dependent network dynamics. We showed that the co-evolution of the network and the state dynamics could be cast as a primal-dual optimization algorithm for a nonlinear program in which the primal updates capture the state dynamics, and the dual updates capture the network evolution. In particular, the constrained Lagrangian function serves as a Lyapunov function for the state-network dynamics. We considered our framework under two different settings: i) when the network and state dynamics are aligned, and ii) when the network and state dynamics have conflicting objectives. In the first case, we showed that the application of the BCD method with a change of variables could generate a variety of interesting state-dependent network dynamics. In particular, we provided a new technique for handling asymmetry in the network dynamics. In the second case, we reduced the stability of the state-network dynamics to a zero-sum game between the network player and the state player. This approach allowed us to establish the Lyapunov stability of multiagent systems by using saddle-point dynamics and, in particular, by using the subgradient method and the quasi-Newton method. Finally, we extended our results to a continuous-time model and provided a general class of continuous-time, state-dependent network dynamics in terms of generalized gradient flow.

As a future direction of research, one could use *augmented* Lagrangian functions or apply other optimization techniques to generate a broader class of stable state-dependent network dynamics. Moreover, in our analysis, we mainly used a quadratic upper approximation to derive the state updates. Thus, a natural extension would be to use other function approximations that include the quadratic approximation as their particular case, or to use approximations that are suitable for specific applications.

REFERENCES

- [1] S. R. Etesami, "A simple framework for stability analysis of state-dependent networks of heterogeneous agents," *SIAM Journal on Control and Optimization*, vol. 57, no. 3, pp. 1757–1782, 2019.

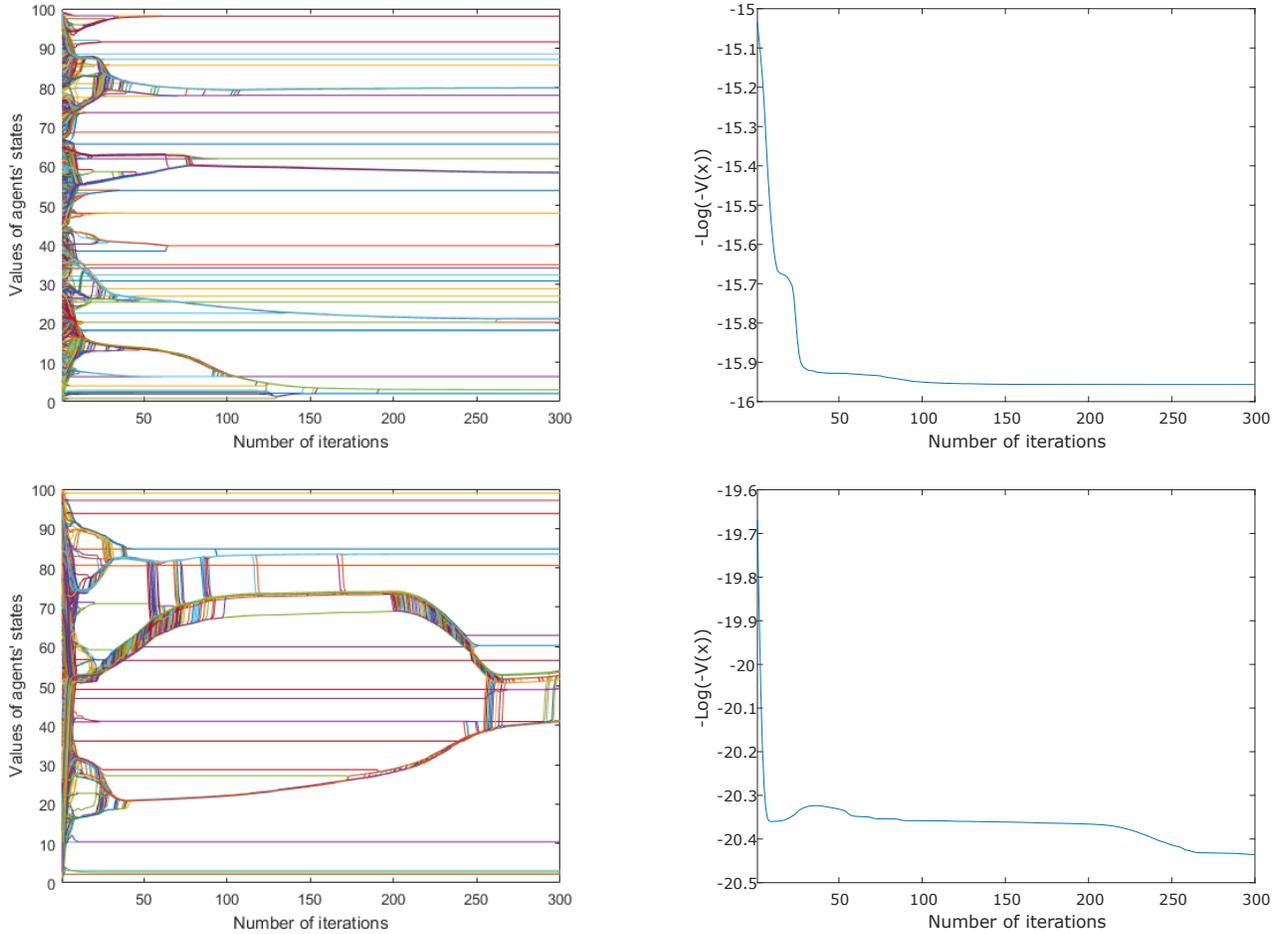


Fig. 6. Evolution of the heterogeneous HK dynamics and their associated $V(x)$ values with uniformly generated confidence bounds from intervals $[0, 10]$ (top figures) and $[0, 50]$ (bottom figures).

- [2] J. Grossklags, N. Christin, and J. Chuang, "Secure or insure?: A game-theoretic analysis of information security games," in *Proceedings of the 17th International Conference on World Wide Web*. ACM, 2008, pp. 209–218.
- [3] F. Bullo, J. Cortes, and S. Martinez, *Distributed Control of Robotic Networks: A Mathematical Approach to Motion Coordination Algorithms*. Princeton University Press, 2009, vol. 27.
- [4] R. Hegselmann and U. Krause, "Opinion dynamics and bounded confidence models, analysis and simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, no. 3, pp. 1–33, 2002.
- [5] D. V. Dimarogonas and K. H. Johansson, "Decentralized connectivity maintenance in mobile networks with bounded inputs," in *IEEE International Conference on Robotics and Automation*, 2008, pp. 1507–1512.
- [6] M. M. Zavlanos and G. J. Pappas, "Potential fields for maintaining connectivity of mobile networks," *IEEE Transactions on Robotics*, vol. 23, no. 4, pp. 812–816, 2007.
- [7] I. Rajapakse, M. Groudine, and M. Mesbahi, "Dynamics and control of state-dependent networks for probing genomic organization," *Proceedings of the National Academy of Sciences*, vol. 108, no. 42, pp. 17 257–17 262, 2011.
- [8] A. Cherukuri, E. Mallada, S. Low, and J. Cortes, "The role of convexity in saddle-point dynamics: Lyapunov function and robustness," *IEEE Transactions on Automatic Control*, vol. 63, no. 8, pp. 2449–2464, 2017.
- [9] M. Mesbahi, "On state-dependent dynamic graphs and their controllability properties," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 387–392, 2005.
- [10] A. Awad, A. Chapman, E. Schoof, A. Narang-Siddharth, and M. Mesbahi, "Time-scale separation in networks: State-dependent graphs and consensus tracking," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 1, pp. 104–114, 2018.
- [11] A. Nedić, A. Olshevsky, and M. G. Rabbat, "Network topology and communication-computation tradeoffs in decentralized optimization," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 953–976, 2018.
- [12] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.
- [13] N. E. Friedkin and E. C. Johnsen, "Social positions in influence networks," *Social Networks*, vol. 19, no. 3, pp. 209–222, 1997.
- [14] A. V. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part I," *Annual Reviews in Control*, vol. 43, pp. 65–79, 2017.

- [15] R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1520–1533, 2004.
- [16] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times*. American Mathematical Soc., 2017, vol. 107.
- [17] A. Olshevsky and J. N. Tsitsiklis, "Convergence speed in distributed consensus and averaging," *SIAM Journal on Control and Optimization*, vol. 48, no. 1, pp. 33–55, 2009.
- [18] A. Nedić, A. Olshevsky, A. Ozdaglar, and J. N. Tsitsiklis, "On distributed averaging algorithms and quantization effects," *IEEE Transactions on Automatic Control*, vol. 54, no. 11, pp. 2506–2517, 2009.
- [19] A. Nedić and A. Olshevsky, "Distributed optimization over time-varying directed graphs," *IEEE Transactions on Automatic Control*, vol. 60, no. 3, pp. 601–615, 2015.
- [20] T. Başar, S. R. Etesami, and A. Olshevsky, "Convergence time of quantized Metropolis consensus over time-varying networks," *IEEE Transactions on Automatic Control*, vol. 61, no. 12, pp. 4048–4054, 2016.
- [21] S. R. Etesami and T. Başar, "Convergence time for unbiased quantized consensus over static and dynamic networks," *IEEE Transactions on Automatic Control*, vol. 61, no. 2, pp. 443–455, 2016.
- [22] S. R. Etesami, *Potential-based Analysis of Social, Communication, and Distributed Networks*. Springer, 2017.
- [23] M. Zhu and S. Martínez, "On the convergence time of asynchronous distributed quantized averaging algorithms," *IEEE Transactions on Automatic Control*, vol. 56, no. 2, pp. 386–390, 2011.
- [24] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," in *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, 2005, pp. 2996–3000.
- [25] J. M. Hendrickx and J. N. Tsitsiklis, "Convergence of type-symmetric and cut-balanced consensus seeking systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 1, pp. 214–218, 2013.
- [26] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Transactions on Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.
- [27] I. M. Sonin *et al.*, "The decomposition-separation theorem for finite nonhomogeneous Markov chains and related problems," in *Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz*. Institute of Mathematical Statistics, 2008, pp. 1–15.
- [28] B. Touri and A. Nedić, "On existence of a quadratic comparison function for random weighted averaging dynamics and its implications," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*, 2011, pp. 3806–3811.
- [29] M. Burger, D. Zelazo, and F. Allgower, "Hierarchical clustering of dynamical networks using a saddle-point analysis," *IEEE Transactions on Automatic Control*, vol. 58, no. 1, pp. 113–124, 2012.
- [30] M. Bürger, D. Zelazo, and F. Allgöwer, "Duality and network theory in passivity-based cooperative control," *Automatica*, vol. 50, no. 8, pp. 2051–2061, 2014.
- [31] A. V. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. Part II," *Annual Reviews in Control*, vol. 45, pp. 166–190, 2018.
- [32] S. Bubeck *et al.*, "Convex optimization: Algorithms and complexity," *Foundations and Trends in Machine Learning*, vol. 8, no. 3-4, pp. 231–357, 2015.
- [33] M. Roozbehani, A. Megretski, and E. Frazzoli, "Lyapunov analysis of quadratically symmetric neighborhood consensus algorithms," in *Proc. 47th IEEE Conference on Decision and Control (CDC)*. IEEE, 2008, pp. 2252–2257.
- [34] P.-E. Jabin and S. Motsch, "Clustering and asymptotic behavior in opinion formation," *Journal of Differential Equations*, vol. 257, no. 11, pp. 4165–4187, 2014.
- [35] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [36] K. J. Arrow, L. Hurwicz, and H. Uzawa, *Studies in Linear and Non-linear Programming*. Cambridge University Press, 1958.
- [37] D. Feijer and F. Paganini, "Stability of primal-dual gradient dynamics and applications to network optimization," *Automatica*, vol. 46, no. 12, pp. 1974–1981, 2010.
- [38] C. Altafini, "Dynamics of opinion forming in structurally balanced social networks," in *Proceedings of the 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 5876–5881.
- [39] A. Mirtabatabaei and F. Bullo, "Opinion dynamics in heterogeneous networks: Convergence conjectures and theorems," *SIAM Journal on Control and Optimization*, vol. 50, no. 5, pp. 2763–2785, 2012.
- [40] B. Chazelle and C. Wang, "Inertial Hegselmann-Krause systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3905–3913, 2016.